In press in: M. van Oostendorp, C. Ewen, E. Hume & K. Rice (eds.) *Companion to Phonology*. Hoboken, N.J. : Wiley-Blackwell.

0 Speech Perception and Phonology*

ANDREW MARTIN

SHARON PEPERKAMP

1 Introduction

The task of speech perception involves converting a continuous, information-rich waveform into a more abstract representation. This mapping process is heavily language-dependent – every language divides up acoustic space differently, and the mapping is distorted by context-dependent phonological rules. This is not an easy job, but it is made easier in two complementary respects. During the first year of life the human perceptual apparatus is gradually optimized to better perceive the distinctions that are crucial in the ambient language's phonological system while ignoring irrelevant variation, and phonological systems themselves are optimized from the perspective of human perception.

This two-way interaction, in which perception adapts to phonology and phonology to perception, has long been of interest to phonologists, but only in recent decades have the tools necessary to explore the connection between the two become available. In this chapter we provide an overview of the debates that have arisen around each issue, as well as the research that bears on each debate (see also Hume & Johnson 2001). §2 discusses how phonology influences speech perception, both in the native language (§2.1) and in non-native ones (§2.2), and how second language perception relates to loanword adaptations (§2.3). §3 addresses the question of how perception influences phonology, beginning with an overview of the relevant typological data (§3.1), and concluding with a comparison of theoretical approaches to the data (§3.2). Finally, in §4 we briefly consider the ramifications of the bidirectional nature of the phonology–perception interaction.

2 Phonological influences in speech perception

2.1 Native language perception

How we perceive speech in our native language has long been a topic of interest. As early as the 1930s, phonologists were interested in the role of abstract phonemes in perception (Chapter 4: Phonemes). Sapir (1933) argued that phonemes are psychologically real, in that native speakers are typically unaware of the allophones in their language. The evidence he provides comes from native speakers transcribing phonemes regardless of their phonetic realization. For instance, native speakers of Nootka fail to transcribe the phoneme /i/ as [e] when it occurs after /h/. Likewise, Swadesh (1934) observed that phonemes are the basic percepts by which native speakers perceive their language.

More recently, the psychological reality of phonemes has been investigated experimentally. For instance, Otake *et al.* (1996) used a phoneme detection task to examine the processing of the moraic nasal consonant by native speakers of Japanese.[i] In Japanese, the phonetic realization of this consonant depends upon the place of articulation of the following consonant: in the words *kanpa* 'campaign', *bando* 'band', *kenri* 'right', and *tanka* 'type of poem', it is realized as [m], [n̪], [n] and [ŋ], respectively. Otake *et al.* found that Japanese listeners respond equally fast and accurately to the four realizations when asked to detect the sound that is spelled N in *rōmaji* (the Japanese writing system that uses the Latin alphabet). Especially revealing is the lack of a difference for the bilabial realization, since in onset position, the presence of [m] is contrastive (cf. *mori* 'forest' and *nori* 'type of seaweed'). Hence, Japanese listeners perceive all phonetic realizations of the moraic nasal consonant as instances of a single underlying unit.

Several other studies have shown perception differences for allophonic as opposed to phonemic contrasts (Chapter 6: Contrast) using discrimination paradigms. In these paradigms, participants are asked, for instance, to indicate whether a given stimulus is identical to another one or not, or to which of two different stimuli a third one is identical. It has thus been shown that listeners have difficulty distinguishing among allophones of the same phoneme, which are typically perceived as more similar than phones in different phonemic categories, even when acoustic distance is equated. For instance, Whalen *et al.* (1997) studied the perception of the allophonic contrast [p–pʰ] (with [p] as in *happy*), compared to that of the phonemic contrasts [b–p] and [b–pʰ]. They found that perception of the allophonic contrast was worse than that of the phonemic ones. Peperkamp *et al.* (2003) showed that, likewise, French listeners have difficulty perceiving the allophonic contrast [ʁ-χ] (both segments being realizations of the French phoneme /r/), compared to the phonemic contrast [m–n].

Of course, what is an allophonic contrast in one language can be a phonemic contrast in another language. Hence, another way of examining the effect of phonological status of a given contrast is to present it to one group of listeners for whom it is phonemic and to another group for whom it is allophonic. Such a cross-linguistic study was carried out by Kazanina *et al.* (2006). They used magneto-encephalographic (MEG) recordings to examine the perception of the contrast between [t] and [d] by native speakers of English and Korean. In English, this contrast is phonemic, whereas in Korean it is allophonic: [d] only occurs as an allophone of /t/ between two sonorant segments (Chapter 121: Laryngeal Contrast in Korean). Kazanina *et al.* revealed a mismatch negativity response (a neural marker for automatic change detection) in English but not in Korean listeners whenever a series of tokens of [ta] was interrupted by a token of [da]. In other words, Korean listeners did not detect the change from [ta] to [da], perceiving the different consonants as two instances of the same category. Similarly, using both a rating task and a discrimination task, Boomershine *et al.* (2008) found that English and Spanish speakers differ with respect to the perception of the contrasts [d–ð] and [d–ɾ]; the former is phonemic in English but allophonic in Spanish, while the reverse holds for the latter. Both groups of participants rated the contrast that is phonemic in their language as more different than the allophonic one and were better at discriminating it.

The fact that listeners have difficulty perceiving allophonic contrasts does not mean that allophones are ignored at all levels of speech processing. On the contrary, listeners appear to use their knowledge about the allophonic rules in their language for the purposes of word recognition. This was shown, for instance, by Lahiri & Marslen-Wilson (1991), who studied the processing of nasal vowels by English and Bengali listeners. In English, nasal vowels only occur as allophones of oral vowels before nasal consonants, whereas in Bengali the distinction between oral and nasal vowels is phonemic. Using a gating task, in which participants have to guess which word they hear as they listen to successively longer stretches of it, Lahiri & Marslen-Wilson found that adult native speakers of English, but not of Bengali, interpret nasality on vowels as an indication that a nasal consonant follows. That is, when listening to a word up to its nasal vowel, English speakers replied with words in which the vowel is followed by a nasal consonant. Likewise, Otake *et al.* (1996), in their study on the Japanese moraic nasal mentioned before, found that Japanese but not Dutch listeners use the phonetic identity of the moraic nasal consonant to anticipate the place of articulation of the next consonant. In particular, they had more difficulty in detecting a given consonant when it occurred after an inappropriate realization of the moraic nasal, i.e. a realization that does not share the following consonant's place of articulation; for instance, they found it hard to detect /k/ when it immediately followed a non-velar realization of the moraic nasal.

Finally, it is important to note that most of the evidence for the phoneme in perception reviewed above concerns unconscious perception processes. These processes should not be confused with (conscious) awareness. Adults who are illiterate or use a non-alphabetic writing system, as well as preliterate children, typically do not have phonemic awareness, that is, conscious knowledge about the phonemes of their language. They are thus unable to carry out metalinguistic

Code de champ modifié

tasks, such as deleting the first sound of a word (Liberman *et al.* 1974; Morais *et al.* 1979; Read *et al.* 1986), although they can perform such tasks with larger units such as syllables (Liberman *et al.* 1974; Morais *et al.* 1986). Importantly, the influence of the native phoneme inventory on speech perception is independent of literacy. Indeed, this influence arises very early in life, long before the question of literacy comes into play. For instance, it has been shown that, like English-speaking adults, 10- to 12-month-old English-learning infants fail to discriminate the non-phonemic contrast between [d] (as in *day*) and [d̥] (as in *stay*), whereas 6- to 8-month-olds do discriminate this contrast (Pegg & Werker 1997). Hence, during the first year of life, the capacity to discriminate non-phonemic contrasts already diminishes.

2.2 *Non-native language perception*

As shown above, our speech perception system is optimized for processing our native language. On the one hand, allophonic contrasts – which never induce differences in meaning – appear perceptually less salient than phonemic ones, even when acoustic distance is equated. On the other hand, the information provided by allophonic variation concerning neighboring segments is exploited for the purposes of word recognition. This optimization comes at a price, though, when we learn a second language later in life. Indeed, we experience difficulty perceiving non-native sounds and sound sequences, just as we find it hard to produce them.

