3

Speech Perception

Mark Antoniou

Introduction

The study of speech perception concerns how an individual’s perception of spoken language is influenced by the history of their interactions with their linguistic environment. In the case of bilinguals, this is doubly complicated as their perception of speech may be influenced by not one, but two (or more) languages. Despite recent advances, there is much that we do not yet understand about bilinguals’ perception of speech. To this day, most modern theories of language, particularly of spoken language, are implicitly monolingual in their assumptions. In the past, researchers studying language have ignored bilingual speakers or treated them as problematic. This is puzzling as it is estimated that half of the world’s population, if not more, is bilingual (Grosjean 2010); that is, bilinguals are just as representative of language users in general as are monolinguals. What is often unacknowledged is that bilinguals are not merely a special population worthy of study in their own right but they actually must be examined in order to provide definitive answers to a number of central questions in psycholinguistics regarding the effects of language experience on speech perception.

In this chapter, the challenge of speech perception facing bilingual listeners will be outlined and how the L1 and L2 may influence one another during perception. The factors that influence bilingual speech perception will be described. Brief overviews will be provided of the leading theoretical models of speech perception, which have been used to explain the perceptual abilities of bilinguals although they were not developed for this purpose, and therefore, we will also highlight their limitations in explaining the perception of fluent bilingual listeners. The limited literature on bilinguals’ perception of suprasegmentals, that is, phonetic variations that occur above the level of segments (consonants and vowels) will also be covered. Finally, we will end the chapter by proposing several future challenges that need to be addressed if we are to advance our understanding of bilinguals’ perception of speech.

Perception of Speech Segments by Bilinguals

Human speech involves a series of intricate, coordinated articulatory gestures that generate complex acoustic signals. People perceive speech segments (i.e., consonants and vowels), as is the case for other objects and events in the environment, as belonging to
categories. Research into bilingual speech perception seeks to use very sensitive perceptual tests and procedures in order to determine whether the speech categories of a bilingual’s two languages influence one another during the act of perception, and if so in which direction and to what degree. This is usually done by comparing the bilinguals’ performance to groups of monolinguals. Performance on a perception task that is equivalent to that of monolinguals is typically interpreted as providing evidence that the bilinguals have successfully kept the L1 and L2 phonological systems separate, or that the L1 and L2 categories coexist in a common space but are free of L1–L2 phonetic interaction. On the other hand, when bilinguals fall short of the performance of monolinguals, such as when they display performance values at a compromise point between those of monolinguals of the L1 and L2, this offers evidence that the L1 and L2 have merged and come to resemble one another in perception. Such comparisons with monolingual baselines are, of course, based on assumptions that should not simply be accepted without scrutiny (and indeed have been discussed in detail elsewhere; see Grosjean, 1998, as well as Chapter 1 in this volume). Nevertheless, comparisons with monolinguals are a common feature of research into bilingual speech perception, including in the studies that are summarized throughout this chapter.

The Flexibility of the Bilingual Phonological System

Bilinguals are remarkably flexible language users. In their daily lives, they are able to use and switch between their languages, seemingly effortlessly without hesitation and without hindering their communication with others. This is particularly impressive when one considers that bilinguals face greater challenges than monolinguals when perceiving speech. For example, a child being raised bilingually will necessarily receive less input in either language than a monolingual child. Nevertheless, bilingual children grow up alongside their monolingual peers, meeting developmental milestones along the way within approximately the same developmental timeframes (see Chapter 8). This is a testament to the resilience and flexibility of the human capacity to learn and use multiple languages.

Although much of the research that has been conducted on bilinguals within the field focuses on the constraints or limits of bilingualism, it does so by deliberately using very sensitive laboratory-based tasks in order to understand the psycholinguistic organization and interrelationships affecting the L1 and L2 systems. These studies tend to focus on contrasts that have been deliberately selected because they are difficult to perceive. As we progress through this chapter and highlight the points at which bilinguals deviate from comparison groups and what this means for their phonological organization, it is important to keep in mind that in reality bilinguals are able to perceive sounds quite well in their languages and communicate successfully in their daily lives.

Persistent L1 Effects

A large number of studies have demonstrated that there is something special about the language learning that occurs in very early childhood. Traces of the early acquired L1 may be observed many years later using speech perception tasks to assess the performance of even fluent bilingual listeners. This has been well documented by various research groups around the world.
A large contribution has been made to this literature by researchers at the University of Barcelona. In a series of studies, these researchers have explored the perceptual capacities of Spanish–Catalan bilinguals born in Barcelona, exposed to their L2 very early in life, and who have used their two languages on a daily basis (Bosch, Costa, and Sebastián-Gallés 2000; Pallier, Bosch, and Sebastián-Gallés 1997; Sebastián-Gallés, Echeverría, and Bosch 2005). A robust observation across these studies has been that Spanish-dominant bilinguals (those born within predominantly Spanish-speaking families) experience some difficulties in perceiving Catalan-specific contrasts. To illustrate, Pallier, Colomé, and Sebastián-Gallés (2001) studied the repetition priming effect in a lexical decision task involving Catalan minimal pair words. In this paradigm, participants perform lexical decisions on lists of stimuli (i.e., they must decide whether a presented stimulus is a word or non-word), some of which appear twice. Importantly, when performing this task, participants respond more rapidly when a word is encountered for a second time. For bilinguals, the crucial comparison rests on minimal pairs of Catalan words that differ in only one Catalan-specific contrast, such as /son/ versus /sɔn/. Indeed, Catalan-dominant bilinguals showed no repetition effects for the distinct Catalan words, whereas Spanish-dominant bilinguals did show priming effects indicating that they perceived the Catalan-specific minimal pairs as homophones (i.e., they heard them as repetitions of the same word). Even though these bilinguals were very fluent in both languages and used them daily, the effects of the earlier acquired L1 influenced the Spanish-dominant bilinguals’ perception of these L2 Catalan contrasts.

In a study involving bilinguals from the same population, Sebastián-Gallés and Soto-Faraco (1999) compared 16 Catalan-dominant and 16 Spanish-dominant bilinguals, all of whom had been born and lived in Barcelona. Each had been exposed to only their home language up until 3–4 years of age (Catalan exposure for the Catalan-dominant, Spanish for the Spanish-dominant). At 6 years of age, all had been exposed to both Catalan and Spanish when they attended kindergarten, and from that point onwards they received a bilingual education and became fluent and biliterate users of Catalan and Spanish. In this experiment, the bilinguals were asked to complete a gating task. In a gating task, a sound is divided into fragments (called “gates”), and successively longer fragments are presented to listeners until they are able to recognize the spoken word or segment. In this study, participants were offered two possible choices after each fragment was played, and importantly the differences between these two alternatives consisted of contrasts that exist in Catalan but not in Spanish: two vocalic contrasts (/e/–/ε/ and /o/–/ɔ/) and two consonantal contrasts (/ʃ/–/ʒ/ and /s/–/z/). The results showed that Spanish-dominant bilinguals needed longer sound fragments to identify the stimuli correctly, and thus performed worse than the Catalan-dominant bilinguals.

