CHAPTER 16

INSIGHTS FROM ACQUISITION AND LEARNING

HOW PHONOLOGICAL REPRESENTATIONS DEVELOP DURING FIRST-LANGUAGE ACQUISITION
KATHERINE DEMUTH AND JAE YUNG SONG

SPEECH PROCESSING IN BILINGUAL AND MULTILINGUAL LISTENERS
PAOLA ESCUDERO

SECOND-LANGUAGE SPEECH LEARNING
RAJKA SMILJANIC
INTRODUCTION

The contributions in this chapter discuss the role of language development in early acquisition, multilingualism, and second-language acquisition, and consider how these inform our understanding of core phonological questions. Together they paint a picture of the critical role of both production and perception in the learning of phonological systems and show how such acquisition studies provide insight into the nature of adult phonological structure.

16.1 HOW PHONOLOGICAL REPRESENTATIONS DEVELOP DURING FIRST-LANGUAGE ACQUISITION

Katherine Demuth and Jae Yung Song

16.1.1 Introduction

Little is known about the phonological representations that underlie children’s early productions, and why variability in production persists even as the child progresses toward the mastery of language. Much of the early research on phonological acquisition focuses on children’s production of segments, showing variability both between and within individuals (e.g., Smith 1973; Ferguson et al. 1992; Vihman 1993, 1996). Some of this research identified children’s early motor-control limitations as a means for understanding their variable productions (e.g., MacNeilage 1980; Lindblom 1992). Others have shown that within-speaker variability is influenced by the frequency of lexical and syllable patterns in the ambient language (e.g., Berkman and Edwards 2000b; Levelt et al. 2000; Roark and Demuth 2000; Edwards et al. 2004; Storkel 2004; Zamuner et al. 2004; Munson et al. this volume). Still others have shown that the phonological contexts in which words and morphemes appear can have an enormous effect on whether a morpheme is apparently produced or not (e.g., Pangos et al. 1979; Bennett and Ingle 1984; Echols and Newport 1992; Gerken and McIntosh 1993; Rvachew and Andrews 2002). Variable processes of coda deletion and coda cluster reduction are also subject to contextual variation within certain dialects. This has been investigated for adults speaking American and

* We thank our collaborators Stefanie Shattuck-Hufnagel and Lucy Menard for their contributions. This work was funded in part by NICHD grant #R01 HD057606.
British English (e.g., Roberts 1997; Foulkes et al. 2005; Docherty et al. 2006), and African-American English (e.g., Moran 1993; Stockman and Vaughn-Cooke 1989; Stockman 2006; Wolfram 1991; Bailey and Thomas 1998; Rickford 1999). Such adult variation is an important issue to keep in mind when examining child phonological and morphological development across dialects.

In this section, we argue that children's phonological representations as evidenced by their productions may be more intact than often assumed. We also suggest that conducting fine-grained acoustic analysis of child (and child-directed) speech holds the potential for better understanding children's developing phonological representations, and the factors that influence variability in production over time. We review below some of the traditional methods used, discussing some of their limitations, and then discuss recent laboratory phonology research examining the development of phonological representations as evidenced through production.

16.1.2 Traditional methods and some limitations

16.1.2.1 Observational/longitudinal studies

Many investigations of children’s phonological development have been observational case studies, where longitudinal data is collected and developmental trends assessed. Some consist of diary studies (e.g., Deville 1891), whereas others consist of tape-recorded and transcribed child speech, using either orthographic (Brown 1973) or phonetic (IPA) form (Smith 1973). Such studies provide useful albeit impressionistic information about a child’s language development, upon which many theoretical claims have been made. In fact, acoustic analysis is critical to fully analyze these data. For example, if the transcription indicates that the child produced no coda consonant on the word dog, it is impossible to know if the child’s representation was really CV, or if there might have been vowel lengthening, indicating that the child has some knowledge of the “missing” coda consonant. Nonetheless, these types of spontaneous, longitudinal corpora can be extremely useful as pilot data for forming hypotheses about aspects of phonological development, which could be investigated under more controlled, experimental conditions. They are also useful in documenting individual differences in phonological development. However, for any specific research question there may not be enough tokens of the right type from spontaneous speech corpora to fully assess the extent of children’s phonological knowledge.

16.1.2.2 Experimental production studies

Some of the concerns about sparse data can be addressed in cross-sectional experimental studies using elicited imitation or elicited production methods. This
provides the opportunity for exploring children’s phonological and morphological knowledge under controlled contexts at a given point in time. For example, Kirk and Demuth (2005) compared 2-year-olds’ acquisition of segmentally similar consonant clusters at the beginnings and ends of words (ski vs ask, ax). They found that children were better at producing consonant clusters word-finally, especially when these decreased in sonority (i.e., ask [ask] was produced more accurately than ax [aks]). However, it is also possible that some of children’s early cluster errors could be due to articulatory difficulty. For example, Kirk (2006) examined 2-year-olds’ coda productions in monosyllabic and disyllabic non-words. They found better coda production in monosyllabic words, and in the final and stressed syllables of disyllabic words. Similarly, Song et al. (2009) found better production of third person singular -s in utterance-final compared with utterance-medial position—for both 1;10-year-olds and 2;3-year-olds (with the older children doing better overall). Taken together, these results suggest that children are more accurate in producing coda consonants in stressed and final syllables, which are longer in duration, thus allowing more time to complete the full articulation. On the other hand, many experimental production studies have not necessarily examined the data from a more fine-grained acoustic perspective.

16.1.2.3 Experimental perception studies

There is a wealth of literature examining the development of infants’ perceptual abilities (see Holt, this volume, Munson et al., this volume, Maye, this volume). Some of this literature has focused on the development of native vs non-native speech contrasts in infants under one year of age, showing that this can be influenced by segmental frequency effects. (e.g., Anderson et al. 2003). It has been found that 19-month-olds have detailed subphonemic phonological representations that encode cues for place, manner, and voicing (White and Morgan 2008). However, mapping novel words onto objects appears to be challenging for 14-month-olds, indicating a heightened processing load that is only overcome around 20 months (Stager and Werker 1997; Swingley and Aslin 2000). There have also been several studies showing cross-linguistic differences in listening preferences for different types of lexical stress (see Nazzi et al. 2006 for review). However, there has been little investigation of infants’ preference for other types of phonological units (though see Jusczyk et al. 2002).

In summary, much has been learned over the past twenty years about the course of phonological development. However, the methods used all exhibit certain limitations. The longitudinal studies have typically lacked an accompanying acoustic record and tend to involve small case studies. Likewise, many cross-sectional production studies have typically not exploited information in the acoustic signal when assessing children’s phonological competence. They therefore miss potential covert contrasts the child may be making, presenting an incomplete and potentially
misleading picture of what children know about phonological structure. They also tend to focus on one age, with little attention to development. Finally, although a few phonological issues have been examined in infant perception studies, these typically use non-linguistic measures, such as listening times or listening preference. Many of the experimental studies also report only group data, making it difficult to assess individual differences in phonological development. Nonetheless, our understanding of how and when children begin to develop grammatical competence at different levels of phonological structure is quickly evolving through the use of more widely available laboratory phonology methods, promising new and exciting results in the coming years.

16.1.3 Contributions from laboratory phonology

As outlined above, one of the challenges to understanding the development of phonological knowledge is that children sometimes make acoustically measurable distinctions corresponding to constraints in adult speech but that are not perceived by the adult. This includes making subtle VOT distinctions for target voiced vs voiceless onset stops, both of which tend to sound voiceless to the adult ear (e.g., Macken and Barton 1980; Scobbie et al. 2000), and extrinsic vowel-duration distinctions before apparently missing voiced vs voiceless codas (Smith 1979; Weismer et al. 1981; Stoel-Gammon and Buder 1999). Young 1–2-year-olds have also been found to use spectral and durational cues to distinguish /gr/ from /gl/ in onset clusters (both heard as /gw/ by adults; Kornfeld 1971). Thus, children may acquire adult-like phonological contrasts earlier than often assumed, despite the fact that their early words often deviate from the adult form. Below we review further such evidence and discuss several possible factors that affect young children’s production.

16.1.3.1 The development of syllable and prosodic word structures

Researchers have noted that children’s early word shapes follow a systematic course of development. Drawing on data from English and Dutch (Fikkert 1994), Demuth (1995b) identified four stages in the development of words, suggesting that similar stages of development might be found in the acquisition of all languages. For example, Dutch-speaking children’s early words expand from core syllables (CV) (e.g., [fa] for olifant ‘elephant’) to minimal words (bimoraic feet, i.e., CVC, CVV(C), or CVCV in Dutch) (e.g., [faut] ‘elephant’), and eventually to larger, more complex phonological words (e.g., [olifant] ‘elephant’) as they progress in language acquisition.