That speech perception depends upon the phonological properties of the listener's native language was already observed in the 1930s (Polivanov 1931; Bloomfield 1933; Swadesh 1934; Trubetzkoy 1939). In the words of Trubetzkoy, the native phonology acts as a sieve during speech perception, in that non-native sounds and sound sequences are perceived as native ones, a phenomenon also called *perceptual assimilation*. Starting in the 1970s, experimental research has documented many cases of perceptual assimilation. Most of this research uses discrimination paradigms, the logic being that if a non-native sound is perceived as a native one, these two sounds are hard to distinguish; similarly, it is difficult to distinguish two non-native sounds if they are perceived as the same native one. An example of the former case is provided by German listeners, who have difficulty perceiving the contrast between the Polish phonemes /ʃ/ and /ɕ/, since they perceive non-native /ɕ/ as an instance of their native consonant /ʃ/ (Lipski & Mathiak 2007). An example of the latter is provided by what is probably the best-known instance of perceptual assimilation: Japanese listeners have difficulty perceiving the contrast between English /l/ and /ɹ/ (Goto 1971; Miyawaka *et al.* 1975). Indeed, Japanese has only one liquid consonant, /ɾ/, which acoustically falls in between [l] and [ɹ]; they thus perceive both English phonemes as instances of their native category /ɾ/ (see also Chapter 28: The Representation of Rhotics and Chapter 29: Lateral Consonants). According to the Perceptual Assimilation Model (PAM; Best *et al.* 1988; Best 1995), this English contrast is particularly difficult for Japanese listeners because [l] and [ɹ] are equally bad (or good) exemplars of Japanese /ɾ/. By contrast, the Polish contrast between /ʃ/ and /ɕ/ should cause less difficulty for German listeners, since it contrasts one good ([ʃ]) and one bad ([ɕ]) exemplar of the German category /ʃ/.

As already noted by Polivanov (1931), perceptual assimilation is not limited to segments, but also concerns suprasegmentals, syllable structure and phonotactics. As for suprasegmentals, much research has been devoted to the perception of tones by speakers of non-tonal languages (Chapter 10: The Representation of Tone). Kiriloff (1969) showed, for instance, that English listeners have difficulty perceiving certain lexical tones used in Mandarin. More recently, the perception of stress and length contrasts has started to be investigated. For instance, it was shown that French listeners have difficulty perceiving vowel length as well as word stress (Dupoux *et al.* 1997, 1999), neither of which is lexically contrastive in French. Concerning syllable structure (Chapter 30: Syllable-internal Structure), an example is again provided by Japanese, which has a very simple syllable structure, disallowing branching onsets and allowing only for codas consisting of a moraic nasal or the first half of a geminate. Japanese listeners find it hard to distinguish between consonant clusters and the same clusters broken up by the vowel /u/ (Dupoux *et al.* 1999). Hence, they perceive non-native clusters with an illusory epenthetic vowel (Chapter 41: Vowel Epenthesis). Finally, perceptual assimilation of non-native phonotactic structure has been studied, for instance, in French listeners; they have difficulty perceiving the contrast between the word-initial clusters /tl/ and /kl/ (Hallé *et al.* 1998; Hallé & Best 2007). In French, /tl/ is not a legal onset cluster (Chapter 56: Onsets), and French listeners tend to perceive it as /kl/.

Perceptual assimilation arises during very early processing stages and does not require the listener's conscious attention. This is evidenced by the fact that compared to native contrasts, non-native contrasts give rise to a reduced mismatch negativity (Näätänen *et al.* 1997); this electrophysiological response component is a neural marker of automatic change detection that arises 150 to 250 ms after the onset of the change when a deviant stimulus interrupts a sequence of repetitive stimuli.

One caveat is in order when interpreting the results of experiments purporting to demonstrate perceptual assimilation. Specifically, it is important to use experimental designs that tap a phonological processing level, where speech is coded in an abstract, language-specific format. Indeed, if listeners are given the opportunity to perform a task using low-level acoustic response strategies, no perceptual assimilation will be observed. For instance, Dupoux *et al.* (2001) used a very robust sequence recall task that yielded a complete separation between French and Spanish listeners as far as the perception of stress is concerned (that is, even the Spanish listener who had the most difficulty with perceiving stress performed better than the best French listener). However, they showed that in the absence of phonetic variability in the stimuli, French listeners performed equally well as Spanish listeners on the very same task. Differences between phonological and acoustic perception can also be shown by transforming the speech stimuli in such a way as to make them sound like non-speech. For instance, after removal of all acoustic information except the F3 component, Japanese listeners perceived the [l–ɹ] contrast as well as native English speakers (Miyawaki *et al.* 1975).

Research with infants has shown that many aspects of the native phonology, including segments, suprasegments, syllable structure and phonotactics, are acquired early in life (Chapter 80: Long-distance Assimilation of Consonants); one-year old infants thus show many of the perceptual assimilation effects that can be observed in their parents. The question as to whether perceptual assimilation can be overcome at a later age has been studied in two types of populations, i.e. bilinguals, and monolinguals who have received specific perception training. In both cases, improvement appears to be possible, with native-like performance generally remaining out of reach, though. To start with training studies, Japanese monolinguals can improve their perception of the [l–ɹ] contrast over several weeks of intensive computerized training, but their performance remains significantly below that of English monolinguals (Lively *et al.* 1993). Similar results have been reported for monolinguals who were trained to improve their perception of non-native suprasegmental contrasts (e.g. Wang *et al.* 1999).

Concerning bilinguals, we can broadly distinguish *late* bilinguals, who have started to learn a second language during or after adolescence, and *early* bilinguals, who have learned their second language during childhood. There is evidence that late bilinguals exhibit the same difficulty as monolinguals, although possibly to a lesser extent (Takagi & Mann 1995; Flege *et al.* 1997; Dupoux *et al.* 2008). For instance, Takagi & Mann (1995) found that Japanese-English bilinguals who had lived for at least twelve years in the US did not yield the same performance as native speakers of English on the [l–®] contrast; their performance was better, though, than that of Japanese monolinguals. Dupoux *et al.* (2008) found that advanced French-Spanish bilinguals were actually indistinct from French monolinguals as far as the perception of stress is concerned. However, these bilinguals had lived on average only four years in a Spanish-speaking country; the possibility that their perception of Spanish stress would improve with more exposure cannot be excluded. Perhaps more surprisingly, even early bilinguals sometimes fail to show native-like perception of their second language (Mack 1989; Pallier *et al.* 1997; Sebastián-Gallés & Soto-Faraco 1999). For instance, Pallier *et al.* (1997) tested fluent Spanish-Catalan adult bilinguals, who had started to learn Catalan before the age of six and who use Catalan on a daily basis. Spanish has only one mid front vowel, whereas Catalan has two, /e/ and /´/. The early bilinguals were found to have difficulty perceiving the Catalan vowel contrast.

Three factors have been proposed that influence the strength of perceptual assimilation in bilinguals. First, in the realm of segmental perception, the Speech Learning Model (SLM) of Flege (1995) states that the greater the perceived dissimilarity of an L2 sound from the closest L1 sound, the better the L2 sound is learned, both in perception and production. Second, L2 features that play a role –

even a non-contrastive one – in the native language, have been argued to be less difficult than others. For instance, McAllister *et al.* (2002) compared native speakers of English and Spanish living in Stockholm. They showed that the former perceive Swedish length better than the latter. Indeed, in English, length is used allophonically, making it easier for native speakers of English to learn the Swedish length feature. Finally, the amount of usage of the native language appears to be correlated negatively with performance in the second language. That is, the less bilingual speakers use their native language, the better they perceive the sounds of their second language, possibly becoming even indistinguishable from native speakers of the latter language (Flege & Mackay 2004).