The two studies above are examples of the many that have been published by the Barcelona group that have demonstrated that Spanish-dominant bilinguals show subtle perceptual deficits when perceiving Catalan-specific contrasts. Based on such observations, Sebastián-Gallés and colleagues have put forth the compelling claim that the L1 exerts a persistent influence on perception of the L2, even after many years of continued daily L2 use from a very early age.

These claims have been corroborated by many other research laboratories around the world. To illustrate, Højen and Flege (2006) compared the English vowel discrimination abilities of Spanish–English early bilinguals to monolingual speakers of Spanish and English. Participants completed an AXB discrimination task for four English contrasts,
one of which was predicted to be easy to discriminate /i/–/u/ and three difficult /ɑ/–/ʌ/, /ɪ/–/eɪ/, /ʊ/–/oʊ/. Discrimination tasks assess whether a listener is able to hear the difference between a speech contrast. Numerous types of discrimination tasks are used by researchers in speech perception experiments (see Strange, 1995). In an AXB discrimination task, three stimuli are presented to the listener on each trial and they must decide whether the first or last stimulus token matches the middle one (i.e., whether A or B matches X). As would be expected, the English monolinguals showed the best discrimination performance, whereas the Spanish monolinguals discriminated the three difficult English contrasts at near chance levels. The bilinguals generally outperformed the Spanish monolinguals, but fell short of the English monolinguals for two out of the three difficult English contrasts. This suggests that the L1 continues to have a persistent effect on L2 speech perception, even though the bilinguals acquired it at a young age.

When Learning an L2 Alters the L1

Although it is well established that for many bilinguals the L1 commonly exerts a persistent influence on the L2, there is also considerable evidence that L2 learning alters the L1. Flege and colleagues have demonstrated such bidirectional L1–L2 effects (sometimes referred to as L2 reorganization) in numerous studies examining speech production (e.g., Flege 1987; Flege and Eefting 1987b; MacKay et al. 2001), and these effects have been observed in perception as well.

A clear demonstration of L2 influence on the perception of L1 segments is reported by Antoniou, Tyler, and Best (2012) who compared Greek–English bilinguals’ perception of Greek and English stop-voicing contrasts to that of Greek and English monolinguals. The Greek–English bilinguals were very accurate, relative to monolinguals, in identifying both L1 and L2 stop consonants in a variety of phonetic environments. The bilinguals did not confuse the phonetically similar short-lag stops of Greek (voiceless) and English (voiced initial stops). Participants were asked to do several things. First, they completed a categorization task combined with a goodness ratings task. Thus, participants were asked to make two judgments: they heard a stimulus token and selected a category label (e.g., they heard /pa/ and selected the label “p”). Then they heard the same stimulus token and assigned a native category goodness-of-fit rating by which they indicated how well the token (e.g., /pa/) matched their notion of that particular syllable (in this case their representation of /pa/). To do this they gave a rating ranging from 1 poor to 7 excellent. Then they were asked to complete a third task, an AXB discrimination task (see the preceding section for an explanation). The results showed the following: for categorization, the bilinguals seemed to attend to high level, abstract, phonological information, and identified the phonetically distinct English and Greek stops as phonological cognates (e.g., English [pʰ] and Greek [p] were both categorized as /p/). The bilinguals were also sensitive to the different L1 and L2 phonetic realizations of the same phoneme (e.g., English versus Greek realizations of /p/), as indicated by their category-goodness ratings. Despite these clear language differences in categorization and the assignment of goodness ratings, in the discrimination task, the bilinguals mirrored the performance of the English monolingual listeners across almost all positional contexts. The observed similarity between the bilinguals and English monolinguals in discrimination suggests that the L2-dominant early bilinguals have undergone perceptual reorganization and have come to resemble monolinguals of the L2 in their discrimination performance despite many years of continued L1 use.
Perception of Non-native Speech Segments by Bilinguals

The studies that we have discussed so far have examined bilinguals’ speech perception of their L1 and L2. Very few studies have examined bilinguals’ perception of foreign, unfamiliar segmental contrasts. Although much has been learned by presenting monolingual listeners with contrasts from foreign and unfamiliar languages, research on bilinguals’ sensitivity to non-native contrasts is only in its infancy. Studies examining bilinguals’ perception of non-native speech segments are concerned with whether bilinguals are sensitive to the phonological systems of both languages during non-native contrast discrimination (Antoniou et al. 2015). For instance, if a non-native contrast falls across a bilingual’s two phonological systems (i.e., one segment corresponds to an L1 category and the other to an L2 category), do bilinguals possess an advantage over monolinguals when discriminating that contrast? The evidence is mixed.

Melguy (2017) examined the discrimination abilities of eight highly proficient Spanish–English early bilinguals and ten American English monolinguals. None of the participants had any prior experience with Nepali and the bilinguals had been raised in the US, acquired English by the age of seven, and were L2-dominant. Participants performed an AXB discrimination task using Nepali stops [tʰa, tʰa, tɑ, da, da, kʰa, ga]. The results revealed that the bilinguals did not discriminate the Nepali contrasts better than the monolinguals. Since the bilinguals did not differ from the English monolinguals, they were judged to have undergone perceptual reorganization and come to resemble monolingual listeners of their L2 in perception of these non-native Nepali contrasts. This conclusion is consistent with Antoniou et al. (2012), who observed that L2-dominant bilinguals rely on the phonetic categories of their dominant language during discrimination.