While exploring four English-speaking children’s attempts to produce coda consonants in monosyllabic CVC words such as dog [dɔɡ], Demuth et al. (2006)
found that two of the children from 1;6 often lengthened the vowel when the coda is omitted, or added heavy aspiration or an epenthetic vowel to codas that were produced (e.g., CV ~ CVC ~ CVC^h ~ CVCV). Similar findings have been reported from corpus studies of other 1-2;6-year-olds (Vihman and Velleman 1989; Goad and Brannen 2003). This raises the question of the nature of children’s early syllabic representations, and whether these include coda consonants at all. Goad and Brannen (2003) proposed that heavy aspiration (typically appearing syllable-initially at this stage) occurring on the final consonant provides support that young children have only CV structure, and that apparent codas are actually onsets to an empty-headed syllable.

In contrast, Demuth et al. (2006) proposed that these children have a highly ranked NoCoda constraint, forcing output forms of CV, CVV, or CVCV. Given the high instance of vowel lengthening in the absence of a coda, they also suggested that English-speaking children may have an early awareness of word-minimality effects, where well-formed English lexical items must take the form of a bimoraic foot (Demuth 1995b; see Fikkert 1994 for similar explanations of early epenthesis in Dutch). Thus, children’s early use of vowel lengthening and the addition of an epenthetic vowel (e.g., dog /dɔg/ → [dɔː] ~ [dɔɡ]) could both be understood in terms of children’s attempts to meet word minimality. Under all these approaches the assumption was that children had early limits on syllabic (and prosodic word) representations, and that this began to change around the age of 2-2;6, as more target-like coda consonants were produced. However, these studies did not conduct acoustic analysis to further investigate these issues.

Some of the first studies to explore these issues acoustically came from an investigation of word productions from 1-2-year-old Japanese-speaking children (Ota 1999). Since Japanese is a mora-timed language, the issue of when these children become sensitive to moraic structure was of interest. Using durational measurements, Ota (1999) found that Japanese learners showed moraic compensation when they omitted the coda consonant, lengthening the vowel, in effect to constitute two moras of structure. Specifically, he showed that the short vowel that preceded a missing coda nasal (CVÔ) was significantly longer than a short vowel in an open syllable (CV) for all three children under investigation. Interestingly, such an asymmetry in vowel duration was not found when onset consonants were deleted, suggesting that the deletion of non-moraic segments does not lead to the compensatory lengthening of vowels. The findings suggest that Japanese children have an awareness of moraic representations or moraic weight of codas even when they cannot reliably produce the word-final consonants.

Similarly, Song and Demuth (2008) examined three English-speaking children’s compensatory lengthening of vowels in the context of missing codas (e.g., dog /dɔg/ → [dɔː]). Languages like English require well-formed content words to contain a bimoraic foot with either a coda consonant (e.g., tin [tın]), or a tense (long or bimoraic) vowel (e.g., tea [ti]) (Hammond 1999). Thus, if lengthening selectively
occurs with lax (monomoraic) vowels but not with tense (bimoraic) vowels, this would support the hypothesis that compensatory lengthening serves to preserve bimoraic or minimal word structure. However, if lengthening occurs across the board, this would indicate that increased vowel duration compensates for the omitted segment. The results showed that 1-2-year-olds lengthened both long and short vowels, suggesting that lengthening was compensating for the missing segment rather than the timing unit, i.e., mora (Stemberger 1992). This suggests that learning some of the language-specific constraints on prosodic word structure may take longer than previously assumed (Demuth 1995). However, it also provides support for the notion that these children have some representation for the missing coda.

16.1.3.2 Limitations on the articulatory control of onset and coda consonants

So far, we have provided evidence that children can exhibit adult-like representations of words even when their word production is not yet adult-like. This raises the question as to the nature of the factors that affect children’s early productions. It is possible that there is a speech-planning explanation for these findings. That is, children might have a coda in their phonological representation, but not yet having the articulatory gestures needed to execute CVC especially within a multi-word utterance. The findings reported above, where morphemes were more accurately produced utterance-finally compared to utterance-medially, provide some support for this position. Furthermore, vowel epenthesis appears most often following voiced codas, and aspiration noise appears most after voiceless codas (Demuth et al. 2006), suggesting that processes of speakers using acoustic cue enhancement might be involved (Keyser and Stevens 2006). That is, the child may be trying to ensure that cues to the voicing of the final consonant are clearly perceived although the cues might not be quite adult-like.

Weismer et al. (1981) found that children who apparently “omit” word-final stops nevertheless produce a stop allophone in word-medial position (e.g., do(g) vs doggy), indicating that /g/ must be part of the lexical representation of dog. This raises the possibility that some children’s early attempted codas may include coda closure, but lack the acoustic cues expected by an adult. We are currently conducting acoustic and ultrasound analyses to see if there is evidence for an incomplete closure gesture at early points in development. If such covert contrasts were found, it would suggest that the acquisition of coda contrasts is a gradient rather than a discrete process, with “quasi codas” produced en route to full coda articulation (cf. Hewlett and Waters 2004). In addition, although most typically developing English-speaking children reliably produce coda releases by the age of 2, there is still some variability in the acoustic realization of coda stops. We are currently pursuing investigation of these issues in the acoustic record of both children and adults to
better determine the development of acoustic cues to phonological contrasts, and the extent of individual variation (Demuth et al. 2009).

Further acoustic evidence of articulatory challenges faced by learners comes from Imbrie (2005), who compared ten children’s variable productions of the onset stops /b, d, g, p, t, k/ at 2;6–3;6 using durational, amplitude, spectral, formant, and harmonic measurements. When these acoustic measures were interpreted in terms of the supraglottal, laryngeal, and respiratory actions that give rise to them, comparison with adult productions of the same words showed that children have acquired appropriate positioning of their primary articulators for producing a stop consonant. However, the children’s gestures were still far from achieving the adult pattern even by the age of 3;6. For example, at this age children are still learning to adjust the tongue body during stop production, and the higher compliance of the articulators, smaller articulator size, and high subglottal pressure results in more tokens that have multiple release bursts and bursts that are shorter than those of the adult speakers. Longer VOT times and highly variable f0 suggest that children are still learning to adjust vocal fold stiffness and glottal spreading, as well as intraoral pressure. High variability in amplitude across an utterance suggests they are also still learning to control subglottal pressure. Thus, these children were less consistent than adults in controlling and coordinating certain aspects of their articulatory gestures, articulator stiffness, and respiration, though some aspects of the children’s speech did become more adult-like over the course of the year of the study (see McGowan and Nittrouer 1988 and Nittrouer et al. 1989 for similar findings for fricatives).

Using the same methods, Shattuck-Hufnagel et al. (forthcoming) examined children’s coda consonant productions, focusing on cues to voicing distinctions. The findings indicate that 2;6–3;6-year-olds exhibit systematic acoustic cues to coda-voicing contrasts (e.g., dog vs duck): an observable voice bar was more likely to precede voiced codas, whereas vowel glottalization was more likely to precede voiceless codas. Results from both 1;6–2;6-year-olds and their mothers’ child-directed speech show similarities; the voice bar appeared more frequently before voiced compared to voiceless codas (Demuth et al. 2009). For mothers, the duration of the voice bar was also longer for voiced codas, and children showed a trend in this direction. However, only mothers showed a significantly higher use of vowel glottalization before voiceless codas. Thus, although younger children produce some acoustic cues to coda-voicing distinctions, other cues take more time to become adult-like.

These findings raise questions regarding the relationship between early articulatory gestures and phonological representations. Regarding tongue gestures, Gick (2007) examined ultrasound recordings of an 11-month-old child imitating productions of /r, l, w/. In accord with results discussed so far, he found that the child’s production employed distinct articulatory traces and acoustic cues for each phoneme, despite the fact that the percept was not completely adult-like. On the
other hand, Ménard et al. (2006) found that French-speaking 4-year-olds’ CVC syllables were produced using different types of lip gestures than those of adults, and that children’s stressed and unstressed syllables were less differentiated than for adults.

Preliminary study of two Canadian French children (aged 1;11 and 2;3) explored these issues in children’s monosyllabic (CV, CVC) and disyllabic (CVCCVC) familiar words (Ménard and Demuth in preparation). The older child produced many word-medial codas, and had a distinct movement of the tongue for final VC as compared with final V. However, the younger child did not produce any codas in the disyllabic words he attempted. Furthermore, his vowels in the resultant CVCV productions were almost twice the duration of other vowels, showing compensatory lengthening. In addition, his tongue moved slightly toward the end of the vowel. This appears to be an articulatory gesture related to the attempted medial consonant, as confirmed by spectral analysis. These studies lay the groundwork for a more comprehensive investigation of young children’s articulatory gestures using ultrasound. They also suggest the importance of conducting close acoustic and gestural analysis of apparently coda-less CVC utterances, for evidence of non-adult-like cue patterns and how they change as children master adult-like pronunciations.

