# *Speech Perception in Phonology*

**Edward Flemming**
Stanford University

## 1. Introduction

The idea that the nature of speech perception plays a role in shaping phonology is not new. There is a substantial literature that proposes and tests perceptual explanations for phonological patterns, e.g. Liljencrants and Lindblom (1972), Lindblom (1986), Ohala (1981, 1990, 1993). Most of this research addresses general tendencies in phonological patterning, e.g. the cross-linguistic tendency for front vowels to be unrounded, which leaves the problem of going from the general to the particular: particular languages may violate the general tendencies, as in the case of a language with front rounded vowels. Recently phonologists have begun to tackle this problem, incorporating principles that invoke properties of human speech perception into models that derive generalizations about phonological systems, but also allow for analyses of individual languages. This area of research has proven very productive, and there is now substantial evidence for the importance of perceptual considerations in phonological theory, but there is little agreement on the proper formalization of the influence of speech perception on phonology. This issue is the organizing theme of the chapter[1].

A key element in the development of this research has been Optimality Theory (OT, Prince and Smolensky 1993), which offers a framework for constructing analyses of individual languages out of constraints expressing general preferences of the kind identified in the works cited above. In OT terms, the central question addressed here is: "What is the form of the constraints imposed on phonology by speech perception?" We will review the main types of evidence that have been used to argue for perceptual constraints in phonology to clarify exactly what kind of constraints they motivate. In the

[1] I would like to thank David Pisoni, Robert E. Remez and Donca Steriade for helpful comments on this chapter.

process, we will also examine the kinds of experimental evidence that have been adduced in formulating analyses.

## 2. Dispersion and Enhancement

The most direct evidence for perceptual constraints in phonology comes from generalizations about inventories of phonological contrasts. Phonetic descriptions distinguish hundreds of sound types, but a typical language has only about 30 contrasting sounds (Maddieson 1984:7). These inventories of contrasting sounds are far from being a random sample of the set of attested speech sounds, rather the observed inventories are subject to many restrictions, some of which can be explained in terms of perceptual constraints.

One well established example involves preferences for particular vowel qualities. There is a strong cross-linguistic preference for vowels to be front unrounded or back rounded unless they are low vowels, as in the common vowel inventories illustrated in figure 1 (low vowels are typically described as central or back and unrounded). In Maddieson's (1984) survey of a genetically diverse sample of languages, 94% of front vowels are unrounded and 93.5% of back vowels are rounded. Where a language does have front rounded, central, or back unrounded vowels, these appear in addition to front unrounded and back rounded vowels.

| | | | | | | |
|---|---|---|---|---|---|---|
| i | | u | | i | | u |
| | | | | e | | o |
| | a | | | | a | |

Figure 1. Two common vowel inventories.

It is hard to imagine any articulatory basis for this relationship between backness and rounding. The tongue and lips are articulatorily relatively independent, so it would appear to be as easy to round the lips with the tongue body forward as with it retracted. On the other hand there is a straightforward perceptual account of the covariation of backness and rounding. The primary perceptual dimensions of vowel quality correspond well to the frequencies of the first two formants (Delattre, Liberman, Cooper, and Gerstman 1952, Plomp 1975, Shepard 1972). Front and back vowels are differentiated primarily by the frequency of the second formant (F2), with front vowels having a high F2 and back vowels having a low F2. Lip-rounding generally lowers F2, so the ordering of front and back, rounded and unrounded vowels, and central vowels in terms of F2 is shown in figure 2. Thus the maximally distinct F2 contrast is between front unrounded and back rounded vowels (Liljencrants and Lindblom 1972, Stevens, Keyser, and Kawasaki 1986). Maximally distinct contrasts are preferred because they are less likely to be confused by listeners.

i  y  ɨ ɯ  u

$\xleftarrow{\hspace{4em}}$

F2

Figure 2. The ordering of vowel qualities on the F2 dimension.

The general preference for maximally distinct contrasts follows from the functionalist hypothesis that phonological systems are well adapted for communication. Efficient communication depends on fast, accurate perception of speech sounds, and listeners are faster and more accurate in identifying the category to which a stimulus belongs if the stimulus is more distinct from contrasting categories (e.g. Ashby, Boynton and Lee 1994, Kellogg 1931, Pisoni and Tash 1974, Podgorny and Garner 1979). We will see that the principle of maximization of distinctiveness is the key perceptual constraint on phonology.

Evidence for this principle has been discussed under a variety of labels. Lindblom and Engstrand (1989) refer to the tendency to maximize the perceptual distinctiveness of contrasting speech sounds as ‘dispersion’, invoking the notion of separation in perceptual space. Similar phenomena have been discussed by Stevens, Keyser, and Kawasaki (1986) under the rubric of ‘enhancement’. They observe that distinctive features are often accompanied by ‘redundant’ features that ‘strengthen the acoustic representation of distinctive features and contribute additional properties which help the listener to perceive the distinction’ (p.426). The relationship between [back] and [round] in vowels is treated as one of enhancement: [round] enhances distinctive [back]. So enhancement essentially involves combining feature differences so as to maximize the perceptual distinctiveness of contrasts. Consequently instances of enhancement also provide evidence for maximization of distinctiveness. Other work providing evidence for dispersion/enhancement includes Ohala (1985:225f.), Diehl (1991), and Flemming (2002:53-56). We will review two further cases here to illustrate the range of phenomena involved.

Another example discussed by Stevens et al (1986) is the enhancement of frication contrasts. Fricatives are distinguished from other sound types by the presence of significant turbulence noise, generated by forcing a jet of air through a narrow constriction. The distinctiveness of this manner contrast can thus be enhanced by increasing the intensity of turbulence noise in the fricative. This is achieved by directing the jet of air against an obstacle downstream, as in the coronal sibilant [s], where a jet of air is directed against the upper teeth (Stevens et al 1986:439, Shadle1991). The greater distinctness of such sibilant fricatives from non-fricatives can explain their cross-linguistic prevalence: in Maddieson’s (1984) survey, 83% of languages have some kind of [s], and if a language has only one fricative it is usually an [s] sound (84%).

Maximization of the distinctiveness of contrasts between sibilants has been argued to explain an otherwise puzzling observation about the realization of post-alveolar fricatives: in English and French, the post-alveolar fricative [ʃ] is accompanied by lip protrusion (Ladefoged and Maddieson 1996:148). There is no articulatory basis for this pattern, but it plausibly serves to make post-alveolar [ʃ] more distinct from the anterior sibilant [s]. These sounds are differentiated by the frequency of the first peak in the noise spectrum. This peak is at the resonant frequency of the cavity in front of the constriction, and so is lower in post-alveolar fricatives, since they have a larger front cavity than dentals and alveolars. Protruding the lips increases this difference by further enlarging the front cavity (Ladefoged and Maddieson 1996:149). Polish provides an interesting variant of this pattern. There are three contrasting sibilants, dental [s̪], alveopalatal [ɕ], and retroflex (apical post-alveolar) [ʂ], and the retroflex is produced with lip protrusion (Puppel, Nawrocka-Fisiak, and Krassowska 1977:157). This is the expected pattern given the goal of maximizing distinctiveness because the retroflex has the lowest front cavity resonance due to the space below the tongue blade. Lowering this resonance further by protruding the lips makes the retroflex more distinct from the other sibilants (Flemming 2002:55f.).

### 1.1 Phonological analyses of dispersion effects

There have been two basic approaches to the analysis of dispersion effects: (i) analyses that incorporate a preference for maximally distinct contrasts into phonological theory, and (ii) analyses that employ standard markedness constraints. The latter approach is in a sense the default option, since it employs only the standard apparatus of phonological theory (as outlined in the next section), but we will see that dispersion effects provide strong evidence for the distinctiveness constraints posited in the former approach.