Antoniou, Best, and Tyler (2013) compared Greek–English bilinguals with Greek and English monolinguals by examining their categorization, goodness ratings, and discrimination of non-native stop contrasts from Ma’di, a Nilo-Saharan language spoken in South Sudan and Uganda. Stimuli were naturally produced pseudo-words /ba, pa, ma, da, ta, da, da/ that were produced by phonetically trained native speakers of Ma’di. The bilinguals were divided into two groups such that half of the bilinguals completed the perceptual tasks in Greek, the others in English, and this had a clear effect on the bilinguals’ categorization (see the section on Language Mode below). For instance, Greek monolinguals and bilinguals completing the task in Greek both assimilated Ma’di /ba/–/pa/ and /da/–/ta/ to separate native categories, that is, they consistently categorized Ma’di /ba, pa, da, ta/ to Greek /b, p, t, d/, respectively. On the other hand, English monolinguals and bilinguals completing the task in English did not. Note that assimilation refers to the perceptual process via which non-native speech segments are categorized according to their perceived similarities to the segments in the native phonological system. In discrimination, Greek listeners discriminated Ma’di /ba/–/pa/ and /da/–/ta/ excellently, whereas the English monolinguals had difficulty. All groups discriminated /ba/–/mba/, /da/–/nda/, /ba/–/ba/, and /da/–/da/ poorly. Most importantly, the two groups of bilinguals showed the same discrimination patterns, regardless of the fact that they completed the task in different languages. Their mean discrimination lay midway between those of the two monolingual groups, suggesting that the bilinguals’ perception of the non-native Ma’di contrasts was a unique configuration resulting from their cumulative L1 and L2 experience, unlike that of either monolingual group.
McKelvie-Sebileau and Davis (2014) asked French–English and English–French bilinguals and English monolinguals to discriminate the Thai contrasts /ba/-/pa/ (phonemic in French, that is, heard as distinct phonemes, but not in English), /pa/-/pʰa/ (phonemic in English but not French), and /ba/-/pʰa/ (phonemic in both languages). The French–English bilinguals were born in France and the English–French bilinguals were born in either Australia or New Zealand, but all were living in Australia at the time of testing. The results showed that the French–English bilinguals discriminated both the native and non-native contrasts well. The English–French bilinguals were able to discriminate the native contrast more accurately than the non-native contrasts, but they performed similarly to the English monolinguals. A possible explanation for this pattern of results may be that the French–English bilinguals had benefited from the native English input that they had received while living in Australia as well as their likely greater use of their L2 relative to the English–French bilinguals. This conclusion leads nicely into the following section in which we will examine the differing factors that shape bilinguals’ speech perception.

**Factors Affecting Bilinguals’ Speech Perception**

Bilingual speech perception is often characterized as incomplete or imperfect. It is sometimes incorrectly assumed that in order to be considered “truly bilingual” an individual must have equal command of both languages akin to two monolinguals in the one person. However, there is now compelling evidence that even very fluent bilinguals will deviate from monolinguals somewhat because under many circumstances a bilingual’s two languages will be activated and will influence one another. The degree to which the languages influence one another is determined by a complex interaction of factors that have to do with a bilingual’s language use at both long and short timescales. These factors have received considerable attention over the past 30 years, particularly from Flege and colleagues. The subsections immediately below introduce the concepts of age of acquisition, L1:L2 usage ratio, language dominance, and language mode, and how they relate to bilingual speech perception.

**Age of Acquisition**

Age of acquisition refers to the age at which an individual is first exposed to their L2. This age of first exposure is generally taken as the starting point of L2 learning, and may vary from very early in infancy to late in life. For example, simultaneous bilinguals are exposed to both languages within the first year of life. In heritage speakers, the age of acquisition commonly corresponds to the age when they first attend preschool or school. In immigrant populations, “age of acquisition” is used alternatively with “age of arrival in the host country”, which may occur during childhood or adulthood depending on individual circumstances. In many studies, Flege and colleagues have taken advantage of such differences in age of acquisition to develop a detailed picture of how it shapes the perception (and production) of L2 speech. Speech perception findings have often been reported in the same articles as those about production, but our discussion here will be restricted to the findings concerning speech perception from those studies.
Flege (1993) examined Chinese participants’ production and perception of /t/ and /d/ in the final position of English words. Note that Chinese words are not differentiated by word-final stop consonants on the basis of voicing. Native English speakers lengthen vowels before /d/ but not /t/ in words such as bead versus beat. Consequently, a relatively long vowel is likely to bias a native English listener to perceive an ambiguous word-final stop as /d/. Thirty of the Chinese participants were late bilinguals, having first been regularly exposed to English following their arrival in the United States, whereas nine were early bilinguals who had arrived in the United States by the age of 10. Continua were used to probe the speech perception abilities of these participants. A continuum refers to a set of stimuli that are created by taking two stimulus tokens and generating a number of intermediate steps that vary along the dimension(s) of interest. In this particular study, the researchers created 17-step continua from naturally produced English words (e.g., bead and beat) by manipulating the vowel durations (/i/ in this case) with the aim of biasing native English listeners to perceive the words as ending in /d/ as in bead (where the vowels were lengthened) or ending in /t/ as in beat (where the vowels were shortened). Participants were presented with continua such as “beat–bead” or “bat–bad,” and were asked to choose the continuum step that was the best exemplar of beat or bead, or bat or bad. Both the early and late Chinese–English bilinguals selected stimulus steps with longer vowels as the best exemplars of English words ending in /d/ compared to those ending in /t/, suggesting that bilinguals showed some sensitivity to the role played by vowel length in determining final stop voicing. However, the late bilinguals exhibited a significantly smaller effect of vowel length than did the native English controls or the early bilinguals, suggesting that their perception of English is constrained by their later age of acquisition.

Flege, MacKay, and Meador (1999) examined the perception of English vowels by Italian–English bilinguals who varied in their age of acquisition (indexed by their age of arrival upon immigrating to Canada). In this study, 72 native Italian speakers were grouped according to their age of acquisition as either early (7 years old), mid (14 years old), or late (19 years old) bilinguals. Participants were asked to discriminate a range of vowel contrasts from the bilinguals’ L1 Italian (/u/-/o/, /e/-/a/, and /u/-/u/), L2 English (/i/-/i/, /u/-/u/, /æ/-/æ/, and /ʌ/-/ʌ/), as well as several between-language L1–L2 contrasts (/æ/-/a/, /ʌ/-/a/, /ʌ/-/o/, and /e/-/e/). The discrimination task presented three stimuli on each trial and required participants to indicate whether stimulus 1, 2, or 3 contained a different vowel or whether all three stimuli contained the same vowel. The English monolinguals discriminated the English–English and English–Italian contrasts more accurately than the groups of bilinguals. Importantly, a strong effect of age of acquisition was observed for the Italian–English bilinguals. Discrimination of English–English and English–Italian contrasts worsened as age of acquisition increased such that the early bilinguals showed the best discrimination, followed by the mid bilinguals, who in turn showed better discrimination than the late bilinguals. Therefore, it appears that as the age of acquisition increases, L2 vowels are perceived less accurately.