16.1.3.3 Context effects on the production and comprehension of grammatical morphemes

Some of the variable production and comprehension of both phonological units and grammatical morphemes may be influenced by the prosodic context and/or utterance position in which they occur. Children have long been known to exhibit within-speaker variability in the production of English inflectional morphemes (e.g., Brown 1973). Many researchers suggest that this is due to incomplete semantic or syntactic representations. However, our recent study of third-person singular -s found that children were much less likely to produce this morpheme when it is a part of phonologically more complex codas (hits vs sees), and in utterance-medial position as compared to utterance-final position (Song et al. 2009). This demonstrates that some of the within-speaker variability in the production of inflectional morphemes may be due to phonotactic complexity and positional effects. Hsieh et al. (1999) raise the possibility that this particular morpheme may be shorter in utterance-medial position. This could mean there is less time to produce it in utterance-medial position, resulting in more omission. Acoustic analysis of our stimuli used for both elicited production and comprehension experiments with 2-year-olds indicates that medial -s is indeed shorter than final -s (Song et al. 2009). This appears to have a negative effect on both production and comprehension of utterance-medial −s (Sundara et al. submitted). Interestingly, the effects of position...
are evidenced at the gestural level as well, in both older children (5-7-years-olds) and in adults (e.g., Nittouer et al. 2005).

Gerken (1996) provided elicited production evidence for 2-year-olds showing earlier production of articles that are prosodically licensed as part of a disyllabic trochaic foot ([hits the]Ft [piggy]Ft vs [catches]Ft the [piggy]Ft). We examined longitudinal data to determine if 1-2-year-olds’ use of articles would exhibit the same prosodic contextual effects in spontaneous speech. The results were confirmed for four of five children (Demuth and McCullough 2009). Interestingly, acoustic analysis of the productions from the fifth child showed a strong connection between prosodic organization and article production; her articles were produced as separate prosodic words at age 1;10, then became prosodified as part of a bimoraic foot (like the other children) at the age of 2. Little is known about young children's prosodic organization of grammatical morphemes, and how this develops over time. It is also unclear if children might go through a stage of development where they leave a “prosodic trace” for unrealized grammatical function items, such as that found in the omission of unfooted syllables for words like (Lu)cinda (Carter and Gerken 2004). Such a finding would provide additional evidence that children have some representation for the syllables and morphemes they omit. This is obviously a rich area for further research, using both longitudinal and cross-sectional methods.

16.1.4 Recent developments and future directions

16.1.4.1 New methods

New technological developments such as more accessible audio/video recording equipment and speech analysis software are beginning to address some of these limitations of previous longitudinal corpora. For example, the CHILDES database (MacWhinney 2000) now allows for both Unicode IPA transcription and the linking of audio/video files to the transcription record. This means that new databases, such as the Providence Corpus (English; Demuth et al. 2006) and the Lyon Corpus (French; Demuth and Tremblay 2008), are being donated with the audio files attached, allowing for a close examination of phonological and morphological development over time. This will permit much more extensive study of the acoustics of child and child-directed speech, and how this develops for the mothers and their children aged 1-3 years. The inclusion of the mother’s speech in these corpora is particularly important, serving as a baseline for understanding the nature of the input.

Ultrasound methods are only now starting to be used to explore the nature of children’s early phonological representations. With a small ultrasound probe placed under the chin, it is possible to collect both acoustic recordings of child speech and
video recordings of tongue movements in a non-invasive manner. This can provide some idea of the types of articulatory gestures being made, and the extent to which these may be incomplete. For example, some children exhibit protracted problems with the production of glides, producing only some of the required articulatory gestures (Bernhardt et al. 2005). This method therefore holds the potential for providing a better understanding about the articulatory underpinnings of phonological development, and possibilities for remediation.

16.1.4.2 Future directions

To adequately address the nature of language acquisition it is critical to know more about the input children hear. Some suggest that child-directed speech is a form of “clear speech”, with larger vowel space (Kuhl, Andruski, Chistovich, Kozhevnikova, Ryskina, Stoljarova, Sundberg, and Lacerda 1997) and less reduction of segments (e.g., want (h)im) than those typically found in adult-directed speech (e.g., Bernstein-Ratner 1982, 1987). However, studies of Dutch child-directed speech suggest more reduction of vowels in grammatical morphemes (van de Weijer 1998). It would be therefore be extremely helpful to know more about the acoustic/phonological properties of child-directed speech and the possible connections with individuals’ phonological development, as well as how both change over time. Computational techniques are currently being developed that could eventually approximate an automatic alignment of phonemes with the acoustic signal, making it possible to examine a large amount of child-directed speech (e.g., Sjölander 2003). This in turn could shed light on the nature of the acoustic input language learners actually hear, providing a better understanding of how and when children develop adult-like phonological representations, both perceptually, and in production.

In summary, we have examined evidence from laboratory phonology research showing that children under the age of 3 have more robust phonological representations of syllable structures and words than evidenced from impressionistic studies of production. Examining the shape of children’s early syllable and word productions is crucial to improving our understanding of the emergence of phonological representations. Several new data sources and methods are now making it possible to conduct laboratory phonology studies of phonological development in ways that were not possible before. This has brought with it an increasing number of studies from different languages, enriching our understanding of the acoustics of phonological development in a cross-linguistic context. The next decade promises to be an exciting one, with much more research on phonological development using laboratory phonology techniques. The results should provide a clearer picture of the course of phonological development at various levels of structure, and the implications this holds for later language development more generally.
16.2 Speech processing in bilingual and multilingual listeners: The relation between sound perception and word recognition

Paola Escudero

16.2.1 Introduction

In this contribution, I discuss the acquisition of speech processing skills in bilingual and multilingual populations. The focus on processing is first motivated by the fact that it has received relatively little attention within the domain of phonology. In addition, it seems reasonable to assume that multilinguals’ abilities to understand their languages should precede their abilities to produce them, and therefore the study of speech comprehension places us closer to multilinguals’ underlying language skills. Smiljanic (this chapter) suggests that the evidence so far shows that improvement in perceptual abilities does not seem to transfer uniformly to the development of production abilities. However, Escudero (2005) reinterprets the same studies and shows that it is likely that perception develops first and needs to be in place before production development can occur.

It is also important to investigate speech processing in speakers of more than one language, since it constitutes a highly complex process even within monolingual populations. This complexity has been emphasized in recent phonological (Escudero 2005; Boersma and Hamann 2009) and psycholinguistic (Cutler 2008) studies which agree on the fact that speech processing involves at least two separate processes, i.e. speech perception and word recognition, and two different representations, i.e. pre-lexical and lexical. Phoneticians commonly study speech perception, which involves a pre-lexical mapping of the raw acoustic signal onto the speech sounds (or phonemes) of a language, while psycholinguists commonly study the lexical mapping of sounds onto the words of a language, i.e. word recognition. Cutler (2008) gives ample evidence confirming the fact that listeners make use of pre-lexical and lexical representations and that these representations are accessed via separate processes. Although these two processes have mostly been studied separately (e.g., Storkel and Morissette 2002), Cutler states that phoneticians and psycholinguists have increasingly been interested in investigating the interrelations between the pre-lexical and lexical components of speech processing. Laboratory phonologists should also take such a comprehensive approach when empirically and theoretically accounting for the workings of speech processing.
What are the specific problems facing a second-language learner, a bilingual or a multilingual speaker when learning to perceive the sounds of a language and to recognize words containing such sounds? In the remainder of this contribution I will address three topics that shed light on this matter. The emphasis is placed on sequential and simultaneous bilinguals, who are more commonly referred to as second-language learners and bilinguals respectively, mainly because, up to date, there have been very few studies in the domain of speech processing which targeted speakers of more than two languages or multilinguals. In section 16.2.2, I review the factors affecting speech perception performance in bilinguals and multilinguals; specifically, bilinguals’ linguistic background and the influence of the experimental tasks with which they are presented. In section 16.2.3, I describe the evidence showing that sound perception and word recognition do not seem to go hand in hand in first-language acquisition and that the interrelation between these processes is also problematic for L2, bilingual and multilingual speakers. The final section deals with the influence of orthography on the perception of sounds and the recognition of words in second and third languages. Despite L2 researchers’ acknowledgement of the strong influence of written systems on speech processing, laboratory phonologists have only recently considered the systematic study of orthographic influences on bilingual and multilingual speech processing. The contribution ends with a summary and brief discussion of future directions for furthering understanding of this important area.