We will first provide a brief overview of Optimality Theory (OT, Prince and Smolensky 1993) and its suitability as a framework for formalizing the influence of speech perception on phonology, then we will turn to the particular proposals for formalizing perceptual constraints[2]. Although the discussion will focus on analyses formulated in OT, the issues raised are relevant to any analysis of these phenomena.

[2] For more detailed introductions to OT, see Kager (1998) and McCarthy (2001).

## 1.2 Optimality Theory

In its basic form, an OT grammar maps input underlying forms onto their surface realizations. For example, in Russian obstruents are devoiced in word-final position, so the morpheme /sad/ ‘garden’ is pronounced [sat] (the underlying voiced final stop surfaces when a vowel-initial suffix is added, as in the dative singular [sadu]). In OT, the mapping between input and output is divided into two components: a mapping from an input form to a set of candidate outputs, and an evaluation function which selects the best member of the candidate set as the actual output. The optimality of candidate outputs is determined by reference to a ranked set of constraints.

Standard OT posits two basic types of constraints: constraints that evaluate the well-formedness of the candidate outputs—markedness constraints—and constraints that require the output to be as similar to the input as possible—faithfulness constraints. These two types of constraints are liable to conflict – satisfying markedness constraints often requires altering the input, which necessarily violates some faithfulness constraint. For example, a simple-minded analysis of the Russian facts above posits a markedness constraint forbidding word-final voiced obstruents, *FINALVOICEDOBSTRUENT. The fully faithful realization of [sad] violates this constraint, but devoicing the final stop, as in [sat] violates the faithfulness constraint IDENT[VOICE] which requires that voicing specifications of input segments should be unchanged in the output.

Conflicts between constraints are resolved by reference to a ranking of the constraints: the higher ranked constraint prevails. So in Russian, *FINALVOICEDOBSTRUENT must outrank IDENT[VOICE] (written: *FINALVOICEDOBSTRUENT >> IDENT[VOICE]) since the voicing of an input stop is changed in order to satisfy the former constraint. If this ranking were reversed the candidate [sad] would win.

OT analyses are typically illustrated using tableaux, as in (1). The input form is shown in the top left cell while the candidate outputs are listed below it in the first column. The constraints are listed in the top row, with higher-ranked constraints on the left. If a candidate violates a constraint, a mark (*) is placed at the intersection of the constraint column and the candidate row. In (1), candidate (a), [sad] violates *FINALVOICEDOBSTRUENT, so a mark is placed under that constraint in row (a). Candidate (b), [sat], satisfies this constraint, so [sad] is eliminated (indicated by the exclamation point after the mark), and [sat] is the optimal output (indicated by the 'pointing hand' in the first column). Note that it is not necessary to satisfy all the constraints in order to be the optimal candidate – candidate (b), [sat], is optimal although it violates IDENT[VOICE]. Indeed, since constraints frequently conflict, it is not usually possible to satisfy them all.

| (1) | /sad/ | *FINALVOICEDOBS | IDENT[VOICE] |
|---|---|---|---|
| a. | sad | *! | |
| b. | ☞ sat | | * |

One of the key strengths of OT is the way in which it relates the analysis of the typology of languages to the analyses of individual languages. The two are connected by the hypothesis that all phonological grammars are constructed from the same set of constraints, but differ in the ranking of those constraints. Typological universals can then be derived from the nature of the universal set of constraints, while the patterns of individual languages are hypothesized to derive from particular rankings of these constraints.

This provides a suitable framework for formalizing the preference for distinct contrasts because any such preference is a universal tendency which may be violated to a greater or lesser extent as a result of conflicting constraints. For example, languages like French and German have front rounded vowels in addition to front unrounded and back rounded vowels. That is, these languages eschew maximally distinct F2 contrasts in favour of distinguishing more contrastive vowels. Conversely, grounding constraints in basic considerations of communicative efficiency, and the nature of human speech perception provides a basis for the universality of those constraints: if a constraint is based on universal properties of communication and perception, it is unsurprising that it is operative in all languages.

### 1.3 The constraints that motivate dispersion

Two kinds of constraints have been proposed in the analysis of dispersion phenomena: basic segmental markedness constraints, and constraints on the distinctiveness of contrasts. A basic markedness constraint in OT prohibits some representational structure, such as a syllable without an onset, or a segment which has the feature combination [-sonorant, +voice]. A number of researchers have suggested that constraints of this form can be motivated by perceptual considerations (e.g. Hume 1998, Côté 2000). Certainly, the most common analysis of the preference for peripheral vowels (i.e. front unrounded and back rounded vowels) has been to propose constraints against other types of vowels, as in (2) (e.g. Calabrese 1988)[3].

(2) *[-back, +round]
*[+back, -round]

Ranking these constraints above faithfulness to [back] or [round] yields a language without non-peripheral vowels because inputs containing these vowels will not be realized faithfully (3-4)[4].

(3)

| | /y/ | *[-back, +round] | *[+back, -round] | IDENT[round] |
|---|---|---|---|---|
| a. | ☞ i | | | * |
| b. | y | *! | | |

[3] It is common to specify central vowels as [+back, -round], in which cases these constraints are sufficient. If central vowels are distinguished from back unrounded vowels, a constraint against this class of vowels is required also.
[4] The dotted line between the top two constraints in (5-6) indicates that the relative ranking of these constraints cannot be determined – that is, either ranking yields the desired result.

| (4) | /ɯ/ | *[-back, +round] | *[+back, -round] | IDENT[round] |
|---|---|---|---|---|
| a. | ☞ u | | | * |
| b. | ɯ | | *! | |

Although these constraints can derive languages in which back and round co-vary appropriately, they do not follow directly from the perceptual considerations behind Liljencrants and Lindblom's (1972) analysis. It was suggested that in order to facilitate speech perception, contrasting sounds should be maximally distinct. This explanation implies a dispreference for F2 contrasts involving non-peripheral vowels because they are less distinct than contrasts between front unrounded and back rounded vowels. The constraints in (2) do not mention contrasts – they simply prohibit front rounded, central, and back unrounded vowels. Liljencrants and Lindblom's proposal is implemented more directly by constraints that penalize less distinct contrasts (distinctiveness constraints), e.g. a constraint ranking along the lines shown in (5), where *X-Y means that words should not be minimally differentiated by the contrast between sounds X and Y (more general formulations are discussed below).

(5) *y-ɯ >> *i-ɯ, *y-u >> *i-u

The crucial difference between these two proposals is that the analysis based on distinctiveness constraints predicts that non-peripheral vowels should be unproblematic as long as they do not enter into front-back (F2) contrasts, whereas the constraints in (2) ban these sound types regardless of what they contrast with. For example, a back unrounded vowel presents no particular perceptual difficulties if the listener knows that it is the only vowel that can appear in the context. It does not violate *i-ɯ or any other distinctiveness constraint because there is no contrast, but it would violate *[+back, -round].

In general, the reasoning outlined above motivates constraints based on the distinctiveness of contrasts between sounds, not on the sounds themselves. Basic markedness constraints as in (2) apply to individual sounds, not contrasts, and so cannot be motivated in this way. More importantly, there is empirical evidence that phonology is in fact subject to constraints on the distinctiveness of contrasts: the markedness of a sound depends on the contrasts that it enters into.

Before evaluating this evidence, it is useful to place distinctiveness constraints in the context of a specific model. The most developed proposal is the dispersion theory of contrast (Flemming 1995, 2001, 2002, Ní Chiosáin and Padgett 2001), which builds on ideas from Lindblom's Theory of Adaptive Dispersion. In this model, the preference to maximize the distinctiveness of contrasts is opposed by two other goals: maximization of the number of contrasts permitted in any given context, and minimization of articulatory effort. Increasing the number of contrasting sounds makes more efficient communication possible by increasing the information content of each sound, since it allows a single segment to differentiate more words. This goal conflicts with maximizing distinctiveness because fitting more contrasts into the finite space of possible speech sounds implies that the sounds must be closer together. Avoiding effortful articulations further restricts the possibilities for realizing distinct contrasts, so this principle also conflicts with maximization of distinctiveness[5]. Thus selecting a set of contrasts that best satisfies these three goals involves finding an optimal balance between them (cf. Lindblom 1986). This optimization is modeled within the framework of OT.