More recently, Stölten, Abrahamsson, and Hyltenstam (2014) investigated the effect of age of acquisition on the categorical perception of stop voicing. Three continua were created for the Swedish voiced versus voiceless contrasts /b/-/p/, /d/-/t/, and /k/-/g/. Forty-one Spanish–Swedish bilinguals were recruited, and all were proficient speakers of the L2 as judged by native speakers of Swedish, but they differed in terms of age of acquisition. There were 31 early Spanish–Swedish bilinguals (who had acquired Swedish
between 1 and 11 years of age) and 10 late Spanish–Swedish bilinguals (who had acquired Swedish between 13 and 19 years). The early and late bilingual groups were compared to 15 native Swedish speakers. Participants were required to categorize whether each of the continuum steps was voiced or voiceless. The native speakers showed a clear category boundary for all three contrasts. For the bilinguals, an effect of age of acquisition was observed on their L2 categorization responses. The late bilinguals’ categorizations deviated most from those of the native speakers, whereas the early bilinguals’ categorization responses were somewhere in between. These results suggest that neither group of bilinguals divided the continua in a native-like way, but the degree of this deviation from the native speakers’ categorization increased with age of acquisition.

**L1:L2 Usage Patterns**

Bilinguals use their L1 and L2 in different social settings, for different communicative purposes (see the Complementarity Principle in Chapter 1). The L1:L2 usage ratio refers to the respective amounts of an individual’s communication that take place in the L1 and L2. Quantifying the amount of L1 and L2 usage is extremely difficult, particularly for bilingual speakers who frequently switch languages. Researchers interested in the pattern of language use in bilingual populations typically rely on self-report. However, such measures should only be used as a general guide rather than an accurate estimate of L1 and L2 use (Grosjean 1998). High levels of continued L1 usage might reflect limited exposure to the L2. Conversely, limited L1 usage may suggest that the L2 is used in a wider variety of social settings. In general, higher usage of a language (be it the L1 or L2) is associated with greater proficiency and has subsequent consequences for speech perception. Long-term patterns of language use have been shown to influence bilinguals’ speech perception. To examine this, it is necessary to compare bilinguals who acquired their languages at the same ages but differ in their L1:L2 usage ratios.

In a series of four experiments, Flege and MacKay (2004) asked native Italian listeners to discriminate several English vowel contrasts. In Experiments 1 and 2, the authors demonstrated that native Italian listeners who had only lived in Canada for up to three months had difficulty differentiating the English vowel contrasts /ɒ/–/ʌ/, /ɛ/–/æ/, and /i/–/ɪ/, and often identified both vowels within a contrast as instances of a single Italian vowel category. In two follow-up experiments, Flege and MacKay probed the effects of age of acquisition and also the amount of L1 Italian use on English vowel discrimination.

Experiment 3 involved 18 Canadian English monolinguals and 72 Italian–English bilinguals who had lived in Canada for years (unlike the recent arrivals described above). The Italian–English bilinguals were all born in Italy and had either immigrated to Canada early (between 2 and 13 years) or late (between 15 and 26 years). They also varied considerably in the amount that they reported using Italian in their daily lives, ranging from low L1 use (between 1 and 15% of the time) to high L1 use (between 25 and 100%). Based on these criteria, the Italian–English bilinguals were divided into one of four groups: early low, early high, late low, and late high. The researchers examined the bilingual groups’ discrimination of English vowels, focusing on the /ɛ/–/æ/, /ɜ/–/ʌ/, and /i/–/ɪ/ contrasts. Before giving the results, we should first mention Experiment 4 in which a different task was used to study the same phenomenon.
In Experiment 4, 64 Italian–English bilinguals from Experiment 3 returned a year later and were supplemented with 16 new participants from the Italian-speaking community. The Italian–English bilinguals had arrived in Canada either early (3–13 years) or late (15–28 years). As in Experiment 3, the low-L1-use participants reported using Italian less frequently (2–15%) than the high-L1-use bilinguals (29–75%). In this experiment, instead of discriminating vowels, participants heard and viewed an English phrase where, in one of the words in the written version, the target vowel was replaced by an asterisk (e.g., d*fficult), and they were asked to indicate if the target vowel in each phrase they heard had been produced correctly or incorrectly. The phrases used contained the vowel contrast /i/–/ɪ/, which does not occur in Italian.

The pattern of findings from Experiments 3 and 4 revealed that (a) early bilinguals were more accurate in discriminating the English vowel contrasts than the late bilinguals and (b) bilinguals who reported low L1 Italian use tended to discriminate the English vowel contrasts better than those with high L1 use. The first finding is consistent with the age of acquisition effects that we have discussed earlier in this chapter. However, the most important contribution of this study is the observation that the early bilinguals who used Italian more frequently (i.e., the early high group), but not the early bilinguals who used Italian seldom (early low group), were found to differ from native speakers of English in perceiving English vowels. These results suggest that the L2 perceptual learning that occurs in early childhood is modulated by L1:L2 usage patterns, such that individuals who use the L1 often are more likely to exhibit L1 influence in their perception of L2 segments, even when the L2 is acquired early in life.

Language Dominance

The implications of these findings are that over the course of years, bilinguals’ L1:L2 usage patterns will affect their command of both the L1 and L2. The perfectly balanced or “truly” bilingual individual who possesses equal command of both languages probably does not exist. According to Silva-Corvalán and Treffers-Daller (2016), bilinguals most often have a dominant, or stronger, language in which they have attained an overall higher level of proficiency, which they use more frequently, and across a wider range of domains. This said, there is little agreement as to how language dominance should be assessed, and determining bilingual dominance can be very difficult (Flege, MacKay, and Piske 2002; Grosjean 1998). Bilinguals may be more comfortable using their non-dominant language in the workplace, although they may actually be less comfortable using that language in other social settings. Furthermore, the language dominance of bilinguals is not fixed, but can change over time. That is, as the bilingual’s environment changes, so will their needs for particular language skills. We do not fully understand how language dominance influences speech perception. The vast majority of research on bilingual speech perception suggests that even fluent bilinguals who have acquired both languages from a young age, and use both languages daily, have a perceptual processing advantage for the L1. However, these studies have tended to recruit bilinguals who are dominant in their L1. It has been suggested that those bilinguals who use their L2 so extensively that they become L2-dominant might be the most likely to suppress interference from the L1 on the L2 (Flege, MacKay, and Piske 2002), although the evidence concerning this prediction is mixed.
Amengual and Chamorro (2015) investigated the role of language dominance in the perception of Galician mid vowel contrasts in order to determine if the Galician vowel system becomes more Spanish-like following extended experience with Spanish. This study involved 54 Spanish–Galician early bilinguals who were born and lived in Vigo and Santiago, Spain, and were highly proficient and used both languages in their daily lives. Language dominance was assessed using the Bilingual Language Profile (Birdsong, Gertken, and Amengual 2012), which is a self-report questionnaire that calculates a score taking into account several factors including age of acquisition, L1:L2 usage, competence, and attitudes to each language. These scores were used to divide the bilinguals into Galician-dominant and Spanish-dominant groups. The 29 Spanish-dominant bilinguals had been exposed to Spanish earlier (and to Galician later) than those who were Galician-dominant, while the 25 Galician-dominant bilinguals used Galician more than Spanish and had more native-like Galician than those who were Spanish-dominant. Participants were asked to categorize the Galician front and back mid vowel contrasts /e/-/ɛ/ and /o/-/ɔ/ by completing a forced-choice minimal pair identification task (listeners chose one of the two elements proposed) and also completed an AX discrimination task. In an AX discrimination task, the listener is presented with two stimuli and is asked to judge if they are tokens of the same segment or different segments. The results from the categorization and AX discrimination tasks revealed that the Spanish-dominant bilinguals had difficulty discriminating the Galician mid vowel contrasts, whereas the Galician-dominant bilinguals did not and showed robust categorical perception. These findings suggest that Galician-dominant (but not Spanish-dominant) bilinguals have established two separate vowel categories for the front and back mid vowels, and thus that language dominance strongly predicts the perception abilities of bilinguals. Note, however, that in this study language dominance was confounded with earlier exposure to that language. At this point, it might be tempting to interpret language dominance to be the factor that determines which language is going to shape a bilingual’s perception of speech; however, this is not always the case.