16.2.2 Factors affecting L2 bilingual and multilingual performance

Here we consider how the age of acquisition, language proficiency, the language in which testing is conducted, and cross-linguistic influences between the bilingual and multilingual’s languages affect speech perception performance.

16.2.2.1 Bilinguals’ age of acquisition and language proficiency

An important question related to bilingual and multilingual linguistic knowledge is whether children or adults who speak two languages have one or two linguistic systems, a question also addressed in Smiljanić (this chapter). A noteworthy variable when considering this question is the age at which the bilinguals’ languages are acquired. Specifically, it seems that a difference needs to be made between sequential bilinguals who acquired their second language after their first, either as

1 The term “sequential bilingual” refers to a speaker of two languages who acquired his or her second language after the first, either as a child or adult. This term contrasts with “simultaneous bilingual” which refers to speakers of two languages who acquire them simultaneously from birth.
children or adults, and simultaneous bilinguals who acquired two languages at the same time from birth. Behavioral studies conducted within the domains of speech perception and production show differential performance for these two types of bilinguals. On the one hand, speech perception studies with sequential bilinguals support one of the main hypotheses in Flege’s Speech Learning Model (Flege 1995, 2003) which states that bilinguals possess a common phonological space for their two languages. For instance, Caramazza et al. (1973), Williams (1979), and Flege and Eefting (1987) found that advanced adult L2 learners have perceptual category boundaries for VOT (Voice Onset Time) with a value that is intermediate between the values of the monolingual VOT boundary in the two languages. Similarly, Pallier et al. (1997) found that Spanish-Catalan bilinguals, who acquire Catalan in their childhood but are dominant in Spanish, did not perform like Catalan-dominant bilinguals because they could not accurately perceive the contrast /ɛ/-/e/ which is found in Catalan but not in Spanish.

On the other hand, Sundara and Polka (2008) found that simultaneous bilinguals, but not early L2 learners, seemed to accurately separate the production of their two languages. Specifically, Canadian English (CE)-Canadian French (CF) bilinguals could discriminate the voicing differences between /dV/ tokens produced by CE speakers and those produced by CF speakers. In contrast, monolingual CE, monolingual CF, and early sequential bilinguals of CF with CE as their first language could not discriminate between CE and CF productions of the same syllable above chance. Similarly, Burns, werker, and mcVie (2003) and Burns, Yoshida, Hill, and Werker (2007) found that the discrimination of the English VOT distinction between /b/ and /p/ in simultaneous bilingual English-French 10–12-month-old infants was similar to that of monolingual English infants of the same age. In addition, Sundara and Polka (2008) found that bilingual 10–12-month-old infants performed similarly to age-matched English monolingual infants in the discrimination of English /d/-/l/.

Thus, speech perception studies show that simultaneous bilinguals’ speech processing cannot be distinguished from that of monolinguals in their two languages, at least when they are very young (10–12 months of age) or as adults.² This, in turn, suggests that they may have separate systems for handling the processing of their two languages.³

² Sundara, Polka and Genesee (2006) found that, although adult simultaneous bilinguals and monolinguals performed similarly in the discrimination of the English /d/-/l/ contrast, 4-year-old bilinguals were poorer at discriminating the same contrast than age-matched monolinguals. In speech production, Sundara, Polka, and Baum (2006) demonstrated that adult bilinguals and monolinguals differed in their production of the stop contrast /d/-/t/. However, the authors also report that monolingual French and English listeners did not distinguish between bilingual and monolingual productions of /d/ and /t/ tokens.

³ Speech production studies show that bilingual children with different linguistic backgrounds produce language-specific differences in VOT for the consonant /t/ (Watson 1990; Khattab 2000; Johnson and Wilson 2002), but not for /d/. Sundara, Polka, and Baum (2006) showed that this
As pointed out by Idsardi and Poeppel (this volume), neurophysiological research can provide great insight into bilingual and L2 speech processing, including the question of whether they possess one or two systems for their languages and whether they perform like monolinguals. Many neurophysiological studies in the domain of bilingual speech perception have been conducted using the mismatch negativity (MMN) component of the event-related brain potential, which can be used to examine how the brain organizes phonological categories (Näätänen et al. 1978). This component is measured pre-attentively, as opposed to the attentive measure of all the behavioral studies reviewed above, through auditory exposure to the target sound distinctions while listeners read a book or watch a silent movie. The results so far suggest that sequential bilinguals, specifically Hungarian immigrants in Finland (Winkler et al. 2003), and simultaneous Swedish-Finnish bilinguals (Peltola et al. 2007) process vowels by means of an intertwined phonological system that handles both of their languages. It seems that the pre-attentive measure of bilingual speech processing reveals that even simultaneous bilinguals do not process their two languages independently. In the next section, it will be shown that these somehow contradictory results between behavioral and neurophysiological studies can be explained by the specific task presented to the bilingual.

Another important factor in bilingual speech processing is the level of proficiency that bilinguals have in each of their languages. Elman et al. (1977) found that, unlike the majority of their bilinguals who exhibit intermediate perception similar to that found in Caramazza et al. (1973), two “strong” or more proficient bilinguals who were sequential adult L2 learners had VOT perceptual boundaries that matched the monolingual perception of each of their languages. Similarly, Escudero and Boersma (2002) and Escudero (2005, 2009) showed that advanced, but not beginning or intermediate, learners of Spanish and Canadian French had vowel perception similar to monolingual listeners for Spanish /i/-/e/ or Canadian French /æ/-/ɛ/ respectively. Both these examples suggest that sequential bilinguals can perform like monolinguals in the perception of second-language consonants and vowels. In the next section, we will see that these seemingly contradictory results depend on the type of language setting in which the bilinguals perform the speech perception task.

16.2.2.2 The task presented to the bilingual

In bilingual infant perception, monolingual-like performance seems to depend on the sensitivity of the task to infants’ perceptual abilities. Using a task similar to the head-turn preference procedure (see Maye, this volume), Bosch and Sebastian-Galles (2003) showed that language-specific phonetic discrimination of the Catalan production problem is developmental because they found that adult bilinguals do produce a VOT difference between English and French /d/. McLeod and Stoel-Gammon (2005) also show that adult bilinguals produce VOT values within monolingual ranges.
/ε/-/e/ contrast is delayed in Spanish-Catalan bilingual infants because at 8 months of age only monolingual Catalan infants were able to discriminate this contrast, while both groups of infants could discriminate the same contrast at 10 months. However, Albareda et al. (forthcoming) show that with a more sensitive paradigm such as the Anticipatory Eye Movement paradigm (McMurray and Aslin 2004), 8-month-old Spanish-Catalan bilinguals are able to discriminate the Catalan contrast.

In line with the differences described in the previous section, psycholinguistic studies demonstrate that the amount of L1 or L2 activation during bilingual speech processing depends on factors such as language proficiency and dominance, and, especially, the language used during the task (Marian and Spivey 2003a, 2003b). Grosjean (2001) suggests that the bilingual’s languages can be activated selectively or in parallel as a function of the amount of use of their two languages during task instructions or in the stimuli presented. For instance, Kroll and Sunderman (2003) and Marian and Spivey (2003b) showed that bilinguals have differential lexical activation depending on which of their two languages is used during the testing session.

Escudero (2005, 2009) found that early and late sequential French-English bilinguals perceived the Canadian French /æ/-/ɛ/ contrast differently depending on whether they listened to it in a testing session solely conducted in their first language (Canadian English), or in another testing session conducted solely in their second language (Canadian French). Similarly, Escudero and Boersma (2002) and Boersma and Escudero (2008) showed that advanced learners of Spanish with Dutch as their first language performed similarly to Spanish monolinguals when classifying Spanish /i/ and /e/ within a solely Dutch or solely Spanish setting.

In the neurophysiological studies reviewed above (Winkler et al. 2003; Peltola et al. 2007), the role of the language setting or context was controlled for in order to investigate the extent to which the bilinguals or L2 learners’ phonological systems relate to one another. However, these pre-attentive results seem to contradict behavioral results which have been gathered controlling for the language used in a testing session, because in pre-attentive studies sequential and simultaneous bilinguals do not seem to have differential performance when perceiving sounds in their two languages. The answer to this controversy is given by Lehtola et al. (2007) who found that a group of simultaneous bilinguals similar to those tested by Peltola et al. (2007) appeared to be able to process the sounds of their two languages by means of two phonological systems when an attentive behavioral task was also included within the pre-attentive MMN testing session. It remains to be seen whether attentive and pre-attentive methods should be combined using the MMN technique, as has been done by Schafer et al. (2005) for monolingual children, in order to get closer to knowing whether bilinguals process their languages using the same or different phonological systems and whether their performance is comparable to that of monolinguals of either language.
16.2.2.3 Cross-linguistic influence and proficiency in multilingual performance

Very few studies have examined the acquisition of third or fourth (L3, L4, etc.) languages and the majority of the studies conducted so far have concentrated on speech production rather than speech processing. Within L2 learning, it is well known that the learner’s L1 prominently influences L2 performance, to the extent that native-like performance can be achieved depending on how the sounds systems of the two languages relate to one another. However, much less is known about whether L3 learning is influenced by the L1, L2, or both. Cenoz et al. (2001) review a number of studies on L3 acquisition, mainly in the domains of syntax and semantics, and suggest that the typological or linguistic distance between the learner’s three languages determines which of the previous two will influence L3 learning. Many studies have shown that there is a tendency to activate an L2 when learning to produce the sounds of an L3, because an L2 has been learned more recently than the L1, which may lead to its prominent use in L3 learning (Williams and Hammarberg 1998; Dewaele 1998; Wrembel 2007).