[5] Articulatory effort is not necessarily equivalent to energy expenditure. Although this is usually assumed to be an important component of articulatory effort (e.g. Lindblom 1983, Nelson 1983, Kirchner 1998), there may also be costs associated with precision, for example. The aspect of effort that is most relevant in the examples discussed here relates to the smoothness of movements – movements are hypothesized to be more difficult if they involve abrupt changes in direction. It has been observed that humans generally employ smooth trajectories in speech production (Perkell 1997:357) and in arm movements (e.g. Flash and Hogan 1985, Uno, Kawato and Suzuki 1989). This preference has been attributed to minimization of energy expenditure (Nelson 1983), but it has also been analyzed in terms of minimizing error in the face of noise internal to the motor control system (Wolpert and Harris 1998).

The preference to maximize the distinctiveness of contrasts is implemented in terms of a ranked set of constraints requiring a specified minimum perceptual distance between contrasting forms (6). Sounds are represented as located in a multi-dimensional perceptual space where closer sounds are more confusable. For example, (7) shows the assumed location of high vowels on the dimension corresponding to F2 frequency, measured in arbitrary units. Assuming for simplicity that these vowels differ on this dimension only, it can be seen that the contrast [i-u] involves a distance of 4, and thus satisfies all the MINDIST constraints in (6), while [i-y] involves a distance of only 1, and thus violates MINDIST = 2 and all lower-ranked constraints. In other words, the less distinct a contrast is, the greater the violation.

(6) MINDIST = 1 >> MINDIST = 2 >>... >> MINDIST = 4

| | | | | | |
|---|---|---|---|---|---|
| (7) F2: | 5 | 4 | 3 | 2 | 1 |
| | i | y | ɨ | ɯ | u |

The preference to maximize the number of contrasts is implemented as a positive constraint, MAXIMIZE CONTRASTS, which is satisfied by the largest inventory of contrasts. The conflict between these two types of constraints is illustrated in (10-11) with the simple example of selecting a set of contrasting high vowels. The balance between maximizing distinctiveness and maximizing the number of contrasts is determined by position of MAXIMIZE CONTRASTS in the hierarchy of MINDIST constraints. In (8), MINDIST=3 outranks MAXIMIZE CONTRASTS, so the largest inventory, (d), is eliminated, because it is does not satisfy MINDIST=3. The most distinct inventory (a), containing front unrounded and back rounded vowels, best satisfies the MINDIST constraints, and hence is the winner. Contrasts involving back unrounded vowels (b), or front rounded vowels (c) are less distinct, and therefore lose to candidate (a).

(8)

| | | MINDIST = 2 | MINDIST = 3 | MAXIMIZE CONTRASTS | MINDIST = 4 |
|---|---|---|---|---|---|
| a. | ☞ i-u | | | ✓✓ | |
| b. | i-ɯ | | | ✓✓ | *! |
| c. | y-u | | | ✓✓ | *! |
| d. | i-ɨ-u | | *! | ✓✓✓ | ** |

In (9), MAXIMIZE CONTRASTS ranks above MINDIST=3 – i.e. the number of contrasts is more important. So the winning candidate is (d) which fits in three contrasting vowels while satisfying the higher-ranked constraint MINDIST=2.

(9)

| | | MINDIST = 2 | MAXIMIZE CONTRASTS | MINDIST = 3 | MINDIST = 4 |
|---|---|---|---|---|---|
| a. | i-u | | ✓✓! | | |
| b. | i-ɯ | | ✓✓! | | * |
| c. | y-u | | ✓✓! | | * |
| d. | ☞ i-ɨ-u | | ✓✓✓ | * | ** |

Effort minimization is assumed to play a negligible role in the selection of F2 contrasts in most contexts[6], but in other cases it may play a role in explaining why languages do not avail themselves of maximally distinct contrasts.

Another consequence of effort minimization is that difficult articulations should only be employed in order to realize more distinct contrasts, so where contrasts are neutralized, considerations of effort minimization are likely to be dominant. This leads to the prediction that preferred vowel qualities should depend on contrastive status: in F2 contrasts, front unrounded and back rounded vowels are preferred (8), but if all vowel F2 contrasts are neutralized, backness and rounding of vowels should be governed by effort minimization. On the other hand, the basic markedness constraints in (2) are insensitive

[6] Effort becomes a more significant factor where vowel duration is very short, and in the environment of consonants that place strong constraints on F2, such as palatalized and velarized consonants (Flemming 2001).

to contrastive status, and consequently predict that peripheral vowels should be preferred in all circumstances.

Flemming (2001) discusses two test cases in which all F2 contrasts are neutralized: 'vertical' vowel inventories, and fully neutralizing vowel reduction in unstressed syllables, as in English reduction to 'schwa'. Both cases conform to the predictions of the dispersion-theoretic analysis: backness and rounding of vowels assimilate to adjacent consonants, often yielding central or centralized vowel qualities which would be highly marked in F2 contrasts, but are favoured by effort minimization.

The best-known examples of 'vertical' vowel inventories, lacking F2 contrasts, are found in Northwest Caucasian languages such as Kabardian and Shapsug (Colarusso 1988, 1992, Kuipers 1960, Smeets 1984). These languages are often described as having only central vowels, but this is a claim about the underlying vowel inventory posited as part of a derivational analysis, not an observation about the surface vowels. On the surface, these languages have a system of five normal length vowels [i, e, a, o, u] (Kuipers 1960:23f., Smeets 1984:123), and 'vertical' system of two extra short vowels, which can be transcribed broadly as [ɨ, ə][7]. However, the precise backness and rounding of these vowels depends on context. They are realized as a smooth transition between the lip and tongue positions of the preceding and following consonants, deviating only to realize the required vowel height (Colarusso 1988:307). An unrelated vertical vowel language, Marshallese, is similar (Bender 1968, Choi 1992). The transitional vowel qualities result from assimilation in backness and rounding to preceding and following consonants, which is plausibly the least effort production strategy. The resulting vowel qualities are often central, back unrounded, front rounded, or short diphthongs involving these qualities – all vowel types which would be highly marked in the presence of F2

[7] Kuipers actually transcribes the Kabardian high vowel as [ə], the mid-vowel as [a], and the 'long' low vowel as [ā], and Colarusso (1988) follows him in this, but their descriptions, Colarusso's phonetic transcriptions, and acoustic data in Choi (1991) all indicate that the vowels are actually high and mid respectively.

contrasts. There are no vertical vowel inventories containing the peripheral vowels that are predicted by the basic markedness constraints in (2) – i.e. there are no inventories such as [i, e, a] or [u, o, a][8].

Neutralization of F2 contrasts is also observed in languages such as English where all vowel quality distinctions are neutralized to a ‘schwa’ vowel in some unstressed syllables. This process is also found in Southern Italian dialects (Maiden 1995) and Dutch (Booij 1995). Phonetic studies of schwa in Dutch (van Bergem 1994) and English (Kondo 1994) indicate that this vowel is comparable to a vertical vowel in that F2 is an almost linear interpolation between values determined by the preceding and following contexts. Again, schwa is a marked vowel where there are quality contrasts – it is often excluded from those positions – but it is the unmarked vowel where all quality contrasts are neutralized. Basic markedness constraints predict that markedness should not depend on contrastive status, so we should expect one of the peripheral vowels, [i, u] or [a], to be the sole vowel in neutralization contexts.