2.3 *Perception and loanword adaptations*

One area of *phonological* research in which perceptual assimilation of non-native sounds and sound sequences appears to be relevant is that of loanword adaptations. In language contact situations, words of one language that are introduced into another language will typically (though not always) be adapted to the phonological pattern of the latter. Loanword adaptations thus transform non-native segments and suprasegments, as well as phonotactic structures, into native ones (see Chapter 100: Loanword Phonology). Given the fact that words must be perceived before they can be produced, it seems necessary that loanword phonology take into account the perception of non-native words by speakers of the borrowing language.

In phonological accounts of loanword adaptations, it is generally assumed that the input to the adaptations is constituted by the surface form of the source language, and that the adaptations are computed by the phonological grammar of the borrowing language. In processing terms, this means that during perception, the phonetic form of the source words is faithfully copied onto an abstract underlying form, and that adaptations are transformations produced by the standard phonological processes in production. Some researchers explicitly argue that foreign words can indeed be perceived without distortions and that speech perception is hence irrelevant to loanword adaptations (Paradis & LaCharité 1997; Jacobs & Gussenhoven 2000; LaCharité & Paradis 2005; see also Chapter 78: Structure Preservation). They are correct in as much as – as shown in §2.2 – at some level of representation non-native contrasts are indeed discriminated (after all, such contrasts present – sometimes quite large – acoustic differences). However, whether discrimination at an acoustic level of representation suffices to faithfully import non-native forms into one's lexicon is a different question. Given the phonetic variability due to inter- and intra-speaker variation, as well as the environmental noise in which speech is embedded, it seems reasonable to expect that listeners should discriminate non-native contrasts at a more abstract level of representation in order to store them faithfully. This is where the psycholinguistic literature becomes relevant, since it is exactly at this abstract level that listeners have been shown to experience difficulty perceiving non-native contrasts.

Several researchers have argued that at least *certain* loanword adaptations take place during perception, due to difficulties with the perception of non-native sound patterns. The first explicit proposal was made by Silverman (1992), based on the adaptations of English loanwords in Cantonese. Silverman proposed a two-stage model of loanword adaptations, containing a perceptual and an operative stage. During the first stage, the surface form of the source language is mapped onto an underlying form in the borrowing language on a segment-to-segment basis. This context-free mapping involves a first part of transformations, that are due to misperception of non-native segments. In particular, segments from the source language that are illegal in the borrowing language are perceived as legal ones (for instance, *bus* would be perceived as [pas], the Cantonese phoneme inventory containing /p/ and /a/, but not /b/ and /ʌ/). The perceptual level is also held responsible for the deletion of segments with a low acoustic saliency (for instance, *lift* would be perceived as [lif]). The output of the perceptual stage serves as the input to the operative stage, during which entire word forms rather than individual segments are evaluated. Adaptations take place whenever the phonological structure of the borrowing language is not respected. For instance, given that fricatives cannot occur in Cantonese syllable codas, the output [lif] of the perceptual level is transformed into [lip]. Other non-native coda consonants, as well as consonant clusters, undergo vowel epenthesis; for instance, the output [pas] of the perceptual level is transformed into [pasi].

Silverman's (1992) article, alongside with Yip's (1993) constraint-based reinterpretation of his model, has provided the impetus to an extensive literature on loanword adaptations, much of which is partly or wholly devoted to the issue of the role of perception (see Chapter 100: Loanword Phonology and references therein). Two questions in particular have been debated, concerning how much of loanword adaptations is due to perception, and how the role of perception is to be modeled. As to the first question, most researchers nowadays defend an intermediate position, arguing that perception does play a role in loanword adaptations, but that it cannot explain all effects. Oftentimes, arguments in favor of a perceptual account of adaptation patterns are impressionistic. For instance, word-final stop consonants are typically argued to lack acoustic saliency (especially if they are unreleased) and therefore to be prone to deletion during speech perception. Explicit reference to the psycholinguistic literature, however, has also been made. Specifically, comparing this literature to loanword data, Peperkamp & Dupoux (2003) observed a number of correspondences between adaptations patterns on the one hand and perceptual distortions of non-native phonological structure on the other hand. Some of these correspondences concern transformations applying at the level of individual segments. For instance, Japanese listeners' difficulty with the perception of the English consonants /ɹ/ and /l/ (Goto 1971) is reflected in loanwords from English, where these consonants are both adapted as /ɾ/ in Japanese (Lovins 1975). Others concern suprasegmental structure. For instance, French listeners' difficulty in perceiving stress contrasts (Dupoux *et al.* 1997) is reflected in loanwords in French, with stress being systematically word-final, regardless of the position of stress in the source word. Still others concern syllable structure. For instance, the illusory vowel that Japanese listeners perceive within consonant clusters (Dupoux *et al.* 1999) is reflected in the epenthetic vowel by which such clusters are broken up in loanwords in Japanese (Lovins 1975). This last example is especially revealing, since it contrasts with the intuition first expressed by Silverman (1993) that perception plays a role in cases of segment *deletion*, but not in those of segment *insertion*.

Speech perception experiments that specifically aim at comparing loanword data to the perception of non-native sound patterns have also been carried out (Takagi & Mann 1994; Kim & Curtis 2002; Peperkamp *et al.* 2008). For instance, Peperkamp *et al.* (2008) examined an asymmetry in French and English [n]-final loanwords in Japanese: the former but not the latter are adapted with a final epenthetic vowel (cf. [kannu] < Fr. *Cannes* [kan] and [peN] < Engl. *pen*). They showed that Japanese listeners perceive a vowel at the end of [n]-final non-words produced by French speakers but not when they are produced by English speakers. Based on these findings, they argued that the asymmetry in the loanword adaptation pattern originates in the perception of French and English words by Japanese listeners.

Concerning the question as to how the role of perception on loanword adaptations should be modeled, Silverman (1992) considered perceptually driven adaptations to be pre-grammatical, in the sense that they are computed before the phonological grammar *per se* comes into play. That is, they are influenced by the phonology of the native language, but not computed by the phonological grammar (see also Yip 1993; Peperkamp *et al.* 2008). Other researchers, however, have modeled perception-driven adaptations with grammatical tools, either as part of a perception grammar (Kenstowicz 2004; Boersma & Hamann 2009), or by incorporating constraints demanding perceptual similarity into the production grammar (Kang 2003; Adler 2006; Kenstowicz & Suchato 2006). The former appear to have the advantage that they can account straightforwardly for the fact that loanword adaptations sometimes conflict with native alternations.[ii] For instance, in Korean, obstruents turn into nasals before nasal consonants (e.g. /kukmul/ → [kuŋmul] 'soup'), but in loanwords, obstruent + nasal sequences undergo epenthesis (e.g. [pʰikʰɨnik] < *picnic*). If loanword adaptations are computed by the phonological production grammar, this grammar thus needs to be able to distinguish between native words and loanwords; by contrast, if loanword adaptations are computed by a perception grammar, conflicts between native alternations and loanword adaptations are a natural consequence of the fact that they are computed by distinct mechanisms, one in production and the other in perception.

3 Perceptual influences in phonology

3.1 The influence of perception on phonology: Evidence

Many phonological processes reduce the chances of a listener incorrectly perceiving certain sounds, particularly in contexts where they might be easily confused with other, highly similar sounds. Phonologies, in other words, appear to be optimized with respect to speech perception. In this section we survey a number of examples of this optimization; in §3.2 we discuss possible reasons for this typological bias.

In the previous section, we looked at evidence that speech perception is shaped by one's native language experience, making some distinctions easier to perceive than others. Despite the language-specific nature of perception, however, there are universal hierarchies of perceptibility imposed by the nature of the human auditory system. Some distinctions, for example those that involve greater acoustic distances, are simply more perceptible, *ceteris paribus*, than others. Although languages may differ in the phonemic contrasts they make use of, some contrasts are thus less marked than others. It is these universal differences in perceptibility that we refer to throughout this section.