Casillas (2015) compared the speech perception abilities of ten English monolinguals with ten Spanish–English late bilinguals who had acquired their L2 English as adults and were Spanish-dominant, and ten early Spanish–English bilinguals who had been exposed to their L1 Spanish as children, but no longer spoke it as adults and had become L2 English dominant. In this study participants were asked to categorize each of the 11 steps of a vowel continuum based on the southwestern American English tense-lax high front vowel contrast /i/-/ɪ/. The continuum steps varied along vowel duration and spectral properties. The English monolinguals and the English-dominant bilinguals were more categorical in their responses than the Spanish-dominant bilinguals, meaning that they were more consistent in assigning category labels, resulting in a sharper category boundary. Subtle differences were also observed between the English-dominant bilinguals and the English monolinguals. The English monolinguals had a sharper category boundary than both bilingual groups. In turn, the English-dominant bilinguals had a clearer category boundary than the Spanish-dominant bilinguals. The groups also differed in their reliance on acoustic cues to make perceptual judgments. Both the English monolinguals and English-dominant bilinguals placed greater weight on vowel spectrum properties than vowel duration, whereas the Spanish-dominant bilinguals relied more on vowel duration. However, the English-dominant bilinguals were not entirely monolingual-like because they relied on vowel duration more so than did the
English monolinguals. Put simply, even though the English-dominant bilinguals had acquired their L2 early in life and did not frequently use their L1, they nevertheless exhibited some perceptual sensitivity to vowel duration, which is commonly observed in native Spanish listeners, although the English-dominant bilinguals did not rely on this cue to the same extent as the Spanish-dominant bilinguals. These findings indicate that the L1 may exert a persistent effect on the perception of L2 segments, even in bilinguals who have become L2 dominant.

**Language Mode**

The factors affecting bilinguals’ speech perception that we have covered thus far describe patterns of language acquisition and usage that may vary over fairly long timescales. Effects arising from differences in age of acquisition, L1:L2 usage patterns, and language dominance encapsulate differences that are typically measured in years. However, bilinguals are very dynamic language users who may use their languages differently across given settings, such as their L1 in the home, their L2 in the workplace, and switching rapidly between the L1 and L2 when talking with bilinguals who speak the same languages. Theories of speech perception have not yet captured the dynamic complexity with which bilinguals use their languages and how such rapid changes may affect speech perception. A very useful contribution has been made in this regard by the Language Mode framework (Grosjean 2001, 2008), which postulates that, at any time, bilinguals are in language modes that correspond to points of language activation on a monolingual–bilingual continuum that are thought to influence all levels of language processing (see Chapters 1, 4, and 6 in this volume for discussions of the effects of language mode beyond speech perception). Bilinguals will be in monolingual mode when interacting with a monolingual speaker of one of their languages. When in monolingual mode, the other language is said to be deactivated, although never completely. Bilinguals will be in the bilingual mode when interacting with a monolingual speaker of one of their languages. Under these circumstances, both languages are activated and code-switching occurs, but one language is used for processing and is, therefore, more active than the other. That is, the main language of the communication (also called the “base language”) is thought to affect the other “guest language” (Grosjean 1988). Code-switching studies have demonstrated that this is indeed the case: the processing of a code-switched sentence is slightly inhibited (Altarriba et al. 1996), comprehension in the guest language is delayed (Li 1996), and passages containing code-switches are read more slowly than single-language passages (Kolers 1966; Macnamara and Kushnir 1971). Thus, language mode is thought to affect both speech perception and production.

A number of factors can position a bilingual speaker or listener at a particular point on the language mode continuum, such as the person being spoken to, the situation, the form and content of the message, the function of the communication, and specific formal situational factors (such as whether the participant has been made aware of the aims of the experiment). Even hinting at the bilingual abilities of a bilingual participant may put them into bilingual mode (Grosjean 1998). Movement along the continuum may occur at any time depending on the factors mentioned above, and this movement is thought to be unconscious, smooth, and effortless. This may be akin to a monolingual’s automatic, smooth change of speech style or register based on changes in context or interlocutor (Giles, Coupland, and Coupland 1991).
In terms of speech perception, bilinguals in differing language modes may select language-specific information from the speech signal and this selectivity in attention to phonetic details will vary depending on the communicative situation. Studies (including some that we have already reviewed, such as Amengual and Chamorro 2015; Antoniou et al. 2012, 2013) have examined the effects of language mode on both native and non-native speech contrasts under systematically varied experimental contexts, with the aim of altering the level of activation of each language, so as to observe the effects on speech perception by bilinguals. Such studies have demonstrated that bilinguals shift their category boundaries when completing perceptual tasks in which the base language has been changed, an important dimension of the language mode concept (Elman, Diehl, and Buchwald 1977; Flege and Eefting 1987a; Hazan and Boulakia 1993).