Distinctiveness constraints and basic markedness constraints are also differentiated by predictions concerning enhancement. Distinctiveness constraints predict that enhancement should only apply to contrasts, since enhancement is analyzed as a consequence of constraints on the distinctiveness of contrasts. This is inherent in Stevens et al (1986) conception of enhancement, but it is not predicted by analyses in terms of basic markedness constraints, because the latter are insensitive to contrast. Evidence on this point comes from enhancement of stop voicing contrasts (Flemming 2001). Stevens et al (1986:439) argue that pre-nasalization can serve as an enhancement of stop voicing. One of the cues that distinguishes voiced stops from voiceless stops is the presence of voicing during the closure, as opposed to the silence of a voiceless stop closure (Stevens

[8] It might be suggested that vertical vowels are phonologically unspecified for [back] and [round] rather than being specified for the marked vowel qualities described here (Choi 1992). However, such unspecified vowels only occur in the absence of F2 contrasts, so this would imply an even more dramatic change from a preference for peripheral vowels in F2 contrasts to a preference for otherwise unattested unspecified vowels where there is no contrast.

and Blumstein 1981), so the distinctiveness of this contrast can be increased by increasing the intensity of voicing. This can be achieved by lowering the velum during the early part of the stop closure, yielding a pre-nasalized stop. It is generally difficult to sustain voicing during a stop because air pressure builds up behind the closure, and when oral pressure approaches sub-glottal pressure, airflow through the glottis ceases, and voicing ceases (Ohala 1983, Westbury and Keating 1986). Lowering the velum during the stop closure allows air to be vented through the nose, slowing the build up of oral pressure, and thus facilitating voicing. In addition, voicing during an oral stop is radiated only through the neck and face, resulting in a low intensity acoustic signal, whereas lowering the velum allows sound to be radiated from the nose, resulting in greater intensity.

Pre-nasalization serves as an enhancement of stop voicing contrasts in Mixtec (Iverson and Salmons 1996), Southern Barasano (Smith and Smith 1971), Guaraní (Gregores and Suárez 1967), and a variety of other languages discussed by Herbert (1986:16ff.) – that is, voiceless stops are contrasted with pre-nasalized stops rather than plain voiced stops. But voiced stops are never enhanced by prenasalization where they do not contrast with voiceless stops. Non-contrastive voiced stops can arise through inter-vocalic voicing, a pattern where voiced stops are found between vowels ([ada], not *[ata], but only voiceless stops occur elsewhere ([ta], not *[da]). However we do not find intervocalic prenasalization of stops (i.e. prenasalized stops between vowels, but only voiceless stops elsewhere)[9].

These generalizations are very difficult to account for with simple markedness constraints. The existence of languages which have pre-nasalized stops but not plain voiced stops shows that some markedness constraint must favour pre-nasalized stops over voiced stops, e.g. PRENASALIZE ‘voiced stops should be prenasalized’. Then a language

[9] See Kingston and Diehl (1994) for a related argument that voicing-dependent perturbations of $F_o$ adjacent to stops are active enhancements of stop voicing contrasts, so these effects are reduced or absent where there is no voicing contrast.

with voiceless stops and prenasalized stops (like Mixtec) would be derived by ranking this constraint above faithfulness to [nasal] (10) so any voiced stops in the input are replaced by prenasalized stops.

(10) PRENASALIZE >> IDENT[nasal]

However, this ranking derives prenasalization of voiced stops even where voicing is not contrastive. For example, if intervocalic voicing of stops follows from ranking a constraint against voiceless stops occurring between vowels (*VTV) above faithfulness to voicing (11), then this ranking can be combined with the prenasalization ranking in (10) to derive the unattested pattern of intervocalic prenasalization, as shown in (12).

(11) *VTV >> IDENT[voice]

(12)

| | /ata/ | *VTV | PRE-NASALIZE | IDENT [nasal] | IDENT [voice] |
|---|---|---|---|---|---|
| a. | ata | *! | | | |
| b. | ada | | *! | | * |
| c. | ☞ a$^n$da | | | * | * |

This consequence is avoided if the constraint PRENASALIZE is replaced by constraints favouring maximally distinct voicing contrasts, e.g. *T-D >> *T-$^N$D (where T, D, and $^N$D represent voiceless, voiced and prenasalized stops, respectively). These distinctiveness constraints only apply to contrasts, so prenasalization of voiced stops is correctly predicted to occur only where there are voicing contrasts. Elsewhere voiced stops are preferred over prenasalized stops because voiced stops are simpler articulatorily.

These, and other examples discussed in Flemming (2001), indicate that phonology includes distinctiveness constraints, as we would expect if considerations of ease of perception influence phonology. Basic markedness constraints do not follow from perceptual considerations and cannot account for dispersion effects because dispersion applies only to contrasts while basic markedness constraints are indifferent to the contrastive status of a sound.

### 3. Licensing by Cue

A second source of evidence for perceptual constraints is the typology of contextual neutralization. Contextual neutralization is a pattern of distribution in which a contrast is permitted in some environments, but is suspended in others. For example stop voicing contrasts may be permitted before sonorants ([**b**a] vs. [**p**a], [**b**la] vs. [**p**la]), but not before obstruents ([a**p**ta], *[a**b**ta]). In a situation like this, the voicing contrast is said to be *neutralized* before obstruents.

Steriade (1995, 1999) observes that different types of contrast have different characteristic environments of neutralization. For example the following are well-attested patterns of distribution for three types of contrasts, following Steriade (1999)[10]:

(13) (i) Obstruent voicing contrasts are permitted only before sonorants (e.g. German, Lithuanian, Russian, Sanskrit).
(ii) Major place contrasts (labial vs. coronal vs. dorsal) are permitted only before vowels (e.g. Japanese, Luganda, Selayarese).
(iii) Retroflexion contrasts (retroflex vs. apical alveolar) are permitted only after vowels (e.g. Gooniyandi, Miriwung, Walmatjari).

Steriade argues that the general characterization of these diverse contexts of neutralization makes crucial reference to perceptual distinctiveness: in each case, the contrasts are neutralized first in environments where 'the *cues* to the relevant contrast would be diminished or obtainable only at the cost of additional articulatory maneuvers' (Steriade 1997:1). Contrasts differ in their distribution of cues so they are subject to different patterns of neutralization. This is dubbed the 'Licensing by Cue' hypothesis –

[10] References: German, Lithuanian, Russian, Sanskrit: Steriade (1997), Japanese: Itô (1989), Luganda: Tucker (1962), Selayarese: Mithun and Basri (1986), Gooniyandi: McGregor (1990), Miriwung: Hamilton (1996), Walmatjari: Hudson and Richards (1969).

the presence of a contrast in a particular environment is licensed by the availability of perceptual cues to that contrast.

For example, the distribution of obstruent voicing contrasts (12i) is analyzed in these terms by Steriade (1997). One of the primary cues to obstruent voicing distinctions is Voice Onset Time (VOT), the lag between the release of the obstruent constriction and the onset of voicing (Lisker and Abramson 1970). Steriade observes that this cue is generally only available where a voiced sonorant follows, and so is absent before obstruents, and in word-final position before pause. Voicing contrasts in these environments can only be realized by cues such as voicing during the consonant constriction, consonant duration, and duration of the preceding vowel, which are hypothesized to be weaker cues than VOT. So according to this analysis, languages like Russian and German disallow voicing contrasts in precisely the environments where a key cue to the contrast, VOT, is unavailable. Given the importance of VOT as a cue to obstruent voicing, it is very plausible that voicing is less confusable before sonorants than before obstruents or word-finally, but there is surprisingly little direct evidence on this point. Studies of voicing perception generally have not directly compared perception of voicing in different contexts.

Similar factors have been argued to explain restrictions on the distribution of major place contrasts (labial vs. coronal vs. dorsal). These contrasts preferentially occur where there is a following vowel, or, failing that, a following approximant. A number of studies have shown that major place distinctions are less confusable in pre-vocalic position than in pre-consonantal or pre-pausal position (Redford and Diehl 1999, Wright 2001). Fujimura et al (1978) and Ohala (1990) have also shown that release cues to major place contrasts dominate over closure cues in stimuli that have been edited so that these cues conflict. This difference in distinctiveness appears to have multiple causes. The greater distinctiveness of prevocalic stops may be attributed to the presence of the release burst which provides cues to place, in addition to the formant transition cues that are also available in post-vocalic position (Dorman et al 1977). Consonants clusters are often articulatorily overlapped so the constriction of a second consonant is formed before the constriction of the first consonant is released. Where the second consonant is an obstruent, this results in the loss or attenuation of the release burst of the first consonant (Henderson and Repp 1982).