As Flemming (2004) points out, perceptual markedness is best understood as a property of distinctions between sounds rather than of individual sounds themselves. Speech perception involves segmenting raw acoustic input and assigning each segment the appropriate category label. The probability that a given segment will be correctly categorized depends on what other categories it might be confused with, and where precisely the boundary between categories lies. It is thus meaningless to claim that a given sound A is difficult to perceive – it is rather the difference between two sounds A and B that is difficult to perceive. A *weak contrast* is a phonemic contrast involving two such hard-to-distinguish sounds.

The problem presented by a weak contrast can be solved in two ways: the weak contrast may be enhanced, making it easier to perceive, or it may be neutralized, making its correct perception unnecessary (Hayes & Steriade 2004). The remainder of this section discusses several examples of each type of process. We also examine the claim that perceptual similarity plays a role in processes that relate different surface forms to each other.

3.1.1 Contrast enhancement

Stevens *et al.* (1986) note that the distinctiveness of contrasts tends to be enhanced by the use of redundant features – high vowels, for example, are typically distinguished not only by their backness but also by rounding (Chapter 6: Contrast; Chapter 24: Phonological Inventories). This is one example of a more general tendency for languages to maximize the acoustic distance between vowels, an observation that has a long history in linguistic theory (e.g. Jakobson 1941; Wang 1971). When the vowels of a language are plotted in F1–F2 space, inventories like that in (1a) are typical, while inventories like the one in (1b) are unattested (although see Lass 1984 for a discussion of possible counterexamples).

(1) a. *Evenly spaced vowel inventory* b. *Unevenly spaced vowel inventory*

i		u			
e		o	e	\	
	a		æ	a	å

Liljencrants & Lindblom (1972) show that a simple quantitative model, which essentially maximizes the total distance[iii] among a given number of points in F1–F2 space, provides a reasonably good fit to attested vowel systems (see also Lindblom 1986; Vallée 1994; Schwartz *et al.* 1997). Actual vowel inventories thus appear to be optimized in the sense that they minimize the risk of misperception by dispersing vowels as widely as possible throughout the available perceptual space.[iv] Similar claims have also been made regarding consonant inventories (Padgett 2001; Padgett & Zygis 2003; Flemming 2004; Gallagher 2008).[v]

Dispersion theory, as it has come to be known, has also been used to analyze historical sound change (Chapter 98: Sound Change). Padgett & Zygis (2003) argue that a change in Polish in which palatalized palato-alveolar sibilants became retroflex sibilants (e.g. [ʃʲija] > [ʂɨja] 'neck') was motivated by the existence of an alveolo-palatal series of sibilants in Polish. Using evidence from perceptual studies, they show that alveolo-palatals are more difficult to distinguish from palato-alveolars than from retroflexes for both native Polish and English speakers. The sound change in Polish was thus perceptually optimizing in the sense that it improved the overall discriminability of sibilants.

Another case of perceptually driven contrast enhancement can be seen in the Sonority Sequencing Constraint, which requires that segments in a syllable's onset and nucleus should occur in order of increasing sonority (Sievers 1881; Jesperson 1904; Saussure 1916; Hooper 1976; Kiparsky 1981; Steriade 1982; Clements 1990; Chapter 50: Sonority). Wright (2004) argues that this cross-linguistically robust generalization can be understood as a way of maximizing the perceptibility of place, manner, and voicing cues in vowels and consonants. A similar analysis is presented by Ohala & Kawasaki-Fukumori (1997), who suggest that perceptibility can also explain the low frequency of the sequences /ji/ and /wu/ in many languages, as well as the fact that the CV syllable is a more common syllable type than VC, because the cues to the place of the consonant are more robust in a transition into a following vowel than in a transition out of a preceding vowel (Fujimura *et al.* 1978).[vi]

A perception-based account of the Sonority Sequencing Constraint explains why sibilants, whose cues do not depend heavily on transitions into vowels, often violate the constraint. Several languages employ metathesis (Chapter 31: Metathesis), whether synchronically or diachronically, in stop–sibilant clusters, converting a stop–sibilant–vowel (TSV) sequence into a sibilant–stop–vowel (STV) sequence, as in the examples from Udi in (2).

(2) *Udi metathesis* (Schultze 2002, cited in Hume 2004)

/tad-esun/	→	[tast'un]	'to give'
/t'it'-esun/	→	[t'ist'un]	'to run'
/etʃ-esun/	→	[eʃtʃ'un]	'to bring'

Metathesis of this kind, which places the stop in a position where its place cues are the strongest, is attested in Udi, Faroese and Lithuanian (Hume 2004), while the opposite pattern (i.e. one that converts STV to TSV) is unattested (but see Blevins 2009 for a critique of this analysis).

3.1.2 Contrast neutralization

Perception can also shape phonologies in ways that involve the elimination of contrasts rather than their preservation (Chapter 84: Merger and Neutralization). One such example is the common process of nasal place assimilation, whereby nasal consonants adopt the place of articulation of the following consonant (Chapter 85: Local Assimilation). The ubiquity of this process may be related to the difficulty listeners have in determining a nasal's place of articulation when it occurs before a consonant (Fujimura *et al.* 1978; Hura *et al.* 1992; Ohala 1990; Beddor & Evans-Romaine 1992, 1995). Listeners who know that their language employs a rule of place assimilation do not need to attend to the place cues of the nasal, thus minimizing the risk of misidentification.

Another example comes from Ohala (1990), who argues that in phonological processes, or cases of historical change, in which heterorganic stop clusters become geminate stops, it is the features of the second stop in the cluster that are typically adopted by the geminate (e.g. Sanskrit *bhaktum* > Pali *bhattum*, not **bhakkum*). He explains this asymmetry as the result of the fact that cues to stop place are more salient in the transition from the stop to a following vowel than they are from a preceding vowel to the stop. The place of the second stop in the cluster is therefore more perceptible, and it is not surprising that when geminates are formed from such clusters, this place is chosen over that of the initial stop. The place contrast in stops is thus eliminated precisely where it is difficult to perceive (Chapter 70: Positional Effects in Consonant Clusters).

Steriade (1999) shows that voicing contrasts in obstruents tend to be permitted intervocalically, where cues to voicing are strongest, and neutralized preconsonantally, where cues are weak (Chapter 47: Final Devoicing and Final Laryngeal Neutralization). Cues to apical contrasts, such as that between [d] and [ɖ], on the other hand, follow a different pattern, being strongest after vowels and weakest after consonants. This correlates with the typology of processes that affect apicality – languages tend to eliminate such contrasts postconsonantally and preserve them postvocalically. Place assimilation, for instance in apical consonant clusters, is overwhelmingly progressive ([ɖd] → [ɖɖ]), while major place assimilation is typically regressive ([np] → [mp]) (Steriade 2001).

3.1.3 Similarity maximization

A third category of perceptual effects in phonology, in addition to contrast enhancement and neutralization, could be called *similarity maximization*. Steriade (2008) observes that phonological processes tend to make changes to underlying forms that are perceptually minimal. For example, although many languages require adjacent obstruents to agree in voicing, every such language repairs disagreeing obstruent clusters through assimilation, by either voicing or devoicing one of the segments in question. No language deletes one obstruent, or inserts a vowel between the obstruents, even though both of these processes would produce a surface form that no longer violates the prohibition on voicing disagreement. Steriade argues that this is because assimilation results in a surface form that is more similar to the underlying form (or to morphologically related surface forms) than would result from deletion or epenthesis.[vii] This bias towards minimal changes demonstrated in the typology of phonological processes can be thought of as optimizing, in that it increases the chances of a listener or learner correctly recognizing multiple realizations of the same morpheme in different contexts (although one must be careful to rule out alternative analyses based on ease of production, which in many cases predict the same types of changes).

Similarity maximization has also been documented extensively by Fleischhacker (2005) in two other phenomena, loanword adaptation and reduplication. Fleischhacker shows that these processes operate in ways that tend to minimize the perceptual difference between the two strings. For example, she shows that in some languages with reduplication, the reduplicant faithfully copies a consonant cluster if it consists of a sibilant and a stop (as in Gothic *ste-stald*), but not if the consonants are an obstruent and a sonorant (as in Gothic *ge-grot*), whereas the opposite pattern is not attested (Chapter 106: Reduplication). Using evidence from perception experiments, partial rhymes in poetry (Zwicky 1976; Kawahara 2007) and imperfect puns (Zwicky & Zwicky 1986), Fleischhacker argues that this asymmetry is due to the fact that (to use the Gothic example) *ge-* and *gr-* are more perceptually similar to one another than *se-* and *st-*.