A more recent demonstration of the effect of language mode, primarily of its base language component, is provided by Garcia-Sierra et al. (2012) who investigated the relationship between speech perception and language mode using event-related potentials. These authors assessed speech perception abilities using the mismatch negativity (MMN), which is a preattentive response to an odd stimulus in a sequence of stimuli. In MMN experiments, participants hear many repetitions of a stimulus and occasionally a deviant stimulus is presented, and if participants are able to hear the difference between the standard and deviant stimuli then the MMN response is observed. In this study, 10 Spanish–English bilinguals were presented with carefully chosen stimulus steps from a /ga/–/ka/ voice onset time (VOT) continuum. Spanish and English differ in their phonetic settings for differentiating between stop consonants in terms of VOT. This is the length of time between the release of a stop consonant and the beginning of the vibration of the vocal folds. Thus, in English, the VOT for a /ka/ is long and for a /ga/ it is short. These authors very cleverly chose continuum steps that would be consistently categorized as /g/ or /k/ by Spanish and English listeners, but also an ambiguous continuum step that Spanish and English listeners would categorize differently (Spanish as /k/, English as /g/). They also manipulated the base language by asking bilinguals to read silently in either Spanish or English while they listened to the speech stimuli during the experiment. The results showed that the MMN changed as a function of the base language. When bilinguals were reading a magazine in Spanish, an MMN response was observed, indicating that they perceived the ambiguous sound as belonging to a different category. Conversely, when the bilinguals were reading in English, no MMN response was observed, indicating that they perceived the ambiguous sound as belonging to the same category. These results demonstrate that experimentally manipulating the language mode (in this instance by having participants read a magazine in one language or another) causes bilinguals to perceive the same physical stimulus as belonging to two distinct categories.

Theories of Cross-Language Speech Perception

Theories of speech perception make specific predictions regarding whether non-native speech contrasts will be perceived accurately, although they do not account for the bilingual flexibility outlined above. Here, we describe the leading models of speech perception and discuss their contributions, but also limitations, in explaining bilingual speech perception.
The Speech Learning Model

To account for the observation that speech production in the L2 is often accented, Flege developed the Speech Learning Model (SLM; Flege 1995, 2002). SLM explains the age-related limits on non-native accentedness in speech production and perception. During L1 acquisition, the language learner becomes attuned to the contrastive phonetic elements of the L1, which also guide proper production of those L1 segments. SLM assumes that the mechanisms that individuals rely on to learn new speech segments remain intact throughout life. However, because of ongoing changes throughout life and the way that the L1 and L2 phonetic systems interact, an L2 learner may never produce or perceive L2 segments in a completely native-like way. Specifically, as age of acquisition and proficiency in the L1 increase, fewer L2 sounds will be pronounced accurately. It is for this reason that late bilinguals exhibit a stronger L1 “accent” than early bilinguals both in their production and perception.

A central tenet of SLM is that the phonetic categories of the L1 and the L2 coexist in a shared phonetic space and influence one another. Production (and perception) difficulties both arise from L1–L2 phonetic similarity within this common space. According to this SLM principle, many L2 production errors actually result from inaccuracies in segmental perception.

SLM proposes two mechanisms to account for these L2 inaccuracies. First is the process of category assimilation, whereby L2 segments that are perceived as sufficiently similar to an L1 category will be perceptually linked to one another, and a merged L1/L2 phonetic category will form, which will differ from that of a monolingual speaker of either language. The single, merged L1/L2 phonetic category will be used to produce and perceive corresponding speech segments in the L1 and L2, and discrimination will be difficult. For example, many Italian learners of English typically have trouble differentiating English vowel length contrasts such as /i/ and /ɪ/, which do not exist in Italian (Flege et al. 1999), suggesting that they have formed just one L1/L2 phonetic category for this high front vowel. The second mechanism comes into play when an L2 segment is perceived as sufficiently dissimilar from the closest L1 category, in which case a new, unique L2 phonetic category will be established. However, this will not result in native-like perception or production because the newly established L2 phonetic category and the nearest L1 phonetic category will shift away from one another in the phonetic space, in a process called category dissimilation. According to SLM, native-like perception of both the L1 and the L2 thus seem to be mutually exclusive, to an extent.

SLM has been enormously influential in the field of cross-language speech perception (and perhaps even more so in speech production). The model accurately captures many of the effects that we have covered in this chapter, including persistent L1 influence on the L2, effects of L2 reorganization on the L1, age of acquisition, and L1:L2 usage patterns. The mechanisms of category assimilation and dissimilation both account for why L1 and L2 production and perception deviate from those of monolingual speakers. Admittedly, however, SLM has focused on accounting for the performance of relatively inexperienced L2 learners rather than on fluent, stable bilinguals. SLM cannot explain why discrepancies are sometimes observed between perception and production, and nor can it account for dynamic changes in bilingual perception, such as those occurring under different language modes.
The Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM; Best 1995) predicts the success with which monolingual adults will discriminate foreign unfamiliar contrasts with which they have little or no experience. A central tenet of PAM is that there are both phonetic (physical, gradient) and phonological (abstract, categorical) levels of speech perception. Therefore, lower level phonetic categories may be linked to higher level phonological categories. This explains how different phonetic segments such as allophones [p] and [pʰ] may both be perceived as exemplars of the phonological category /p/. PAM proposes that non-native speech segments are categorized according to their perceived similarities to, or differences from, the segments in the native phonological system, a process termed perceptual assimilation. That is, perceivers can assimilate non-native segments into native phonological categories, and sensitivity to gradient phonetic information allows them to perceive within-category differences (e.g., foreign-accented versus native-accented consonants and vowels). PAM describes numerous assimilation types that in turn lead to specific predictions concerning discrimination. These are covered in detail elsewhere and will only be briefly summarized here.

When a naive listener hears an unfamiliar language, non-native segments may be:

1) assimilated to a native phonological category as a good exemplar, an acceptable but not ideal exemplar, or notably deviant exemplar of that category; or may be
2) perceived as speech but not fitting within any existing native category; or
3) heard as non-speech, such as an isiZulu click being heard as a cork popping.

How about when a listener has to discriminate between two non-native segments and not just listen to individual segments? There are six possibilities (called assimilation types) although here we describe only the three most important ones (for a detailed description of all assimilation types, see Best 1995). In general, discrimination will be better when two non-native segments fall on opposite sides of a native category boundary. This is the case in Two-Category assimilation, in which the non-native segments are assimilated into different native phonological categories and discrimination will be excellent. When non-native segments are assimilated to the same native category, however, discrimination success will vary in terms of category goodness-of-fit. To illustrate, let us consider a Category-Goodness difference in which both non-native segments are assimilated to the same native phonological category, but one is perceived as a better exemplar of that category than the other, and discrimination will range from moderate to very good depending on the perceived category-goodness difference. Antoniou et al. (2012) observed such a Category-Goodness pattern for Greek monolingual listeners who perceived English /b/ as a good exemplar of the Greek /p/, and English /p/ as a moderate exemplar of the Greek /p/, resulting in good discrimination (74.1%). In contrast, a Single-Category assimilation occurs when both non-native segments are assimilated to the same native category as equally good or poor exemplars, and discrimination will be poor. Antoniou et al. (2013) found this pattern for English monolingual listeners who perceived both Ma’di /b/ and /ɓ/ as good exemplars of English /b/, resulting in poor discrimination (53.5%).