Another factor that has been suggested to contribute to the greater distinctiveness of prevocalic place contrasts is the nature of the peripheral auditory system (Wright 1996, 2001). Auditory nerve fibres respond most strongly to rapid rises from low intensity within their frequency band, and the transition from a consonant to a vowel often involves rapid onsets of this kind, especially where the consonant is an obstruent (Delgutte and Kiang 1984, Greenberg 1995). This effectively amplifies release formant transitions and stop bursts. As noted by Ohala (1990:261f.), experiments by Fujimura et al (1978) support an auditory-perceptual basis for the greater distinctiveness of onset consonants: they found that in stimuli with conflicting cues to place, release cues dominated closure cues, even when the stimuli were played backwards – i.e. the release cues were reversed closure transitions. However, Redford and Diehl (1999) also found that the formant transitions of onsets were more distinctly articulated than word-final consonants, so production differences may play a role in explaining the observed difference in distinctiveness.

The patterns of distribution of obstruent voicing and major place contrasts are broadly similar in that both preferentially occur before sonorants, but there are differences of detail, some of which follow from differences in the nature of the cues to these two types of contrast. When obstruent voicing contrasts are permitted before sonorant consonants, they are allowed before all sonorants, whereas major place contrasts are usually subject to further restrictions. For example, many languages, including English, do not allow coronal stops before coronal laterals, although labials and velars contrast in this environment: *plan, clan, *tlan* (Kawasaki 1982:14).

The insensitivity of voicing contrasts to the nature of a following sonorant is expected given that the primary cue to voicing is VOT. The realization of VOT depends only on the presence of a voiced sonorant of sufficient duration; place of articulation, nasality, and laterality make little difference. On the other hand, primary cues to stop place contrasts are the release burst and formant transitions. Approximants and vowels allow the realization of both, but simply realizing a burst and formant transitions is not adequate to support contrast: the burst and/or formant transitions must be distinct for contrasting places of articulation. The distinctiveness of these cues can be affected by coarticulation with the following vowel or approximant.

Kawasaki (1982:157f.) and Flemming (2002:132ff.) argue that these factors underlie the restrictions on coronal stops before laterals. That is, coarticulation effects make the burst and formant transitions of coronals insufficiently distinct from velars in this context. The lateral constrains the position of the tongue tip and body, so the formant transitions in coronal-lateral and velar-lateral clusters are very similar, while a labial is generally distinguished by lower F2 due to lip constriction (Kawasaki 1982:67ff., Olive, Greenwood, and Coleman 1993:284). The coronal and velar closures are at or behind the location of the lateral constriction, so in both cases frication noise is generated at this lateral constriction at release, resulting in acoustically similar bursts.

A more striking example of how distribution of contrasts differs depending on the nature of the cues involved comes from the comparison between major place contrasts and retroflexion contrasts (Steriade 1995, 2001). The contrast between retroflex and apical alveolar consonants is found in many Australian and Dravidian languages. It is commonly restricted to positions following a vowel, so it is neutralized word-initially and following consonants (Steriade 1995). This is in sharp distinction from most other place contrasts, which, as we have seen, occur preferentially *before* vowels. Steriade argues that this difference follows from differences in the distribution of cues to these types of contrasts. Retroflexes are distinguished from apical alveolars by a low third formant at closure (Stevens and Blumstein 1975). However, the tongue tip moves forward during the closure of a retroflex and is released at the alveolar ridge, so these sounds are articulatorily and acoustically very similar at release (Dave 1977, Butcher 1995, Anderson 1997, Spajic, Ladefoged and Bhaskararao 1994). Closure transitions are only available where the consonant is preceded by a vowel, hence this cue is missing in other environments, making the contrast less distinct (Anderson 1997). So the retroflexion contrast differs from other place contrasts in that it is realized most distinctly on a preceding vowel rather than a following vowel, but given this difference we can see that all place contrasts are liable to neutralize in environments where it would be difficult to make them distinct.

It should be noted that the patterns of distribution described for major place and obstruent voicing contrasts have often been analyzed as involving neutralization of contrasts in the coda of syllables (e.g. Itô 1989, Vennemann 1972). Steriade (1997, 1999) and Côté (2000) present detailed arguments in favour of the 'Licensing by Cue' analysis. In the present context, the important weaknesses of a coda-neutralization account are that it does not extend to retroflexion contrasts, which are neutralized in word-initial and post-consonantal onsets, but are permitted in codas and intervocalic onsets, and that it cannot relate the patterns of distribution to the nature of the features involved.

While the analyses sketched above indicate that considerations of distinctiveness play a central role in accounting for the distribution of contrasts, it is clear that other constraints are important also. For example, stop bursts will only be absent before obstruents if some constraint requires the stop closure to overlap with the following consonant. One general phenomenon that implicates additional constraints is word-final neutralization. For example, in German, obstruent voicing is neutralized preceding obstruents, and in word-final position. For words spoken in isolation, these are both environments in which VOT cues are unavailable, because there is no following sonorant, but in phrase-medial position, a word-final obstruent might be followed by a sonorant, allowing the realization of VOT differences. If contrast is governed strictly by the availability of cues, the voicing contrast should be permitted in this context, but in German, and many similar languages, voicing is neutralized in word-final position, regardless of phrasal context. So the analysis in terms of licensing by cue must be supplemented by additional constraints relating to morphosyntactic structure. Steriade (1997) analyzes this pattern as resulting from generalization of the citation form of words. That is, there is a general preference to give words a uniform pronunciation in all contexts, and this is modeled on the pronunciation of the word spoken in isolation. This analysis is formalized in terms of Output-Output Correspondence constraints (Benua 1997, Kenstowicz 1996, Steriade 2000). A comparable distinction between word-internal and cross-word sequences must be made in syllabification-based analyses in order to block syllabification of a word-final consonant as an onset to a following vowel-initial word.

### 3.1 Formalizing Licensing by Cue

Steriade (1997, 1999) formalizes the Licensing by Cue hypothesis in terms of constraints on the distinctiveness of contrasts. Although the specifics are rather different from dispersion theory, the general conception is very similar, so the same constraints motivated above in the analysis of enhancement can be used to analyze patterns of contextual neutralization (Flemming 2002:40ff.)[11].

#### *3.1.1 Distinctiveness constraints*

In the case of obstruent voicing, we will assume that there is a perceptual dimension corresponding to VOT, which takes a value of 0 for voiced and 1 for voiceless

[11] Steriade proposes constraints of the form *□voice/C that penalize obstruent voicing contrasts in a particular context, C. These constraints are ranked according to the richness of cues to voicing available in that context. These constraints are replaced here by MINDIST constraints that refer directly to the cues that differentiate contrasting obstruents. This allows for variability in the cues realized in a given context, depending on the production strategy adopted (cf. Steriade 1997, Koontz-Garboden 2002).

obstruents[12]. Languages that restrict voicing contrasts to pre-sonorant positions require a VOT difference for the contrast to be adequately distinct. In other words, MINDIST=VOT:1 ranks above MAXIMIZE CONTRASTS. This sets a threshold for minimum distinctiveness that can be met in pre-sonorant position, so a voicing contrast is permitted in that environment (14).

| (14) | _V | MINDIST = VOT:1 | MAXIMIZE CONTRASTS | *[+voice, -son] |
|---|---|---|---|---|
| a. | ☞ dV-tV | | ✓✓ | * |
| b. | dV | | ✓! | * |
| c. | tV | | ✓! | |