3.2 The influence of perception on phonology: Mechanisms

Although there is widespread agreement that the facts of speech perception shape phonologies, there is less consensus on the causal mechanisms involved. Two types of explanation have been proposed. A *misperception* account explains perceptually optimizing processes as accidental results of language learners' confusing sounds in contexts where the cues needed to distinguish them are weak or absent (Ohala 1981, 1993; Hale & Reiss 2000; Blevins 2004, 2006, 2007). On this view, listeners' errors accumulate in the lexicon, resulting in a state in which a given process is instantiated in the data to which future generations of learners are exposed. These misperceptions could occur in a number of ways (Blevins 2004): coarticulation effects may be misinterpreted as phonological alternations (resulting in assimilation), listeners may attempt to undo a coarticulatory effect that they believe has taken place (resulting in dissimilation), or listeners may hear a range of variant pronunciations of a sound, and misidentify which is the prototypical variant.

Under a misperception account of nasal place assimilation, for example, learners would tend to misperceive [anpa] as [ampa]; once enough of these mistakes have been made, learners will posit a process of place assimilation, even if no such process was employed by the previous generation. Of course, for misperceptions to result in perceptually optimizing phonologies, confusion between sounds must be asymmetric; in the example above, listeners must mishear [np] as [mp] more often then they mishear [mp] as [np]. In an experiment carried out by Ohala (1990), English speaking subjects did indeed misperceive heterorganic nasal–stop clusters as homorganic clusters 93% of the time (see Beddor & Evans-Romaine 1995 for a similar experiment with an 80% homorganic error rate). However, it should be noted that the fact that nasal–stop clusters in English are overwhelmingly homorganic (Hay *et al.* 1998) may have biased the subjects' perception; as discussed in §2.2, there is evidence that categorical native-language phonotactics can affect listeners' perception of ambiguous stimuli (Massaro & Cohen 1983; Moreton 1999; Moreton & Amano 1999), and it is plausible that gradient phonotactics may have a similar effect (Chapter 5: Gradience vs. Categoricality; Chapter 000: Frequency Effects). Further research is called for to identify asymmetries in misperception which are independent of language experience, and determine the extent to which they align with the directionality of common phonological processes.

Another possible critique of this approach challenges the claim that all perceptual effects are the outcome of misperceptions. Flemming (2005) argues that misperception can only account for processes that eliminate contrasts, as in the nasal place assimilation example, and not for processes that enhance contrasts, as in the dispersion of vowel inventories. Indeed, it is hard to see at first how perceptual errors could result in making a distinction clearer. Much recent work using computational models, however, has shown that under the right conditions *self-organization* (Nicolis & Prigogine 1977) can lead to contrast enhancement even when the perception of categories is disturbed by random noise. De Boer (2001) demonstrates that a group of simulated agents which attempt to imitate each others' vowel productions evolve realistically dispersed vowel inventories over time, even in the absence of any agent-internal biases towards optimization.

In de Boer's model, an agent represents each vowel category as a cloud of tokens, representing all of the vowels heard by the agent which were classified as belonging to that category. It thus strongly resembles *exemplar theory*, the hypothesis that highly phonetically detailed tokens of sounds (or larger units such as words) are stored by listeners, and that phonological categories consist of sets of these tokens (Pierrehumbert 2001). Blevins & Wedel (2009) identify two processes that can maintain category distinctness in an exemplar model: *variant trading*, in which ambiguous exemplars (i.e. those near the border between two categories) are more likely to be misidentified than prototypical exemplars, and *variant pruning*, in which ambiguous exemplars are more likely to fail to be assigned to any category at all. Both processes have the effect of pushing categories apart, since exemplars further from the border between the categories have a higher change of surviving (i.e. being correctly identified), and thus a greater influence over the center of the category as a whole. This work, together with other recent research using agent-based simulations (Wedel 2006, 2007; Kochetov & So 2007; Boersma & Hamann 2008), has shown that surprising and counterintuitive effects can emerge globally in a system of very simple interacting agents, and provides an important 'proof of concept' for the misperception approach to perceptual effects on phonology.

An alternative to the misperception account is what we will call the *cognitive bias* approach, which sees phonological processes as being implemented *in order to* minimize the possibility of listener errors (Hayes 1999; Steriade 1999, 2001, 2008; Hayes & Steriade 2004). On this view, optimization is an explicit goal of phonological systems. A cognitive bias account would explain the ubiquity of perceptually optimized phonological processes as the result of knowledge on the part of learners which biases learning in some way. There are several ways these biases might manifest themselves. In the case of nasal place assimilation, for example, learners who know that homorganic NC clusters are 'better' than heterorganic clusters, for example, might simply be unable to learn a language in which [np] is legal, but [mp] is not. Alternatively, learners may come equipped with a soft bias against the pathological heterorganic-clusters-only language – they could learn such a grammar, but only when given much more evidence than would be required to learn a homorganic-clusters-only language. If these biases are universal, then even if sound change were random and unconstrained, those languages that conform to learners' expectations would flourish at the expense of those that do not.

One way to formalize these cognitive biases is to posit optimality-theoretic (OT) markedness constraints which prohibit problematic sequences, which interact with faithfulness constraints to produce the observed typological distribution of languages (Chapter 66: Markedness and Faithfulness Constraints). Nasal place assimilation could be modeled as the effects of a constraint against heterorganic clusters and the absence of a corresponding constraint banning homorganic clusters. This type of analysis, however, would not work in other cases. How could dispersion effects be explained with constraints on output forms? In languages with one high vowel, it will be central, but languages with two high vowels ban the central vowel in favor of two peripheral vowels. The grammar cannot 'know' whether a front high vowel is marked or not without knowing what other vowels the language uses. The problem stems from the fact, mentioned earlier, that perceptual difficulty is a property of contrasts between sounds, not of individual sounds or sound sequences. Under standard formulations of OT (Prince & Smolensky 2004), however, phonological contrasts are emergent properties of an entire constraint ranking, and as such cannot be referred to directly by markedness constraints. Two examples of ways around this problem are represented by Flemming's (2002) *distinctiveness constraints* and Steriade's *P-map* (Steriade 2008).

Flemming (2002) solves the problem by expanding the definition of OT markedness constraints to allow them to refer to contrasts, or to entire sound inventories. These distinctiveness constraints are of two types, MAXIMIZECONTRASTS constraints, which prefer a greater number of distinct sounds, and perceptually motivated MINIMIMUMDISTANCE constraints, which ban inventories in which sounds are not sufficiently distinct. A learner equipped with these constraints will be able to posit a ranking that will generate a typical, well dispersed vowel inventory, but no ranking will generate an unattested inventory, such as one in which there are only mid and low vowels.

Steriade's approach is to posit a 'map' of perceptibility, realized as a matrix of sound pairs and contexts, with a similarity value assigned to each pair in each context. This 'P-map' is essentially a representation a speaker's knowledge of perceptual similarity (whether innate or somehow deduced by the speaker). The P-map is held to regulate rankings among faithfulness constraints, and it is these fixed rankings that result in perceptually optimizing typological biases. Recall the case discussed in §1.3 of voicing assimilation in obstruent clusters, which is preferred over deletion as a solution to voicing disagreement. According to Steriade, speakers faced with a cluster like /kd/ know that the difference between [k] and [g] is smaller than the difference between [k] and Ø. This translates into the fixed constraint ranking MAX-C >> IDENT[voice], with the consequence that no possible ranking will generate the unattested change /kd/ → [d].