PAM has made an important contribution to our understanding of speech perception by making testable predictions concerning how the phonological categories of the perceiver’s native language determine how unfamiliar, non-native phonetic contrasts are perceived. If perceptual sensitivity to the L1 constrains non-native perception, then it
seems reasonable that it might also interfere with acquisition of an L2. Subsequently, PAM has been applied in studies that have examined L2 perception as well. However, PAM was not designed to explain bilinguals’ perception of non-native contrasts, that is, whether the L1 and L2 interact to enhance or inhibit discrimination of non-native contrasts. Recall the study of Flege et al. (1999), who presented Italian–English bilingual participants with contrasts from their L1 Italian, L2 English, as well as between-language L1–L2 contrasts. PAM does not mention L1 and L2 categories, and thus cannot specify whether L1 or L2 categories would be used to discriminate such contrasts. Nor does PAM make a priori predictions about which set of a bilingual’s categories will be activated under different language modes and the resulting shifts in perception (e.g., Antoniou et al. 2013).

**PAM-L2**

PAM-L2 (Best and Tyler 2007) is an extension of PAM that, like SLM, predicts how extended L2 experience will change a phonological system that has developed based on native language input. Similar to SLM, it assumes that perceptual learning is possible at all ages but will be influenced by the entire language-learning history of the individual.

According to PAM-L2, L2 segments are first assimilated into already existing L1 categories or dissimilated from existing categories and established as new categories firstly on a phonetic level. As the learner’s L2 vocabulary expands they attune to the higher-order organization of the L2 and segments begin to be distinguished more and more on the basis of meaningful categorical differences. Improvement in L2 segmental perception may reach an asymptote very early in L2 learning, while the L2 vocabulary is still fairly small, and a larger L2 vocabulary may curtail further phonetic learning. Though seemingly paradoxical, this is thought to occur because a larger and rapidly growing vocabulary forces the learner to settle on a particular (L1-influenced, in many cases) version of the L2 phonology in the perception of the L2 phones in words of their second language. Consequently, PAM-L2 predicts that L2 learners will learn to perceive L2 segments with varying degrees of success.

PAM-L2 extends PAM principles to the perception of L2 segments and development of L2 categories. According to PAM-L2, a common L1/L2 phonological category can be mapped to distinct phonetic categories for each language, and those language-specific phonetic categories may evolve without necessarily influencing one another. This type of theoretical framework is useful for conceptualizing how bilinguals are able to maintain L1–L2 phonetic differences but simultaneously recognize that they are both instances of the same phonological category. To illustrate, recall the study of Antoniou et al. (2012) in which Greek–English bilinguals categorized both English [pʰ] and Greek [p] as /p/, but assigned different category-goodness ratings to the L1 and L2 phonetic realizations of that same phoneme (e.g., English versus Greek realizations of /p/). PAM-L2 does not, however, offer any a priori predictions concerning why a bilingual is likely to perceive the same physical stimulus differently under different language modes (e.g., Garcia-Sierra et al. 2012).

**Limitations of these models**

The leading models of cross-language speech perception described above predict how naive listeners will perceive nonnative contrasts (Best 1995) and how people who are acquiring an L2 will perceive L1 and L2 segments (Best and Tyler 2007; Flege 1995).
However, these models were not designed to account for the performance of fluent bilingual speakers. The aforementioned models of cross-language speech perception do not offer explanations as to why bilinguals may show differences between speech perception and production, as is often observed in studies involving bilinguals. The models simply specify the existence of categories that are thought to influence both speech perception and production, but do not specify a mechanism that may be responsible for discrepancies that may occur based on the communicative tasks facing bilinguals. Although SLM and PAM-L2 are useful for predicting the success with which individuals who are acquiring their L2 will discern L2 contrasts, neither addresses the performance of fluent, stable bilingual perceivers. Fluent bilinguals are not actively acquiring either of their languages. They have already reached a level of language stability and are comfortable using both of their languages every day, often interchangeably. This, of course, does not imply that they do not have a dominant language, nor does it imply that bilinguals ever stop learning. The factors that we have covered within this chapter such as the L1:L2 usage ratio, language dominance, and language mode are effects arising from differences in language use and are evidence that bilinguals are constantly adapting to the communicative needs of their environment. However, this lifelong learning differs from L2 acquisition, and so if we are to capture the complexity and dynamics of bilingual language use then new models of speech perception will need to be developed in the future.

The Role of Suprasegmentals in Speech Perception

Suprasegmental variation, so named because it occurs above the segmental (consonant, vowel) level, refers to how structure such as rhythm, stress, and intonation is realized in the acoustic dimensions of pitch, amplitude, and duration. This realization is highly language specific and therefore it affects speech perception in language-specific ways. For example, language rhythm, among its other functions, assists with the separation of continuous speech signals into units of stored meaning (a process called word segmentation). This means that in languages such as Spanish, Catalan, and French, where the syllable is the basic metrical unit, syllabic segmentation will be useful for this process, whereas in stress-timed languages such as Dutch and English, the rhythmic distinction between strong and weak syllables will serve the same purpose (Cutler 2012). To date, suprasegmentals have been poorly integrated into models of speech perception (and especially in the case of theories specifically dealing with L2 speech processing and bilingualism), and indeed relatively few studies have examined cross-language perception of suprasegmentals at all.