Pre-pausally, no VOT difference is possible, and a difference in closure voicing alone is insufficient, so a voicing contrast is not permitted (15, # indicates a word boundary). Given that there is no contrast, obstruents are realized with the least-effort laryngeal state. In pre-pausal position, this is voiceless, given the difficulties involved in maintaining vocal fold vibration during an obstruent (cf. §2.3). This preference is formalized as a constraint against voiced obstruents, *[+voice, -sonorant].

| (15) | V_# | MINDIST = VOT:1 | MAXIMIZE CONTRASTS | *[+voice, -son] |
|---|---|---|---|---|
| a. | Vd#-Vt# | *! | ✓✓ | * |
| b. | Vd# | | ✓ | *! |
| c. | ☞ Vt# | | ✓ | |

Voicing contrasts are also neutralized before obstruents, because VOT differences cannot be realized in this position either (17). However, in this case, the neutralized stop is voiced, assimilating to the following obstruent (in this case [g]). A plausible analysis of this pattern is that it is especially difficult to initiate voicing during an obstruent – due to hysteresis effects it is easier to maintain voicing from a sonorant into a following

[12] In fact there are two basic types of obstruent 'voicing' contrasts: fully voiced vs. voiceless unaspirated, and voiceless unaspirated vs. aspirated, so a more general analysis requires three levels of VOT (Flemming 2002).

obstruent than it is to initiate voicing during an obstruent following a voiceless sound (Westbury and Keating 1986). Thus we can posit the constraint in (16), named *TD for brevity, universally ranked above *[+voice, -son].

(16) *TD: *[-voice][+voice, -sonorant]

This constraint forces assimilation in obstruent sequences, as shown in the following tableau:

| (17) | V_gV | MINDIST = VOT:1 | MAXIMIZE CONTRASTS | *TD | *[+voice, -son] |
|---|---|---|---|---|---|
| a. | VdgV-VtgV | *! | ✓✓ | * | * |
| b. | ☞ VdgV | | ✓ | | *! |
| c. | VtgV | | ✓ | *! | |

Languages with broader distribution of obstruent voicing rank MAXIMIZE CONTRASTS above MINDIST=VOT:1, tolerating less distinct contrasts in order to realize more contrasts. But no language will prefer less distinct contrasts over more distinct contrasts of a similar type.

Neutralization of retroflexion is analyzed along similar lines: a MINDIST constraint requiring an F3 difference is ranked above MAXIMIZE CONTRASTS, so the contrast is neutralized where it is not possible to realize this cue.

*3.1.2 Faithfulness constraints*

An alternative approach to formulating the perceptual constraints that account for these generalizations about the distribution of contrasts makes use of faithfulness constraints (Steriade 1995, 2001, Jun 1995, Boersma 1998). This is a natural move since faithfulness constraints play a central role in the regulation of contrasts in standard OT. Essentially, a faithfulness constraint like IDENT F, where F is a feature, favours preserving underlying differences – if the input contains [+F], the output should contain [+F], if the input contains [-F], the output should contain [-F]. So if IDENT F is satisfied, an underlying difference between [+F] and [-F] is preserved on the surface, and the language has a contrast in F.

Perceptual factors are introduced by distinguishing IDENT F constraints for different contexts, then ranking them according to the distinctiveness of an F contrast in that context. For example, we might posit the ranking of IDENT[voice] constraints in (18).

(18) IDENT[voice]/ _ [+son] >> IDENT[voice]/ _ # >> IDENT[voice]/ _ [-son]

The distribution of voicing contrasts is then determined by the position of a constraint against voiced obstruents, *[+voice, -son]. For example, the ranking in (19) derives neutralization everywhere except before sonorants (the German pattern)[13]. If *[+voice, -son] is ranked lower, then the contrast is permitted in more positions, but again contrasts are permitted first in more distinct environments. These constraints predict that neutralization always yields voiceless obstruents, so an additional constraint, such as *TD, is required to derive assimilation to following obstruents.

(19) IDENT[vce]/ _ [+son] >> ***[+voice, -son]** >> IDENT[voi]/ _ # >> IDENT[voi]/ _ [-son]

This approach works elegantly in simple cases, but it has some limitations that make it incapable of providing a comprehensive account of perceptual effects. The fundamental limitation of faithfulness constraints is that they can only block change between input and output, they cannot motivate change. This is problematic because there are various phenomena that have been argued to be perceptually-motivated which crucially involve unfaithfulness to input forms, including the dispersion phenomena discussed in §2. For example, a language with only the peripheral vowels [i, e, a, o, u] must unfaithfully map non-peripheral input vowels such as [y, ɯ] onto one of these vowels. Ranking IDENT[round] low in the constraint hierarchy, for example, makes it relatively acceptable

---

[13] This analysis is structurally very similar to the one proposed in Lombardi (1995), but Lombardi employs a faithfulness constraint specific to pre-sonorant onsets.

to realize [y, ɯ] as [i] and [u] respectively, but it does not favour these realizations. Unfaithful mappings can only be motivated by markedness constraints, and as we have seen above, the markedness constraints that best account for this pattern are distinctiveness constraints implementing a preference for maximally distinct F2 contrasts. The same applies to other cases of dispersion and enhancement. For example, enhancement of voicing contrasts by pre-nasalizing voiced stops (§2.3) implies unfaithful realization of input voiced stops as pre-nasalized stops, which must be motivated by a markedness constraint.

More generally, perceptually-ranked featural faithfulness constraints can only account for patterns of neutralization, but arguably neutralization is just one way of avoiding an otherwise indistinct contrast. That is, an indistinct contrast may be avoided by giving up the contrast (neutralization), or by making the contrast more distinct (enhancement). We have seen that distinctiveness constraints can be used to derive both patterns, but perceptually-ranked faithfulness constraints can only derive neutralization.

This limitation applies not only to the analysis of segment-internal enhancements of the kind just discussed, but also to modification of the environment of a contrast (cf. Côté 2000:175f., Hume and Johnson 2001:8f.). For example, it has been suggested that vowel epenthesis is often motivated by the need to make consonant contrasts more distinct (e.g. Wright 1996:40, Côté 2000). One such pattern is epenthesis into clusters of three consonants, exemplified from Yawelmani Yokuts (Newman 1944, Kisseberth 1970) in (20). Similar patterns are observed in Cairene Arabic (Broselow 1976) and Lenakel (Lynch 1978).

| | | | | |
|---|---|---|---|---|
| (20) | /paʔt+mi/ | → | [paʔitmi] | ‘having fought’ |
| | cf. /paʔt+al/ | → | [paʔtal] | ‘might fight’ |
| | /lihm+mi/ | → | [lihimmi] | ‘having run’ |
| | cf. /lihm+al/ | → | [lihmal] | ‘might run’ |

Côté (2000) analyzes this pattern in terms of the markedness constraint C↔V: 'A consonant is adjacent to a vowel' – that is, epenthesis applies to ensure that every consonant is adjacent to a vowel, which is not the case in a triconsonantal cluster. Formally, epenthesis is derived by ranking C↔V above DEPV, the faithfulness constraint that is violated by inserting a vowel[14] (see Kager 1998:107ff. for a similar analysis based on syllabification constraints).

As Côté argues, it is perceptually desirable for consonants to be adjacent to a vowel because many consonantal contrasts are best realized in this position. As noted above, formant transitions are important place cues that are best realized on a vowel. The contrast between presence and absence of a consonant is also more distinct adjacent to a vowel because the change in constriction between consonant and vowel results in salient spectral discontinuities (Liu 1996, Ohala 1980, Stevens 1998:245f.). The nature of the spectral change, e.g. the rate and magnitude of change in different frequency bands, may also provide cues to consonant manner (Stevens 1985, Liu 1996)[15].

This analysis cannot be implemented in terms of perceptually-ranked faithfulness constraints. Ranking constraints against consonant deletion (MAXC) according to the strength of the cues to the presence of a consonant can only allow deletion of poorly-cued consonants, it cannot motivate epenthesis to improve the cues to a consonant. The unfaithful insertion of a vowel can only be motivated by a markedness constraint violated by triconsonantal clusters, such as C↔V.