Another factor which has been suggested to play a role in L3 acquisition is the level of proficiency in the learners’ languages. Gonzalez Ardeo (2001) found an L3 speech production advantage for bilinguals with high proficiency or high exposure to their L2. However, in the domain of speech perception, Gallardo del Puerto (2007) found no effect of language proficiency on the English consonant and vowel perception of Spanish-Basque bilinguals. The author suggests that this finding is due to the fact that both Spanish and Basque have similar vowel and consonant systems and that therefore none of the two languages can be used to aid L3 acquisition of English. Simon et al. (2010) and Escudero, Broersma, and Simon (under review) show that Spanish (L1) learners of Dutch (L3) who are highly proficient in English (L2) are more accurate in perceiving Dutch vowels than Spanish learners who have only basic knowledge of English. Again it seems that the degree of similarity between L1 and L2 does affect the potential benefit to L3 learning.

16.2.3 The relation between speech perception and word recognition

Here, we consider whether the difficulties with speech processing in monolingual, bilingual, and multilingual language acquisition are found at the pre-lexical or at the lexical levels. First, the complexity of the relationship between speech perception and word recognition in monolingual first-language acquisition is addressed and it is demonstrated that children do not fully master the two processes involved in speech processing until later in life. Then, it is shown that simultaneous and
sequential bilinguals may have problems with one or both processes but that their performance can be close to that of monolinguals. Finally, the few available studies on L3 sound perception and word recognition are discussed.

16.2.3.1 Learning of minimal pairs in monolingual and bilingual children

Werker and colleagues have shown that infants younger than 17 months are unable to learn to associate two pictures to two different words if the words differ in a single consonant, i.e. if they constitute a minimal pair as in, e.g., /bin/ and /din/ (Werker, Cohen, Lloyd, Casasola, and Stager 1998; Werker et al. 2002). Importantly, the same studies show that these infants have no trouble distinguishing the minimal pair in a purely discrimination task where no word learning and recognition is involved. According to the speech-processing approaches within phonology and psycholinguistics reviewed in the introduction, these results may indicate that young infants have trouble accurately mapping pre-lexical representations which are minimally different to their lexical counterparts.

Unlike the consonant studies described above, infants younger than 17 months seem to be able to learn and recognize words that differ in some vowel contrasts (Curtin et al. 2009). Specifically, they could learn to associate the minimal pair /dit/ and /dıt/ to two different novel objects but not /dit/ and /dıt/ or /dıt/ and /dit/. Additionally, it seems that the difficulty in learning minimally different words or lexical neighbors continues later in life, as shown by Swingley and Aslin’s (2007) lexical competition study with 1;6-year-olds, Storkel’s (2009) word-learning study with 1;4–2;6-year olds, and Giezen et al.’s (under review) word-learning study with 5–6-year-olds. Importantly, Escudero and Benders (2010) review a number of recent studies showing that the type of testing paradigm used to examine infant and children's early word recognition seems to heavily influence their performance.

Most of the studies reviewed in 16.2.2 suggest that bilingual children have similar speech perception performance to age-matched monolinguals. The question is whether they have similar difficulties as monolinguals when learning words that constitute a minimal pair. Fennell et al. (2007) taught the novel words /bı/ and /dı/ to monolingual English, English-French, and English-Cantonese infants and found that the monolingual but not the bilingual infants could learn the minimally different words at 17 months. In contrast, Mattock et al. (2010) report that English-French bilinguals and not French or English monolinguals succeeded at learning the words /bos/ and /gos/ at 17 months. To explain these contradictory findings, Mattock et al. suggest that the type of contrast and its phonetic realization across languages may be an important factor in bilingual infants’ performance. In their study, the contrast in their study is phonemic in the bilinguals’ two languages, while the consonants of the words /bı/ and /dı/ in Fennell et al.’s study show considerable phonetic variation across the languages and the vowel /ı/ is only phonemic in English and not in French or Cantonese.
16.2.3.2 Recognition of minimal pairs in sequential bilinguals and multilinguals

It has been suggested that sequential bilingual’s difficulty in perceiving L2 sounds resides in the fact that they do not have distinct lexical representations for those sounds. Pallier et al. (2001) found that Spanish-Catalan bilinguals activated both lexical entries /dona/ (s/he gives) and /dôna/ (woman) when presented with either of the Catalan words, which may be due to the fact that Spanish has the vowel /o/ but not /ô/. Cutler and Otake (2004) found similar results when Dutch listeners were presented with words containing /æ/ and /ɛ/, probably because Dutch only has /ɛ/. The authors of both studies interpret their findings as evidence of a lexical problem, i.e. the bilinguals have a single lexical representation for both words containing two different sounds. However, the problem may reside in their pre-lexical processing rather than in their lexical representations and processing.

In an attempt to separate the role of word recognition and perception, Curtin et al. (1998) found that English listeners could discriminate the two Thai distinctions voiced vs voiceless-unaspirated and voiceless-unaspirated vs voiceless-aspirated depending on the nature of the task: if the task was perceptual they could discriminate the latter contrast better than the former, but if the task was lexical the opposite was true. The authors claim that this is because English has a voicing distinction, represented lexically by the feature [±voice], which distinguishes voiced and voiceless-unaspirated stops, but not an aspiration distinction. However, as was shown in 16.2.2, the task presented to the bilinguals matters: in a follow-up to Curtin et al., Pater (2003) found that when the word recognition and perception tasks were more equal, English listeners’ performance was comparable for both Thai contrasts.

Perhaps a more sensitive methodology to investigate the interrelation between speech perception and word recognition in bilinguals is that of Weber and Cutler (2004), who used eye-tracking technology to measure how listeners evaluate incoming auditory input over time. The authors used words whose first syllables minimally differ in the English contrast /æ/–/ɛ/, which was previously shown to be difficult for Dutch-English sequential bilinguals (Cutler and Otake 2004; Cutler and Otake 2004). The results showed that Dutch listeners looked longer and more frequently at a picture of, for instance, a pencil when the target word was *panda* than at a less confusable distractor (e.g., *beetle* when the target word was *bottle*), which may suggest that they have the same representation for words containing the two English vowels. However, when the Dutch listeners heard, for instance, *pencil* they did not look at the picture of the *panda*. Thus, the authors infer that these bilingual listeners have encoded the /æ/–/ɛ/ contrast lexically because they show an asymmetry in their inaccurate patterns of recognition of words containing these vowels. This means that these bilinguals may have no problem with differentiating the first syllables of words containing this contrast at a lexical level but that they perceive the two vowels as equal, i.e. they have different lexical representations but
a single perceptual representation for the two vowels. The question that emerges from these results is how learners can encode lexical contrasts that they cannot auditorily perceive. An answer to this question will be given in the next section.

As for the learning of minimally different words by multilinguals, Simon et al. (2010) and Escudero, Broersma, and Simon (under review) show that Spanish learners of Dutch, who have English as their second language, learn words containing Dutch vowel contrasts that do not exist in Spanish less accurately than words containing Dutch vowel contrasts that have a similar counterpart in Spanish. In addition, other studies have shown that these listeners have problems perceiving novel vowel contrasts (Escudero and Wanrooij forthcoming). As mentioned in the previous section, a high proficiency in L2 English facilitated the learning of Dutch words containing sound contrasts that exist in Dutch but not in Spanish, and that have similar counterparts in English. That is, the Dutch vowel pairs used by Simon et al. and Escudero, Broersma, and Simon had acoustic properties which matched the multilinguals’ L2 vowel system rather than their L1 vowel system. In this case, it seems that sequential multilinguals are able to beneficially transfer vowel categories from a previously learned language.