Evaluating which of these theories is most likely to be true is made difficult by the fact that their empirical predictions largely overlap. We conclude this section by considering ways in which they may be teased apart. The first concerns parsimony. As many proponents of the misperception approach argue, if typological asymmetries can be shown to emerge without the need for a complex set of synchronic cognitive biases, then the diachronic analysis is the simplest, and should therefore be preferred. It is customary that in the absence of any empirical evidence, simpler explanations should be adopted over more complex ones, and it is surely healthy to question assumptions about the supremacy of synchronic over diachronic analyses. But ultimately, the question will be decided on empirical grounds.

A more empirical way of distinguishing the two approaches begins by independently establishing the plausibility of the mechanisms proposed by each theory. In the misperception case, this can take two forms. First, experiments can be used to determine under what conditions listeners in fact make perceptual errors, and the results can be compared to the typological facts. Several experiments (Fujimura *et al.* 1978; Ohala 1990; Hura *et al.* 1992) have shown, for example, that nasal place cues are difficult to identify before consonants, the same environment which frequently triggers nasal place assimilation. Guion (1998) shows that the sequences [ki] and [ʧi] are highly confusible when compared to similar sequences, which may explain why velar palatalization before [i] is a common process synchronically and diachronically. Kochetov & So

(2007) show that the perceptibility of released stops in various environments correlates with cross-linguistic patterns of place assimilation.

Second, because misperception accounts rely on effects that emerge globally within a network of speakers and listeners, rather than inside the head of each individual, computer simulations involving multiple speakers can be used to explore how biases develop and propagate throughout a speech community. As discussed above, de Boer's (2001) work on the evolution of vowel systems falls under this heading, as does Wedel's (2006) research into the emergence and maintenance of linguistic contrasts. Boersma & Hamann (2008) show that optimally dispersed sibilant inventories can evolve over multiple generations when agents produce sounds using the same grammar used for perception. Simulations such as these do not by themselves provide direct evidence of the misperception hypothesis, but they are a crucial step in delineating the precise empirical predictions made by the theory.

Similarly, proponents of the cognitive bias model have sought to establish its plausibility by looking for evidence for these biases in domains for which diachronic explanations do not apply. Kawahara (2007) shows, for example, that the lyrics of Japanese rap songs employ a type of line-final rhyme which has no precursor in Japanese poetic tradition. Kawahara shows that the rhymes attested in a corpus of lyrics adhere to a hierarchy of perceptual similarity which plays no role in the rest of Japanese phonology, suggesting that the Japanese speakers who composed these lyrics possess a knowledge of perceptual similarity relations that is not learned.

Wilson (2006) comes to a similar conclusion using an artificial language learning experiment. He shows that English-speaking subjects who are taught a velar palatalization rule which changes /k/ to [ʧ] before [e] generalize that rule to contexts in which the vowel is [i], while subjects taught a rule that only applies before [i] do not generalize it to [e]. Wilson attributes these results to a learning bias that is motivated by the subjects' (presumably unconscious) knowledge that [ki] and [ʧi] are more perceptually similar than [ke] and [ʧe] (Guion 1998). Skoruppa *et al.* (in press) also use an artificial language learning experiment to show that subjects find it easier to learn phonological rules which implement small changes.

Of course, evidence that adults use biases when faced with novel phonological situations does not itself constitute evidence that children use the same biases when acquiring their native language (Chapter 101: Experimental Approaches in Theoretical Phonology). The next step, for both theories, is to look for evidence that their proposed mechanisms are in fact operating in the real world of language learning and use. How often do language learners, outside of a laboratory, actually misperceive sounds, and how frequently must misperception occur in order to allow a given sound change to spread throughout a speech community? What evidence is there that children employ learning biases in the early stages of acquisition?[viii] To what extent are typological patterns are driven by historical forces other than sound change, for example the preferential adoption or retention of some words over others (Martin 2007)?

Answering these questions would be difficult, but not impossible, and the relevant research has for the most part yet to be done. Data could come from detailed examinations of diachronic changes, either those currently in progress or those for which there is a written record, or from experiments involving numbers of interacting subjects that more closely resemble an actual language learning environment.

In a sense, both theories agree on the fact that linguistic sound systems are optimized for human perception – they disagree only on how this optimization takes place. In the misperception approach, optimization takes place globally, as the overall effect of many 'innocent misapprehensions'. The cognitive bias theory claims that optimization is local, occurring inside the head of each speaker. The two theories are not mutually incompatible (see Moreton 2008 for discussion). It is certainly possible that biases in diachronic change shape phonological systems, and that learners have also become biased so as to more easily learn the types of phonological patterns that they are most likely to encounter.

4 Conclusion

Learning the phonology of a language involves, among other things, adjusting one's expectations in accordance with the language one has already heard. This makes the learner better able to perceive the distinctions used by her native language, albeit at the cost of a reduced ability to perceive non-native distinctions. On this view, the fact that the native phonology influences perception is not surprising – it is simply a consequence of the trade-off involved in tuning the perceptual system to expect input of a certain type. In this chapter, we surveyed evidence that the perception of phonemic and allophonic categories is biased by the phonemic categories and phonotactic restrictions of one's native language, and that these factors also influence the adaptation of words borrowed from other languages. Further research on this topic promises to give us a clearer picture of exactly which properties of the native language influence speech perception.

We also discussed evidence of influence in the other direction, from perception to phonology. This influence operates over historical time, changing languages in the direction of eliminating difficult-to-perceive contrasts, whether by neutralizing the problematic contrasts, or making them more distinct, ultimately resulting in a linguistic typology which appears to be perceptually optimized. Current debate on this topic is centered on the question of what drives this historical change – synchronically active learning biases, or perceptual errors on the part of language learners. Both hypotheses make strong claims regarding the fundamental structure of phonological grammars, and the results of this debate have the potential to shape the direction phonological theory will take in the coming decades.

REFERENCES

Adler, Allison. 2006. Faithfulness and perception in loanword adaptation: A case study from Hawaiian. *Lingua* 116. 1024–1045.
Beddor, Patrice & David Evans-Romaine. 1992. Acoustic-perceptual factors in nasal place assimilation. *Journal of the Acoustical Society of America* 91. 2473.
Beddor, Patrice & David Evans-Romaine. 1995. Acoustic-perceptual factors in phonological assimilations: A study of syllable-final nasals. *Rivista di linguistica* 7. 145–174.
Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 171–206.
Best, Catherine T., Gerald McRoberts & Nomathemba Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14. 345–360.
Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.
Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32(2). 117–166.
Blevins, Juliette. 2007. Interpreting misperception: beauty is in the ear of the beholder. In Maria-Josep Solé, Patrice Streeter Beddor & Manjari Ohala (eds.) *Experimental approaches to phonology*, 144-154. Oxford: Oxford University Press.
Blevins, Juliette. 2009. Phonetic explanations for recurrent sound patterns: Diachronic or Synchronic? In Eric Raimy & Charles E. Cairns (eds.) *Contemporary views on architecture and representations in phonology*, 325–336. Cambridge, MA: MIT Press.
Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26. 143–183.
Bloomfield, Leonard. 1933. *Language.* New York: Holt.
Boer, Bart de. 2001. *The origins of vowel systems.* Oxford: Oxford University Press.
Boersma, Paul & Silke Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25. 217–270.
Boersma, Paul & Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In Andrea Calabrese & W. Leo Wetzels (eds.) *Loan phonology*, 11–58. Amsterdam: & Philadelphia: John Benjamins.
Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume & Keith Johnson. 2008. The impact of allophony versus contrast on speech perception. In Peter Avery, B. Elan Dresher & Keren Rice (eds.) *Phonological contrast: Perception and acquisition*, 146-172. Berlin & New York: Mouton de Gruyter.
Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In Kingston & Beckman (1990). 283–333.
Cristia, Alejandrina & Amanda Seidl. 2008. Is infants' learning of sound patterns constrained by phonological features? *Language Learning and Development* 4. 203–227.
Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578.

Dupoux, Emmanuel, Christopher Pallier, Núria Sebastián-Gallés & Jacques Mehler. 1997. A destressing 'deafness' in French? *Journal of Memory and Language* 36. 406–421.