Perceptually-ranked faithfulness constraints favour perceptually minimal changes between input and output. This arrangement predicts that indistinct contrasts are more likely to be lost because they can be neutralized by perceptually minimal changes, but it

[14] DEPV must also be outranked by MAXC, the offending consonant is deleted rather than being rescued by vowel epenthesis. Deletion in triconsonantal clusters is observed in a number of languages, e.g. Korean (Kim and Shibatani 1976).

[15] It should be noted that these kinds of considerations properly motivate constraints requiring consonant *contrasts* to be realized adjacent to vowels (i.e. distinctiveness constraints), as discussed at length in §2.3. The limitations of basic markedness constraints are addressed further in §3.1.3.

does not implement a general preference for distinct contrasts, and so cannot account for the observed range of perceptually-motivated phenomena. However, there is evidence that perceptually minimal change between input and output is preferred in alternations (i.e. contextual variation in the realization of morphemes) (Steriade forthcoming), so perceptual ranking of faithfulness constraints may be motivated on independent grounds.

### *3.1.3 Sound change via misperception*

The limitations of perceptually-ranked faithfulness constraints are shared by some theories that locate perceptual constraints in the process of sound change rather than in synchronic grammars (e.g. Blevins and Garrett 1998, Ohala 1990). According to these accounts indistinct contrasts appear to be dispreferred in languages because they are more likely to be lost over time through misperception on the part of language users. For example, Ohala (1990) argues that consonants often assimilate in place to a following consonant (e.g. anka > aŋka) because the unassimilated cluster is easily misperceived as the assimilated cluster. This is related to the observation above that post-vocalic major place contrasts are relatively indistinct, but according to Ohala this pattern results from '"innocent" misapprehension' on the part of listeners, so no dispreference for indistinct contrasts needs to be encoded in grammars.

Sound change through misperception, like perceptually-ranked faithfulness constraints, can only hope to account for neutralization, not dispersion or enhancement. For example, at least some cases in which stop voicing contrasts are enhanced by prenasalization of voiced stops (§2.3) seem to have arisen via a sound change from earlier voiced stops to prenasalized stops (Herbert 1986:16ff.). This change cannot be attributed to misperception, rather prenasalization seems to be a strategy that speakers have hit upon to make stop voicing contrasts more distinct, so a preference for distinct contrasts is necessary to account for this pattern. In general, a mechanism of sound change via misperception only predicts that less distinct contrasts are more likely to be lost, it cannot account for cases in which speakers appear to take measures to increase the distinctiveness of contrasts – i.e. dispersion and enhancement phenomena (cf. Steriade 2001:233ff. for a similar argument).

Relating sound change directly to patterns of misperception also incorrectly predicts some unattested sound changes. For example, a study of vowel confusions in French (Robert-Ribes et al 1998) found that [i] is confused with [y] much more frequently than it is confused with [u]. This difference in distinctiveness is expected, and is hypothesized to lie behind the cross-linguistic preference for contrasts like [i-u] over front rounding contrasts like [i-y]. An 'innocent misapprehension' model might attribute this preference to the greater tendency for [i-y] contrasts to be lost through misperception. However, the study found that [i] is misidentified as [y] at about the same rate as the converse misidentification of [y] as [i]. So if sound changes arise from misperception, we would expect a change [i] > [y] to be as likely as [y] > [i], but while the latter change is well attested (e.g. in Old English (Lass and Anderson 1975:286ff.) and Greek (Newton 1972:19)), unconditioned rounding of front vowels is unattested. Significantly, unrounding front vowels increases the distinctiveness of front-back contrasts, while the unattested change would reduce distinctiveness without any compensatory benefits.

### *3.1.4 Basic markedness constraints*

A final approach to formalizing the perceptual constraints responsible for contextual neutralization is to use basic markedness constraints. For example, in the analysis of voicing neutralization reviewed above, Steriade (1997) proposes a distinctiveness constraint against obstruent voicing contrasts appearing where there is no following sonorant. The closest equivalent basic markedness constraint would be a constraint requiring voiced obstruents to be followed by sonorants (cf. Lombardi 1995, 1999). Constraints of this kind are widely used in the analysis of contextual neutralization (McCarthy 2002:87), but usually without appealing to any perceptual motivation.

However, some researchers have used basic markedness constraints to formalize perceptually motivated constraints (e.g. Côté 2000, Hume 1998).

We saw in section 2.3 that basic markedness constraints are inadequate for the analysis of dispersion effects, and are difficult to motivate on perceptual grounds because perceptual considerations disfavour indistinct contrasts, not individual sounds. Similar difficulties face the use of basic markedness constraints in the analysis of Licensing by Cue effects. A basic constraint on obstruent voicing must ban [+voice] or [-voice] rather than the contrast between them. This is not only perceptually unmotivated, it leads to empirical difficulties. For example, it is common for the result of neutralization to be phonetically distinct from either of the sounds that occurs in positions of contrast (cf. Trubetzkoy 1939:71-3). This is the case in the neutralization of retroflexion contrasts, for example. Butcher (1995) studied several Australian languages that contrast retroflex and apical alveolar consonants, and found that neutralization of this contrast in word-initial position yields an intermediate consonant, generally postalveolar (unlike apical alveolars), but apical rather than sub-laminal (the contrastive retroflexes are sub-laminal). This intermediate status is reflected in uncertainty among Australianists as to the appropriate transcription for these sounds (Butcher 1995, Steriade 1995). If the distinction between retroflexes and apical alveolars is treated as binary (e.g. [+/-anterior]), then formulating a constraint against either retroflexes or apical alveolars in word initial position predicts that the other articulation should be favored in neutralization, which is not accurate, since an intermediate articulation is actually observed. If we make a three-way distinction between apical alveolars, retroflexes, and an intermediate articulation, then it is possible to formulate constraints against either extreme appearing in word-initial position, but it would also be necessary to prevent the intermediate place from surfacing in environments of contrast. These problems are avoided if we recognize that it is the *contrast* between retroflexes and apical alveolars that is problematic in word-initial position. In the absence of contrast, the intermediate articulation is preferred as less effortful than a sub-laminal retroflex, but more distinct from laminal coronals than an apical alveolar.

**4. Conclusions**

The evidence reviewed here leads to the conclusions that (i) speech perception does play a role in shaping phonological patterns, and (ii) the relevant constraints are constraints on the distinctiveness of contrasts.

We have examined two types of phonological patterns that have been related to the perceptual properties of speech sounds: dispersion/enhancement and contextual neutralization. Both phenomena can be analyzed in terms of a preference for more distinct contrasts and the converse dispreference for indistinct contrasts. Dispersion of contrastive sounds in perceptual space is a direct consequence of maximization of distinctiveness, while enhancement phenomena simply reflect the fact that greater distinctiveness is often achieved by covarying physiologically unrelated articulations such as tongue body backness and lip rounding. Contextual neutralization also follows from the preference for distinct contrasts given the fact that the distinctiveness of a contrast type varies according to context. For example, obstruent voicing contrasts are more distinct before a sonorant than in other environments, so some languages only allow voicing contrasts before sonorants, neutralizing the contrast elsewhere.

So the two patterns are fundamentally similar: a language with front unrounded and back rounded vowels avoids the less distinct contrasts between front rounded and back rounded vowels, and a language that only allows obstruent voicing contrasts before sonorants avoids the less distinct contrasts involving obstruent voicing in other contexts. Alternative analyses in terms of basic markedness constraints, perceptually-ranked faithfulness constraints or sound change through misperception cannot provide adequate accounts of the full range of perceptually-based phonological phenomena.

## 5. References


Anderson, Victoria (1997). The perception of coronals in Western Arrernte. *Proceedings of Eurospeech '97: Fifth Conference on Speech Communication and Technology* 1, 389-392.