Cutler et al. (1992) examined whether French–English bilinguals segment speech into words in the same way as French and English monolinguals. As just noted, English and French differ in the rhythmic patterns that support segmentation. The authors sought participants with approximately equal command of both languages and gave them tasks that reveal segmentation patterns in each language. For French, this was a syllable monitoring task, in which spoken words were presented and a specified syllable target had to be detected; French responses are typically faster if the syllable corresponds exactly to a syllable in the carrier word, while English responses are not sensitive to this target-word relationship. For English, there was a word detection task, which reveals
stress-based segmentation if words spread across a strong syllable boundary prove hard to spot. Cutler et al. observed that the bilinguals were typically unable to use both rhythmic segmentation procedures. Rather, some bilinguals showed the French-typical procedure and others the English-typical procedure. Performance did not correlate with the mother’s native language, the father’s native language, or the country where the testing took place (France or England). Eventually, participants were split into two groups based on their response to the question: “Suppose you developed a serious disease, and your life could only be saved by a brain operation which would unfortunately have the side effect of removing one of your languages. Which language would you choose to keep?” Interestingly, those participants who elected to “keep” French used the French-typical segmentation procedure when presented with French words, but those who elected to keep English did not. Conversely, those who elected to keep English used the English-typical segmentation procedure with the English words, but those who elected to keep French did not. This pattern of results led Cutler et al. to conclude that even though these bilinguals were fluent in both English and French, they could only employ a single rhythmic segmentation procedure (the procedure typical of French or that typical of English) which they then used. Cutler et al. conjectured that such procedures are acquired in infancy (where sensitivity to language rhythm is known to be strong) and perhaps just one procedure suffices to launch word segmentation ability.

In a series of studies, Broersma (2005, 2008, 2010) examined how Dutch–English bilinguals’ perception of duration influences their perception of word-final obstruents. For native English listeners, the duration of the preceding vowel is the primary cue to syllable-final voicing. Note that although both Dutch and English contain contrasts such as /f/–/v/ and /s/–/z/, Dutch does not distinguish between voiced and voiceless obstruents in word-final position because all Dutch syllable-final obstruents are voiceless. The first study in this series (Broersma 2005) showed that Dutch–English bilinguals categorized the English final voicing contrasts as accurately as English monolinguals. This indicates that the Dutch–English bilinguals were able to accurately perceive word-final obstruent voicing contrasts even though they do not occur in that position in Dutch. In Experiment 2, vowel duration was held constant and voiced–voiceless continua were created for each of the English obstruent contrasts. This experimental design meant that vowel duration was no longer an informative cue. The results showed that English monolinguals tried to use vowel duration when categorizing the obstruents, and this resulted in less sharp categorization functions. In contrast, the Dutch–English bilinguals were able to make use of the cues to voicing in the consonant itself, showing that unlike the native listeners they did not rely solely on the preceding vowel duration. This observation was robust across Broersma’s series of experiments: English monolinguals relied more on vowel duration than Dutch–English bilinguals, even when that cue became uninformative. These studies suggest that Dutch–English bilinguals are, in a sense, more flexible listeners than monolinguals because they are able to make use of alternative cues to guide the perception of speech under less than ideal listening conditions (such as when the primary cue becomes uninformative).

Cutler (2009) examined 24 Dutch–English bilinguals’ perception of L2 English words and their sensitivity to suprasegmental cues to stress. The bilinguals’ performance was compared to that of English monolinguals who completed the same procedure in a prior study (Fear, Cutler, and Butterfield 1995). A cross-splicing procedure was used to replace the vowels within words, for example, unstressed (but unreduced) vowels were
replaced by stressed versions of the same vowels. The Dutch–English bilinguals showed sensitivity to suprasegmental information and rated these words as more unacceptable than did monolingual listeners in the original Fear et al. study. These results suggest that Dutch–English bilinguals may be more sensitive to suprasegmental cues to stress than English monolinguals. A possible explanation for this might be that unstressed unreduced vowels are more common in Dutch than English. These results corroborate those of Broersma (2005, 2008, 2010) in suggesting that properties of the L1 may in some instances improve L2 listening and that bilinguals may outperform monolinguals on tasks requiring perceptual flexibility.

These studies offer a brief snapshot of the research being conducted on bilinguals’ perception of suprasegmentals. Fundamental questions remain unanswered regarding how suprasegmentals affect cross-language speech perception, and future research is needed to better understand how bilinguals process suprasegmentals and how this affects speech perception.

**Future Challenges and Directions**

The studies covered within this chapter raise a number of important questions to be answered by future research and theories of bilingual speech perception. Part of the challenge facing researchers is accounting for the enormous variability that is characteristic of bilingual speech perception, which may depend on patterns of language acquisition as well as long- and short-term patterns of language use. In the past, different types of bilinguals have been conflated into individual groups, or mixed into diverse and rather amorphous groups, and consequently results have often been misinterpreted. The field is moving away from the idea that bilinguals are part of a homogeneous group, but rather bilingualism is now being treated as a continuous variable because bilingual individuals differ on multiple dimensions (Luk and Bialystok 2013). The development of standardized instruments, such as the Bilingual Language Profile (Birdsong, Gertken, and Amengual 2012) and Language Experience and Proficiency Questionnaire (Marian, Blumenfeld, and Kaushanskaya 2007), offer a degree of rigor and comparability across studies that will benefit the field and lead to a fuller understanding of the factors affecting speech perception that we have covered.

In this chapter, we have only scratched the surface of how speech perception is affected by factors such as age of acquisition, L1:L2 usage patterns, language dominance, and so on. Developing a comprehensive and coherent theoretical explanation that accounts for all of these factors is a daunting challenge, but one that the field must ultimately face if we are to advance our understanding of bilingual speech perception. Extending existing theories of speech perception (e.g., SLM, PAM/PAM-L2) is likewise crucial to a deeper understanding of bilingual speech perception. These theories assume that language learning mechanisms remain intact throughout life. In conjunction with research demonstrating perceptual malleability even in monolinguals (Cutler 2012) and evidence of a strong effect of language usage and context on the speech perception of bilinguals, it seems necessary that any theory must explain the dynamic nature of bilingual speech perception, taking into account variations in performance stemming from differences in age of acquisition, L1:L2 usage patterns, language dominance, as well as short-term shifts in perceptual attention (language mode). The development of such a
framework for bilingual speech perception will likely require an adaptive and multilayered theoretical conceptualization that permits dual mapping of L1 and L2 segments to a higher level of phonological abstraction (Samuel and Larraza 2015), in order to simultaneously account for how bilinguals recognize L1 and L2 phonetic realizations of the same phoneme but remain sensitive to the phonetic differences of the L1 and L2.

Summary

In this chapter, we have presented an overview of speech perception in bilingual listeners. We have discussed studies that have examined bilinguals’ perception of segments in their L1 and L2, as well as those from foreign, unfamiliar languages. We have described the factors that have been shown to affect bilinguals’ perception, including age of acquisition, L1:L2 usage patterns, language dominance, and language mode. An overview has been given of the leading models of cross-language speech perception that have commonly been used to explain the perceptual patterns of bilinguals, but the limitations of these models have also been discussed in accounting for the dynamic ways in which bilinguals use their languages. An overview has also been provided of some findings from the literature on bilinguals' perception of suprasegmentals.

References