16.2.4 The role of orthography in bilingual and multilingual speech processing

It is well known that sequential adult bilinguals are influenced by the orthography of their first language (L1). Much of foreign speech accent seems to have its source in L1 spelling conventions. Apparently, L1 grapheme-phoneme correspondences are quite entrenched in our linguistic knowledge. Recently, Escudero and Wanrooij (forthcoming) found that orthography influenced sound perception by Spanish adult learners of Dutch with low and high proficiency in Dutch. That is, when orthography was available in the response categories, listeners were more accurate in the perception of the novel Dutch contrast /a/-/A/ than when they heard tokens of these vowels in a task where the target stimuli and the response categories were presented only auditorily. The authors show that this positive difference is due to the fact that the Dutch vowels differ not only in vowel quality but also in length, and that their spelling represents such a duration difference; i.e., aa and a. Spanish learners of Dutch seem to be able to exploit this vowel duration difference (cf. Escudero et al. 2009). In addition, the authors show that for other contrasts the influence of orthography may be negative, i.e. it leads to lower accuracy.

In a first attempt to empirically examine the role of orthography in bilingual word recognition, Escudero et al. (2008) used the same eye-tracking paradigm to test Weber and Cutler (2004) and Cutler et al.’s (2006) hypothesis that Dutch-English bilinguals acquired differential lexical representations through orthography. Escudero et al. taught two groups of native Dutch speakers who had a high
proficiency in English twenty English non-words, which followed the same pattern as the words used in Weber and Cutler, i.e. the first syllables contained the English /æ/-/e/ contrast. One of the groups learned the words only by listening to their auditory forms and looking at their pictures, while the other was also presented with their orthography. Only the group who learned the words with their orthography looked at the picture of a word containing /e/ and not at the picture of the word containing /æ/, which suggests that the availability of spelled forms results in the establishment of lexical contrasts that can be used in auditory word recognition. In addition, the results suggest that learners may not be able to encode a lexical contrast for auditorily confusable L2 words if they are learned only on the basis of their auditory forms.

The studies mentioned above show that adult sequential bilinguals seem to be able to transfer their L1 orthographic representations when learning a novel contrast with orthographic representations that match those of their first language. This suggests a tight link between pre-lexical and lexical auditory representations and orthographic representations. This type of auditory and visual connection has been previously demonstrated in the domain of visual word recognition (e.g. Van Orden 1987; Ota et al. 2009).

As for multilingual listeners, it would be interesting to investigate how the spelling forms of the learners’ L1 and L2 languages influence the learning of words in a third language. Escudero et al. (in preparation) set out to examine whether the orthographic effects in speech perception found by Escudero and Wanrooij extend to word learning by varying the availability of orthographic information in the same word-learning task as that used by Escudero, Boersma, and Simon (under review), which was reviewed in the previous section. Preliminary results show that Spanish learners of Dutch who have English as their second language are both positively and negatively influenced by orthographic information in similar ways as was found for speech perception, i.e. orthographic information leads to higher accuracy for some contrasts while it leads to lower accuracy for others.

16.2.5 Summary and final remarks: What do we know so far and what is still to come?

This contribution has illustrated the existing empirical evidence for a number of aspects of bilingual and multilingual pre-lexical and lexical speech processing. It was shown that simultaneous bilinguals, but not early or late sequential bilinguals, can perform like monolingual listeners when perceiving vowels and consonants. This was the case when they were tested within behavioral speech perception studies but not when using neurophysiological pre-attentive methods, such as electroencephalography. Further, early and late sequential bilinguals can perform like monolinguals if the study is conducted solely in the language in which
they are being tested, which avoids the activation of their other language and promotes monolingual-like performance. Additionally, neurophysiological studies which include an attentive task yield monolingual performance in simultaneous bilinguals. Studies conducted with speakers of more than two languages show that the typological closeness of the multilinguals’ languages determines cross-linguistic interactions between them during L3 acquisition.

The learning of minimally different words is a complex matter because it involves the mastery of both pre-lexical and lexical processing. In that respect, it seems that children master the learning and processing of minimal lexical pairs after their second year of life. Importantly, monolingual and bilingual early recognition of minimal pairs is influenced by the type of sound contrast involved, i.e. either vowels or consonants, and the sensitivity of the testing paradigm to reveal children’s perceptual abilities. As for adult sequential bilinguals and multilinguals, it seems that they could have problems with perceiving sound contrasts that only exist in their L2 or they could have problems encoding lexical differences between words containing those contrasts. More studies comparing different contrasts and using different methodologies for examining the performance of the same type of bilinguals or multilinguals would shed more light on this issue. In addition, individual differences seem to be the norm rather than the exception in L2, bilingual and multilingual populations; these ought to be brought to light and explained.

Since bilingual and multilingual speech processing can be positively or negatively influenced by orthography, future research should further examine orthographic influence, as it applies to both pre-lexical and lexical processing, and whether it needs to be modeled as part of the knowledge underlying speech processing (cf. Simon and Herreweghe forthcoming). Another important question for additional research is whether sources of non-orthographic visual information can result in the same speech-processing effects, positive or negative, as shown with orthography.

16.3 Second-language speech learning

Rajka Smiljanic

16.3.1 Introduction

During the process of mastering their first language, infants become uniquely attuned to the distributional patterns of the sounds in their ambient language and less attentive and less sensitive to the phonetic dimensions of sound contrasts
not found in the language input (Werker and Tees 1984a). As a consequence, adult second-language learners who already have a system of phonological contrasts in place as part of their linguistic knowledge encounter difficulties in acquiring and processing a non-native language. Second-language acquisition thus represents a quantitatively and qualitatively different process from first-language acquisition, involving often effortful retuning and realigning of the existing linguistic system to the sound structure of the non-native language. A broad goal of research on second-language learning is to understand the processes by which a language learner comes to perceive and produce speech sounds in a non-native language. Furthermore, this research seeks to understand how these processes change over time and what role exposure to a second language plays in the learning process. Second-language learning research, in general, draws on and informs cognitive, biological, developmental, social, linguistic, and educational perspectives.

The section begins by considering significant findings over the past few decades in the second-language speech perception and production domains. We focus on some important insights in the segmental domain and some newer results from investigations of connected speech and suprasegmental phenomena. Theoretical frameworks that have been proposed to account for a wide range of empirical findings are also discussed. As will become evident, substantial progress has been made in advancing our knowledge in all of these domains. However, these advances also raise a number of new questions. Throughout the review, we highlight some of these open questions. We end by noting some additional active and future areas of second-language speech-learning research.

16.3.2 Second-language speech perception

Early cross-language research amply demonstrated the profound effect that linguistic experience has on second-language (L2) learning. In the perceptual domain, which had been a focus of a lot of early cross-language speech studies, it has been repeatedly shown that adult listeners have difficulty perceptually discriminating phonetic contrasts that are not used distinctively or are phonetically realized differently in their first language (L1) (see Strange 1995 for review). Such perceptual advantage for native over non-native speech sound contrasts was taken to demonstrate that the adult speech perception mechanism is geared to process the native language in the most efficient way while at the same time contributing to adult learners’ difficulty in acquiring L2.

Although the basic premise of an L2 processing disadvantage remains valid, other work has since uncovered considerable variation in perceptual difficulties with non-native sound contrasts, showing that discrimination accuracy can vary from chance to native-like levels (Polka 1991, 1992; Best 1993, 1994; Best et al. 1988, 2001, 2003).
For instance, Farsi velar and uvular stops (Polka 1992) and Zulu voice and click place contrasts (Best et al. 1988) are easily distinguished by native English speakers, although these contrasts are not present in their L1 and these listeners had no prior experience with them. In contrast, English listeners’ discrimination of Hindi retroflex and dental stops and Nthlakampx velar and uvular ejectives is near chance (Werker et al. 1981; Werker and Tees 1984a). Some non-native vowel contrasts are difficult for L2 learners to distinguish even when they are similar to L1 contrasts in terms of their phonological features (e.g., Gottfried 1984). In contrast, Polka (1995) demonstrated native-like ease in discrimination of the German tense /u/-/y/ vowel contrast by English listeners without any previous experience with that contrast. Even though the listeners did not have two distinct L1 categories to map the non-native contrasts onto, they were sensitive to some phonetic aspects of the distinction between the members of the contrastive pair (they rated them as “good” and “poor” exemplars of a single native category). Findings such as those reported in Polka (1995) clearly demonstrated that non-native listeners’ discrimination abilities are not constrained exclusively by phonological distinctiveness in their L1. Non-native speech perception is affected by fine-grained phonetic similarities (that do not reflect phonological contrasts in L1) and dissimilarities between the two sound systems in contact. Similar sensitivity to fine-grained phonetic similarities and differences between L1 and L2 categories has been demonstrated for non-native consonant contrasts (Best et al. 1988), phonotactics (Flege and Wang 1989; Dupoux et al. 1999; Halle et al. 2003), coarticulatory patterns (Beddor et al. 2002; Bohn and Steinlen 2003; Levy and Strange 2008) and segmental context (Strange et al. 2001; Lively et al. 1993; Sheldon and Strange 1982; Schmidt 1996).