Dupoux, Emmanuel, Sharon Peperkamp & Núria Sebastián-Gallés. 2001. A robust method to study stress 'deafness'. *Journal of the Acoustical Society of America* 110. 1606–1618.

Dupoux, Emmanuel, Núria Sebastián-Gallés, Eduardo Navarrete & Sharon Peperkamp. 2008. Persistent stress 'deafness': the case of French learners of Spanish. *Cognition* 106. 682–706.

Flege, James Emil. 1995. Second-language speech learning: theory, findings, and problems. In Winifred Strange (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 229–273.

Flege, James Emil, Ocke-Schwen Bohn & Sunyoung Jang. 1997. The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics* 25. 437–470.

Flege, James & Ian MacKay. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26. 1–34.

Fleischhacker, Heidi. 2005. Similarity in phonology: Evidence from reduplication and loan adaptation. Ph.D. dissertation, University of California, Los Angeles.

Flemming, Edward. 2002. *Auditory representations in phonology*. London & New York: Routledge.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Hayes *et al.* (2004). 232–276.

Flemming, Edward. 2005. Speech perception in phonology. In David B. Pisoni & Robert E. Remez (eds.) *The handbook of speech perception*, 156–182. Malden, MA & Oxford: Blackwell.

Fujimura, Osamu, Marian Macchi & Lynn Streeter. 1978. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21. 337–346.

Gallagher, Gillian. 2008. The role of contrast in laryngeal cooccurrence restrictions. *Proceedings of the West Coast Conference on Formal Linguistics* 27. 177–184.

Gbéto, Flavien. 2000. *Les emprunts linguistiques d'origine européenne en Fon (Nouveau Kwa, Gbe: Bénin)*. Cologne: Rüdiger Köppe Verlag.

Goto, Hiromu. 1971. Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. *Neuropsychologia* 9. 317–323.

Guion, Susan. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55. 18–52.

Hale, Mark & Charles Reiss. 2000. 'Substance abuse' and 'dysfunctionalism': Current trends in phonology. *Linguistic Inquiry* 31. 157–169.

Hallé, Pierre & Catherine T. Best. 2007. Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *Journal of the Acoustical Society of America* 121. 2899–2914.

Hallé, Pierre A., Juan Segui, Uli Frauenfelder & Christine Meunier. 1998. Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* 24. 592–608.

Hay, Jennifer, Janet B. Pierrehumbert & Mary E. Beckman. 2004. Speech perception, well-formedness and the statistics of the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.) *Phonetic interpretation: Papers in laboratory phonology VI*. Cambridge: Cambridge University Press. 58–74.

Hayes, Bruce. 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Frederick Newmeyer,

Michael Noonan & Kathleen Wheatley (eds.) *Functionalism and formalism in linguistics.* Vol. 1: *General papers,* 243–285. Amsterdam & Philadelphia: John Benjamins.
Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.) 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.
Hayes, Bruce & Donca Steriade. 2004. Introduction: The phonetic bases of phonological markedness. In Hayes *et al.* (2004). 1–33.
Hooper, Joan B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.
Hume, Elizabeth. 2004. The indeterminacy/attestation model of metathesis. *Language* 80. 203–237.
Hume, Elizabeth & Keith Johnson (eds.) 2001. *The role of speech perception in phonology*. San Diego: Academic Press.
Hura, Susan L., Björn Lindblom & Randy L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35. 59–72.
Jacobs, Haike & Carlos Gussenhoven. 2000. Loan phonology: Perception, salience, the lexicon and OT. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: Phonology, syntax, and acquisition*, 193–210. Oxford: Oxford University Press.
Jacquemot, Charlotte, Christophe Pallier, Denis LeBihan, Stanislas Dehaene & Emmanuel Dupoux. 2003. Phonological grammar shapes the auditory cortex: A functional Magnetic Resonance Imaging study. *Journal of Neuroscience* 23. 9541–9546.
Jaeger, Jeri. 1980. Testing the psychological reality of phonemes. *Language and Speech* 23. 233–253.
Jakobson, Roman. 1941. *Kindersprache, Aphasie, und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell. Translated (1968) by A. R. Keiler as *Child language, aphasia and phonological universals*. The Hague & Paris: Mouton.
Jespersen, Otto. 1904. *Lehrbuch der Phonetik*. Leipzig & Berlin: Teubner.
Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20. 219–273.
Kawahara, Shigeto. 2007. Half rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics* 16. 113–144.
Kazanina, Nina, Colin Phillips & William J. Idsardi. 2006. The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences* 103. 11381–11386.
Kenstowicz, Michael. 2004. Review article: The role of perception in loanword phonology. (Review of Gbéto 2000.) *Studies in African Linguistics* 32. 95–112.
Kenstowicz, Michael & Atiwong Suchato. 2006. Issues in loanword adaptation: A case study from Thai. *Lingua* 116. 921–949.
Kim, Soohee & Emily Curtis. 2002. Phonetic duration of English /s/ and its borrowing into Korean. In Noriko M. Akatsuka & Susan Strauss (eds.) *Japanese/Korean Linguistics 10*, 406–419. Palo Alto: CSLI.
Kingston, John & Mary E. Beckman (eds.) 1990. *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.
Kiparsky, Paul. 1981. Remarks on the metrical structure of the syllable. In Wolfgang U. Dressler, Oskar E. Pfeiffer & John R. Rennison (eds.) *Phonologica 1980*, 245–256. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
Kochetov, Alexei & Connie K. So. 2007. Place assimilation and phonetic grounding: a cross-linguistic perceptual study. *Phonology* 24. 397–432.

LaCharité, Darlene & Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36. 223–258.

Lahiri, Aditi & William Marslen-Wilson. 1991. The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition* 38. 245–294.

Lass, Roger. 1984. Vowel system universals and typology: Prologue to theory. *Phonology Yearbook* 1. 75–111.

Liberman, Isabelle, Donald Shankweiler, F. William Fischer & Bonnie Carter. 1974. Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology* 18. 201–212.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48. 839–862.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala & Jeri J. Jaeger (eds.) *Experimental phonology,* 13–44. Orlando: Academic Press.

Lipski, Silvia & Klaus Mathiak. 2007. A magnetoencephalographic study on auditory processing of native and nonnative fricative contrasts in Polish and German listeners. *Neuroscience Letters* 41. 90–95.

Lively, Scott E., John S. Logan & David B. Pisoni (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94. 1242–1255.

Lovins, Julie. 1975. *Loanwords and the phonological structure of Japanese*. Bloomington: Indiana University Linguistics Club.

Mack, Molly. 1989. Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception and Psychophysics* 46. 189–200.

Martin, Andrew. 2007. The evolving lexicon. Ph.D. dissertation, University of California, Los Angeles.

Massaro, Dominic W. & Michael M. Cohen. 1983. Phonological context in speech perception. *Perception and Psychophysics* 34. 338–348.

McAllister, Robert, James E. Flege & Thorsten Piske. 2002. Influence of L1 on acquisition of Swedish quantity in Spanish, English and Estonian. *Journal of Phonetics* 30. 229–258.

Miyawaki, Kuniko, Winifred Strange, Robert Verbugge, Alvin Liberman, James Jenkins & Osamu Fujimura. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics* 18. 331–340.

Morais, José, Paul Bertelson, Luz Cary & Jesus Alegria. 1986. Literacy training and speech segmentation. *Cognition* 24. 35–64.

Morais, José, Luz Cary, Jésus Alegria & Paul Bertelson. 1979. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7. 323–331.

Moreton, Elliott. 1999. Evidence for phonological grammar in speech perception. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.) *Proceedings of the 14th International Congress of Phonetic Sciences*, 2215-2217. Berkeley: Department of Linguistics, University of California, Berkeley.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25. 83–127.

Moreton, Elliott & Shigeaki Amano. 1999. Phonotactics in the perception of Japanese vowel length: evidence for long-distance dependencies. In *Proceedings of the 6th European Conference on Speech Communication and Technology, Budapest*. 2679–2682.

Moreton, Elliott, Gary Feng & Jennifer L. Smith. 2008. Syllabification, sonority, and perception: new evidence from a language game. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 41. 341–355.