The notion that non-native listeners are sensitive not only to phonetic details that signal phonological contrastiveness, but also to fine-grained phonetic variation within categories, led to the exploration of underlying mechanisms in L2 phonetic learning. Research with native listeners has shown their sensitivity to contextual variation and within-category structure (e.g., Allen and Miller 2001; Volaitis and Miller 1992). Furthermore, learning L1 affects the weighting of perceptual cues such that perceptual sensitivity along an acoustic dimension is reduced near the distributional peaks of L1 category prototypes (Iverson et al. 2003; Kuhl 1991). Acquiring one’s L1 can also lead to greater perceptual sensitivity for some dimensions than for others (e.g., Francis and Nusbaum 2002; see also Holt, this volume). Such learned L1-appropriate weighting of acoustic cues is one source of perceptual difficulties in L2 learning. For instance, Japanese listeners’ problems in differentiating between the English /r/-/l/ contrast relate to their lack of sensitivity to changes in the third formant frequency (F3), a primary cue for the native English listener, and their focusing instead on duration and changes in F2 (Miyawaki et al. 1975; Iverson et al. 2003). This leads to miscategorization of L2 sounds and a difficulty in production and perception of the novel contrast.
Other possible underlying sources of difficulties for L2 learners have been identified through examining the discrimination of English tense vs lax vowels. While native English listeners seem to rely predominantly on spectral differences and only partially on the durational differences between members of the tense/lax pairs (Hillenbrand et al. 1995, 2000), non-native listeners of various language backgrounds, ranging from Spanish, Portuguese, Catalan, German, and Russian to Mandarin and Japanese, seem to weight duration more heavily (Flege et al. 1997; Rauber et al. 2005; Kondarova and Francis. 2008, forthcoming; Cebrian 2006; Escudero et al. 2009). Importantly, some of these languages do not use phonemic vowel length in their L1 to differentiate vowel contrasts, suggesting that exposure to English, allophonic use of duration in L1 and/or some universal preference for duration (cf. Bohn’s desensitization hypothesis, 1995) guide this perceptual bias. Combined, these results show that L2 listeners employ different cue-weighting patterns in L2 perception, compared with native listeners and sometimes with the patterns observed in their own L1. An important issue is how the experience-based and universal preferences interact in shaping different L1 and L2 cue-weighting strategies and accuracy in the processing of L2 contrasts.

The effect of linguistic experience and variability in levels of discriminability of non-native contrasts extend to phonotactics and prosody as well (Dupoux et al. 1997, 1999, 2001; Burnham and Mattock 2007; Aoyama and Guion 2007; Hallé et al. 2004; Francis et al. 2008). Looking at sound sequences, Dupoux and colleagues (1999, 2001) showed that in both identification and discrimination tasks, Japanese listeners hear, for instance [ebzo], with an illegal CC cluster, as [euzo], a phonotactically permissible CVC sequence in their L1. The frequency of such a “repair” epenthesis may differ for other L1s and for different sound sequences (Davidson 2007b). Illustrating a difficulty of non-native speech perception at the level of non-native stress, Dupoux et al. (1997, 2001) showed that French listeners performed more poorly when perceiving contrastive stress compared with Spanish listeners, presumably reflecting the property of French fixed stress rather than the variable stress used contrastively in Spanish. The authors also found that the adult learners’ difficulties varied depending on the task (ABX vs AX discrimination task) and level of speech processing. Finally, with regard to the perception of tonal contrasts by speakers of a tonal vs non-tonal languages, Halle et al. (2004) showed that French listeners’ judgments were largely psychophysically based. In contrast, Taiwanese Mandarin listeners’ judgments were based on a set of contrastive categories and were more categorical. The observed variation across types of prosodic contrasts and listener background languages further demonstrates that the perceived similarity between phonetic properties of the native and non-native sound structure plays a crucial role in adult learners’ L2 perception patterns. More work is needed focusing on perception of phonotactic and prosodic phenomena and how they interact with other levels of linguistic processing.
16.3.3 Second-language speech production

Besides learning and tuning their responses to the relevant acoustic cues for L2 phoneme discrimination and identification, L2 learners need to learn how to produce new contrastive L2 sounds, their specific phonetic targets, coarticulatory and co-occurrence patterns in syllables and in words, and novel prosodic patterns, including stress, intonation, and rhythm. All these aspects of L2 pronunciation present L2 learners with significant challenges, resulting in pervasive accented speech patterns. In an early exploration of consonant production by Spanish learners of English, Flege (1991) found that those bilingual speakers who learned English as adults produced voice onset time (VOT) of English /t/ with intermediate values, i.e., between those found in monolingual Spanish and monolingual English speakers. The results were taken to indicate that these late learners did not succeed in establishing a new L2 category but rather used a different phonetic implementation rule for a single (i.e., merged English and Spanish /t/) phonetic category. Similar production difficulties were found for a variety of non-native consonant and vowel contrasts and L1-L2 pairings (e.g., Flege et al. 1992, 1995; Aoyama et al. 2004; Bohn and Flege 1992, 1997; Ingram and Park 1997; Flege et al. 1999a; Tsukada et al. 2005).

Analogous to the perceptual difficulty with novel sound sequences, production of novel sound sequences presents another level of difficulty for adult L2 learners. For example, Hansen (2004) showed that Vietnamese speakers had trouble producing /s/, /f/, /v/, /l/, and /ʃ/ phonemes in coda positions of English words, presumably due to their distributional restrictions in Vietnamese (see also Davidson et al. 2004; Davidson 2006b for English speakers’ difficulty in producing novel consonant sequences in onsets not found in English). Interestingly, these studies revealed asymmetries in the degree of difficulty for these sequences, i.e., not all consonants in coda positions or consonant clusters were found equally difficult, despite the fact that all of them were absent from L1. These results suggest that language-specific generalizations derived over the classes of phonemes and phoneme sequences are crucial in accounting for the observed production patterns (see Davidson 2006b for discussion). Finally, difficulty in acquiring L2 rhythm and intonation, very salient aspects of foreign accent, has also been investigated, although this remains a largely understudied area of second-language learning (e.g., White and Mattys 2007; Jilka 2007). These studies demonstrated the need for better understanding of speech production beyond the level of individual segments, as potential sources of difficulty and foreign accent for L2 learners.

An important question that all second-language learning production and perception studies address concerns the nature of the phonological system(s) in L2 learners. Results from behavioral perception and production studies indicate that the two sound systems, L1 and L2, are interrelated. Most of the studies reviewed so far clearly show the effect of L1 on production and perception patterns in L2. An
interesting and less explored aspect of this interaction is the effect that L2 exhibits on L1 (MacKay et al. 2001; Guion 2003; Cebrian 2006). For instance, Flege and Hillenbrand (1987), Flege (1987), and Mack (1990) found that French learners of English produced their native language stops with longer VOT than is characteristic of French, although not with as long a VOT as that produced by native speakers of English. The effect of L2 on L1 appears to manifest itself in two ways: the L1 segment is “modified” in such a way as to make it more dissimilar from the new L2 category (Flege and Eefting 1987), or the L1 segment becomes more similar to the new L2 sound (Flege 1987). Some evidence suggests that the degree of interrelatedness of the two systems depends on the age of acquisition with simultaneous bilinguals more likely to develop two independent monolingual-like sound systems compared with early and late bilinguals (Guion 2003; Kang and Guion 2006, Escudero this chapter). Note, though, that not all studies found that L1 and L2 systems are independent in early or simultaneous bilinguals (Sundara, Prika, and Baum 2006).

Another longstanding issue in L2 speech learning as well as in speech science and experimental phonetics concerns the nature and the relationship between the perception and production systems (see McMurray and Farris-Trimble, this volume). A common assumption in both first- and second-language acquisition is that the development of adult (or native)-like phonetic perception precedes production abilities. An interesting counterexample to this assumption, as discussed by Goto (1971), Sheldon and Strange (1982), and Yamada et al. (1994), is that some Japanese learners were successful in producing distinct English /r/ and /l/ categories despite the fact that they could not reliably identify native tokens, i.e., their production abilities exceeded their perception abilities. Bradlow et al. (1997) tested the production-perception link by exploring whether success in perceptual training with a wide range of naturally produced stimuli varying across talkers and syllable positions, i.e., high variability training, led to an improvement in speech production by adult L2 learners. As expected, they found that training in /r/-/l/ identification resulted in perceptual learning of the novel contrast by Japanese learners. Importantly, they found that the knowledge of the new contrasts gained through perceptual training improved the learners’ productions of the same contrasts as judged by the native English listeners. Finally, the results showed a high degree of individual variation in the level of learning in the two domains; i.e., there was not a uniform amount of improvement that transferred from learning through perceptual training to the production patterns by all learners. The finding that improvement in production occurred through perceptual exposure only, rather than through explicit instruction, was interpreted to support a unified mental representation for production and perception mechanisms consistent with the motor theory of speech perception (Liberman et al. 1967; Liberman and Mattingly 1985, 1989) and direct-realist approach (Fowler 1986; Best 1995). While these theories, as well as second-language learning models (Best et al. 1988; Best 1994, 1995; Flege 1987, 1992, 1995), provide frameworks for considering the transfer of perceptual learning to production, they
still need to be refined to account for the lack of correlation between the degrees of learning in the two domains.