Näätänen, Risto , Anne Lehtokoski, Mietta Lennes, Marie Cheour, Minna Huotilainen, Antti Iivonen, Martti Vainio, Paavo Alku, Risto J. Ilmoniemi, Aavo Luuk, Jüri Allik, Janne Sinkkonen & Kimmo Alho. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385. 432–434.

Ní Chiosáin, Máire & Jaye Padgett. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and representations*. 118–156. Cambridge: Cambridge University Press.

Nicolis, Gregoire & Ilya Prigogine. 1977. *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations.* New York: Wiley.

Ohala, John J. 1981. The listener as a source of sound change. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 17(2). 178–203.

Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. In Kingston & Beckman (1990). 258–275.

Ohala, John J. & Haruko Kawasaki-Fukumori. 1997. Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In Stig Eliasson & Ernst Håkon Jahr (eds.) *Language And Its Ecology: Essays In Memory Of Einar Haugen. Trends in Linguistics. Studies and Monographs, Vol. 100*. 343-365. Berlin: Mouton de Gruyter.

Ohala, John J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13. 155–161.

Otake, Takashi, Kiyoko Yoneyama, Anne Cutler & Arie van der Lugt. 1996. The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America* 100. 3831–3842.

Padgett, Jaye. 2001. Contrast dispersion and Russian palatalization. In Hume & Johnson (2001). 187–218.

Padgett, Jaye & Marzena Zygis. 2003. The evolution of sibilants in Polish and Russian. *ZAS Working Papers in Linguistics* 32. 155–174.

Pallier, Christophe, Laura Bosc & Núria Sebastián-Gallés. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64. B9–B17.

Paradis, Carole & Darlene LaCharité. 1997. Adult and infant perception of two English phones. *Journal of the Acoustical Society of America* 102. 3742–3753.

Peperkamp, Sharon. 2005. A psycholinguistic theory of loanword adaptation. *Proceedings of the Annual Meeting, Berkeley Linguistics Society* 30. 341–352.

Peperkamp, Sharon, Michèle Pettinato & Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: the role of perception. In Barbara Beachley, Amanda Brown & Francis Conlin (eds.) *Proceedings of the 27th Annual Boston University Conference on Language Development*, 650–661. Somerville, MA: Cascadilla Press.

Peperkamp, Sharon, Inga Vendelin & Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: Experimental evidence from Japanese. *Phonology* 25. 129–164.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.) *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam & Philadelphia: John Benjamins.

Polivanov, Yevgeny. 1931. La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4. 79–96.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar.* Malden, MA & Oxford: Blackwell.

Read, Charles, Zhang Yun-Fei, Nie Hong-Yin & Ding Bao-Qing. 1986. The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition* 24. 31–44.

Saffran, Jenny R. & Erik D. Thiessen (2003). Pattern induction by infant language learners. *Developmental Psychology* 39. 484–494.

Sapir, Edward. 1933. La réalité psychologique des phonèmes. *Journal de Psychologie Normale et Pathologique* 30. 247–265.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale.* Lausanne & Paris: Payot.

Schultze, Wolfgang. 2002. Functional grammar of Udi. Unpublished ms., University of Munich.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée & Christian Abry. 1997. The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics* 25. 255–286.

Sebastián-Gallés, Núria & Salvador Soto-Faraco. 1999. On-line processing of native and nonnative phonemic contrasts in early bilinguals. *Cognition* 72. 112–123.

Sievers, Eduard. 1881. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen.* Leipzig: Breitkopf & Härtel.

Silverman, Daniel. 1992. Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology* 9. 289–328.

Skoruppa, Katrin, Anna Lambrechts & Sharon Peperkamp. Forthcoming. The role of phonetic distance in the acquisition of phonological alternations. *Papers from the Annual Meeting of the North East Linguistic Society* 39.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. Ph.D. dissertation, MIT.

Steriade, Donca. 1999. Phonetics in phonology: The case of laryngeal neutralization. *UCLA Working Papers in Linguistics* 2: *Papers in Phonology* 3. 25–146.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Hume & Johnson (2001). 219–250.

Steriade, Donca. 2008. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Kristin Hanson & Sharon Inkelas (eds.) *The nature of the word: Studies in honor of Paul Kiparsky*, 151–180. Cambridge, MA: MIT Press.

Stevens, Kenneth N., Samuel Jay Keyser & Haruko Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell & Dennis H. Klatt (eds.) *Invariance and variability in speech processes*, 426–449. Hillsdale, NJ: Lawrence Erlbaum.

Swadesh, Morris. 1934. The phonemic principle. *Language* 10. 117–129.

Takagi, Naoyuki. 2002. The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies. *Journal of the Acoustical Society of America* 111. 2887–2896.

Takagi, Naoyuki & Virginia Mann. 1994. A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words. *Journal of Phonetics* 22. 343–356.

Takagi, Naoyuki & Virginia Mann. 1995. The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics* 16. 379–405.

Trubetzkoy, Nikolai S. 1939. *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht.

Vallée, Nathalie. 1994. Systèmes vocaliques: De la typologie aux prédictions. Thèse de Doctorat, Université Stendhal, Grenoble.

Wang, William S.-Y. 1971. The basis of speech. In Carroll E. Reed (ed.) *The Learning of Language*. 267–306. New York: Appleton-Century-Crofts.

Wang, Yue, Michelle M. Spence, Allard Jongman & Joan Sereno. 1999. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America* 106. 3649–3658.

Wedel, Andrew B. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23. 247–274.

Wedel, Andrew B. 2007. Feedback and regularity in the lexicon. *Phonology* 24. 147–185.
Werker, Janet F. & Richard C. Tees. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75. 1866–1878.
Whalen, D. H., Catherine T. Best & Julia R. Irwin. 1997. Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics* 25. 501–528.
Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.
Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Hayes *et al.* (2004). 34–57.
Yip, Moira. 1993. Cantonese loanword phonology and Optimality Theory. *Journal of East Asian Linguistics* 2. 261–291.
Zwicky, Arnold M. 1976. Well, the rock and roll has got to stop. Junior's head is hard as a rock. *Papers from the Annual Regional Meeting, Chicago Linguistic Society* 12. 676–697.
Zwicky, Arnold M. & Elizabeth Zwicky. 1986. Imperfect puns, markedness, and phonological similarity: With fronds like these, who needs anemones? *Linguistica* 20. 493–503.

NOTES

* Research for this chapter was supported by the Agence Nationale de la Recherche (ANR-BLAN-0065-01) and the Centre National de la Recherche Scientifique. We are grateful to Katrin Skoruppa for comments and discussion.

[i] There is at least one earlier study on the processing of abstract phonemes: Jaeger (1980) used a conditioning as well as a concept-formation paradigm to examine whether English listeners perceive different allophones of /k/ to be part of a single category. Unfortunately, the results of these experiments were inconclusive (mainly because of lack of data).

[ii] See Peperkamp *et al.* (2008) for an extensive list of examples, and Chapter 100: Loanword Phonology for discussion.

[iii] Technically, their model minimizes the sum, over all possible vowel pairs, of the inverse square of the perceptual distance between each pair.

[iv] Other factors, such as homophony avoidance, are presumed to regulate the number of vowels, preventing languages from adopting a single-vowel inventory, which is ideal from a purely perceptual standpoint (Flemming 2004).

[v] For more on dispersion theory, including discussion of possible formalizations, see Flemming (2002, 2004, 2005); Padgett (1997); Ní Chiosáin & Padgett (2001).

[vi] Although see Moreton *et al.* (2008) for experimental evidence of an asymmetry between CV and VC syllables that cannot be explained by this perceptual account.

[vii] Steriade defines similarity on the basis of the perceptual confusability of two surface forms, although of course alternative definitions which take into account more abstract properties of the output are possible.

[viii] The experimental studies in Saffran & Thiessen (2003) and Cristià & Seidl (2008) suggest that phonotactic learning in infants is indeed biased in favor of certain types of pattern.