Finally, exploring the role of the environment and learner-related variables in determining how successful second-language learners are in achieving native-like levels in L2 processing has shed important light on non-native production and perception patterns. As described above, non-native speakers can have extreme difficulties with producing and perceiving certain non-native segmental and suprasegmental contrasts. Foreign accent can persist even for proficient speakers and even for early L2 learners (e.g., Flege and Hillenbrand 1987; Flege et al. 2006). Some adult learners, however, manage to achieve native-like levels in proficiency and pronunciation (Bongaerts 1999; Bongaerts et al. 2000; Birdsong 1992, 2007). These seemingly contradictory results underscore the importance of better understanding the role of variables, such as the age of acquisition, length of residence in L2-speaking country, relative amount of L1 and L2 use, quantity and quality of input from native L2-speakers, gender, motivation, social stigma associated with speaking with an accent, musical training etc. (e.g., Flege 1999; Flege et al. 1995, 2006, 1999b; MacKay et al. 2006; Piske et al. 2001; Bialystock and Hakuta 1999; Cebrian 2006; Gottfried 2007; Mayr and Escudero 2010; see also Escudero this chapter); and also numerous studies cited above).

16.3.4 Theoretical models

Several theoretical frameworks have been proposed to account for the underlying mechanisms that shape learners’ difficulties with novel sound contrasts. One common assumption shared by these models is that an acquired native sound system and the native speech experience serve as organizing principles that systematically relate to the adult learners’ non-native perception and production processes. One such model, the Native Language Magnet model (NLM), focuses on characterizing developmental changes in auditory perception from a universal to a language-specific perception during the first year of life (Kuhl 1991, 1992; Iverson and Kuhl 1996; Kuhl et al. 1992; Kuhl and Iverson 1995). These changes reflect a reorganization or “attunement” of infants’ phonetic perception to the contrasts that are linguistically functional in the ambient language. The ambient language input “warps” the underlying auditory-phonetic “space” in which phonological categories reflect the distributional properties of the native language system. One of the defining tenets of the model is that a prototype category acts as a magnet in that it “pulls” acoustically similar sounds towards it, simultaneously decreasing discriminability of tokens close to the prototype and increasing sensitivity to across-category differences. The listeners’ ability to differentiate the phonetic variation near the prototype is in that way diminished compared to discrimination around non-prototypes. Applied to second-language learning, the “pull” of the native prototypes on similar
non-native sounds results in a diminished discrimination of non-native categories. The inability to “carve up” the acoustic space along the dimensions relevant for L2 sound contrasts, along with the emphasis on acoustic cues transferred from L1, can lead to the formation of “wrong” category representations and longer processing times in second-language processing (Iverson et al. 2003).

The Speech Learning Model (SLM), which was developed specifically to account for second-language acquisition phenomena, proposes that perceptual similarity between native and non-native sound categories affects the degree to which L2 learners will be successful in producing and perceiving L2 sounds (Flege 1987, 1995). According to the model, “equivalence classification” determines the degree of similarity between L1 and L2 sounds, with similar L2 sounds being approximated more quickly at the beginning of the learning process due to their assimilation to L1 categories. However, more successful formation of new L2 categories and more accurate production and perception will arise with L2 sound categories less similar to the existing L1 categories, presumably due to less “interference” from the L1 categories. An important assumption of SLM is that the language acquisition processes remain intact over the lifespan, allowing L2 learners to apply the same processes to L2 acquisition rather than losing these abilities at some critical point during development (Lenneberg 1967).

Another model developed to account for non-native speech perception, the Perceptual Assimilation Model (PAM), provides an assessment of L2 learners’ perceptual difficulties in discriminating non-native sound contrasts within the existing native phonological system of contrasts (Best 1994, 1995; Best et al. 2001; Best and Tyler 2007). Patterns of assimilation of L2 contrasts to L1 categories present various scenarios by which non-native sound discrimination difficulties can be predicted. The direction and the degree of assimilation are determined by phonetic similarities between L1 and L2 sounds. For instance, two non-native sounds can be assimilated to two different L1 categories (two-category assimilation) or to a single L1 category (single-category assimilation), depending on the degree of the perceived similarity with L1 categories. Discrimination of non-native contrasts is expected to be very good in the former and poor in the latter case. Varied discrimination is expected for assimilation of two non-native sounds to a single L1 category in cases where one L2 sound is a “better” exemplar of the native category compared with the other member of the contrastive pair (category goodness difference). Finally, non-native sounds can be heard as speech but not be assimilable to any L1 categories (uncategorizable) or they can even be heard as non-speech sounds which fall outside the native phonetic space (non-assimilable). Discrimination in the last two cases can also vary widely depending on the saliency of the acoustic differences between the target L2 sound categories.

Despite many shared assumptions, these models differ in how they conceive of some aspects of the nature and mechanisms that underlie L2 speech-learning difficulties. For instance, SLM assesses adult learners’ difficulties in acquiring single
novel L2 sounds with an emphasis on production and on relatively experienced learners. On the other hand, PAM addresses perceptual difficulties of non-native contrasts, rather than single sound categories, through their assimilation to native categories. Furthermore, PAM was originally conceived to provide a framework for cross-language speech perception, i.e., perception of naïve non-native listeners, and only recently some details have been provided that account for L2 perception phenomena as well (Best and Tyler 2007). Unlike the other two models, PAM makes specific claims about the nature of the underlying representations based on articulatory phonology (Browman and Goldstein 1986, 1989). On this view non-native listeners assimilate non-native sounds to native sounds based on detection of similarities in the articulatory gestures. All of the models, regardless of their different foci, have contributed greatly to our understanding of the processes shaping adult second-language learners’ production and perception. Importantly, they have generated specific questions and predictions that resulted in substantial research as discussed above. A variety of questions remain for the models and the research to address: Can these models be extended to account for perception/production of sound sequences and prosodic phenomena? How do the models of L2 sound processing link up to other levels of linguistic processing (e.g., lexical recognition)? How does acquiring an L2 lexicon exert influence over sound perception and category formation? How can individual variation in L2 attainment be incorporated into the models?

16.3.5 Final remarks

This review brings together some of the most important and most recent second-language-learning empirical findings and theoretical developments. Additional promising research areas include studies exploring the plasticity of speech perception and production mechanisms through various training paradigms (e.g., Pisoni and Lively 1995; Strange and Dittmann 1984; McCandliss et al. 2002; McClelland et al. 2002; Lively et al. 1993; Yamada 1995; Bradlow et al. 1997, 1999; Iverson et al. 2005; Francis et al. 2008; Kondaurova and Francis forthcoming), the notion of the cross-language phonetic similarity of sounds evoked by the theoretical models (e.g., Strange 2007; Park and de Jong 2008; Bradlow et al. 2007, under review), effect of noise on non-native speech perception (e.g., Mayo et al. 1997; Van Wijngaarden et al. 2002; Van Engen and Bradlow 2007), the interaction of low-level acoustic information with information at higher-level structural and contextual information (e.g., Cutler et al. 2004; Bradlow and Alexander 2007), the effect of foreign-accented speech on intelligibility and speech processing (e.g., Munro and Derwing 4

4 Recently, Second-Language Linguistic Perception Model (L2LP) was developed to address the entire developmental L2 perception process (Escudero 2005). This model also allows for the assessment of individual variation in L2 learning tasks.
1995; Rogers et al. 2004, 2006; Bradlow 2008; Smiljanic and Bradlow 2007) and of multilingualism on speech processing (see Escudero, this chapter), to name just a few. Another area contributing significantly to our understanding of L2 processing and facilitating raising of new questions concerns physiological and brain-imaging studies (see Idsardi and Poeppel, this volume). Future research should be extended to second-language processing in more naturalistic communication settings; i.e., exploring the role of social interactions and audiovisual information in second-language acquisition, looking at spontaneously produced speech, exploring how perception and processing are affected by more realistic goals and demands of everyday communication situations. This will allow us to find out whether insights discussed here extend to situations outside of common laboratory conditions. Finally, it is important to think about how our research findings can be used to inform language pedagogy, second-language teaching, and language policy, which have practical implications for everyday functioning of the growing population of non-native speakers.