Ashby, F. Gregory, Geoffrey Boynton, and W. William Lee (1994). Categorization response time with multidimensional stimuli. *Perception and Psychophysics* 55, 11-27.

Bender, Byron W. (1968). Marshallese phonology. *Oceanic Linguistics* 7, 16-35.

Benua, Laura (1997). *Transderivational Identity*. Ph.D. dissertation, University of Massachusetts, Amherst.

Blevins, Juliette, and Andrew Garrett (1998). The origins of consonant-vowel metathesis. *Language* 74, 508-556.

Boersma, Paul (1998). *Functional Phonology*. PhD dissertation, University of Amsterdam.

Booij, Geert (1995). *The Phonology of Dutch*. Oxford University Press, Oxford.

Broselow, Ellen I. (1976). *The Phonology of Egyptian Arabic*. PhD dissertation, University of Massachusetts, Amherst.

Butcher, Andrew (1995). The phonetics of neutralization: the case of Australian coronals. Jack Windsor Lewis (ed.) *Studies in General and English Phonetics: Essays in Honour of Professor J.D. O'Connor*. Routledge, New York, 10-38.

Calabrese, Andrea (1988). *Towards a Theory of Phonological Alphabets*. Ph.D. dissertation, MIT.

Choi, John D. (1992). *Phonetic Underspecification and Target Interpolation: An Acoustic Study of Marshallese Vowel Allophony (UCLA Working Papers in Phonetics 82)*. Ph.D. dissertation, University of California, Los Angeles.

Colarusso, John (1988). *The Northwest Caucasian Languages: A Phonological Survey*. Garland Publishing, New York.

Colarusso, John (1992). *A Grammar of the Kabardian Language*. University of Calgary Press, Calgary.

Côté, Marie-Hélène (2000). *Consonant Cluster Phonotactics: A Perceptual Approach*. Ph.D. dissertation, MIT [distributed as *MIT Working Papers in Linguistics and Philosophy*]

Dave, Radhekant (1977). Retroflex and dental consonants in Gujarati. A palatographic and acoustic study. *Annual Report of the Institute of Phonetics, University of Copenhagen* 11: 27-156.

Delattre, Pierre C., Alvin M. Liberman, Franklin S. Cooper, and Louis J. Gerstman (1952). An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word* 8, 195-210.

Delgutte, Bertrand, and Nelson Y.S. Kiang (1984). Speech coding in the auditory nerve IV: Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America* 75, 897-907.

Diehl, Randy L. (1991). The role of phonetics in the study of language. *Phonetica*, 48, 120-134.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. , *22*, 109-122.

Flash, Tamar, and Neville Hogan (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience* 5, 1688-1703.

Flemming, Edward (1996). Evidence for constraints on contrast: The dispersion theory of contrast. Chai-Shune K. Hsu (ed.) *UCLA Working Papers in Phonology* 1, 86-106.

Flemming, Edward (2001). Contrast and perceptual distinctiveness. To appear in Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.) *The Phonetic Bases of Markedness*. Cambridge University Press, Cambridge.

Flemming, Edward (2002). *Auditory Representations in Phonology*. Routledge, New York.

Fujimura, Osamu, M. Macchi, and L.A. Streeter (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21, 337-346.

Greenberg, Steven (1995). Auditory processing of speech. Norman J. Lass (ed.) *Principles of experimental phonetics*. Mosby, St. Louis.

Gregores, Emma, and Jorge A. Suárez (1967). *A Description of Colloquial Guaraní*. Moton, The Hague.

Hamilton, Philip (1996). *Constraints and markedness in the phonotactics of Australian aboriginal language*. Ph.D. dissertation, University of Toronto.

Harris, Christopher M., and Daniel M. Wolpert (1998). Signal-dependent noise determines motor planning. *Nature* 394, 780-784.

Herbert, Robert K. (1986). *Language Universals, Markedness Theory and Natural Phonetic Processes*. Mouton de Gruyter, Berlin.

Hudson, Joyce, and Eirlys Richards (1969). The phonology of Walmatjari. *Oceanic Linguistics* 8, 171-189.

Hume, Elizabeth (1998). The role of perceptibility in consonant/consonant metathesis. *Proceedings of the West Coast Conference on Formal Linguistics* 17, 293-307.

Hume, Elizabeth, and Keith Johnson (2001). A model of the interplay of speech perception and phonology. Elizabeth Hume and Keith Johnson (eds.) *The Role of Speech Perception in Phonology*. Academic Press, New York.

Itô, Junko (1989). A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7, 217-259.

Iverson, Gregory K., and Joseph C. Salmons (1996). Mixtec prenasalization as hypervoicing. *International Journal of American Linguistics* 62, 165-175.

Jessen, Michael (1998). Phonetics and phonology of tense and lax obstruents in German. John Benjamins, Amsterdam.

Jun, Jongho (1995). *Perceptual and Articulatory Factors in Place Assimilation: An Optimality-Theoretic Approach*. PhD dissertation, UCLA.

Kager, René (1998). *Optimality Theory*. Cambridge University Press, Cambridge.

Kawasaki, Haruko (1982). *An acoustical basis for universal constraints on sound sequences*. Ph.D. dissertation, University of California, Berkeley.

Kellogg, W.N. (1931). The time of judgment in psychometric measures. *The American Journal of Psychology* 43, 65-86.

Kenstowicz, Michael (1996). Uniform exponence and base identity. J. Durand and B. Laks (eds.) *Current Trends in Phonology*. CNRS and University of Salford Publications, 363-393.
Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *J Acoust Soc America, 73(5)*, 1779-1793.
Kim, Kong-On, and Masayoshi Shibatani(1976). Syllabification phenomena in Korean. *Language Research* 12, 91-98.
Kingston, John, and Randy L. Diehl (1994). Phonetic knowledge. *Language* 70, 419-454.
Kirchner, Robert (1998). *An Effort-Based Approach to Consonant Lenition*. PhD dissertation, UCLA.
Kisseberth, Charles (1970). On the functional unity of phonological rules. *Linguistic Inquiry* 1, 291-306.
Kondo, Yuko (1994). Targetless schwa: is that how we get the impression of stress timing in English? *Proceedings of the Edinburgh Linguistics Department Conference '94*, 63-76.
Koontz-Garboden, Andrew (2002). Voicing contrasts in medial stop clusters. Ms., Stanford.
Kuipers, Aert H. (1960). *Phoneme and Morpheme in Kabardian (Eastern Adyghe)*. Janua Linguarum, series minor, no. 8. Mouton, The Hague.
Ladefoged, Peter, and Ian Maddieson (1996). *The Sounds of the World's Languages*. Blackwell, Oxford.
Lass, Roger, and John M. Anderson (1975). *Old English Phonology*. Cambridge University Press, Cambridge.
Liljencrants, Johan, and Björn Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48, 839-62.
Lindblom, Björn (1983). Economy of speech gestures. Peter. F. MacNeilage (ed.) *The Production of Speech*. Springer-Verlag, New York, 217-246.
Lindblom, Björn (1986). Phonetic universals in vowel systems. J.J. Ohala and J.J. Jaeger (eds.) *Experimental Phonology*. Academic Press.
Lindblom, Björn, and Olle Engstrand (1989). In what sense is speech quantal? *Journal of Phonetics* 17, 107-121.
Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563-567). Prague: Academia.
Liu, Sharlene (1996). Landmark detection for distinctive feature-based speech recognition. *Journal of the Acoustical Society of America* 100, 3417-3430.
Lombardi, Linda (1995). Laryngeal neutralization and syllable wellformedness. *Natural Language and Linguistic Theory* 13, 39-74.
Lombardi, Linda (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17, 267-302.
Lynch, John (1978). *A Grammar of Lenakel*. Pacific Linguistics, Series B, no. 55. Asutralian National University, Canberra.
Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge University Press, Cambridge.
Maiden, Martin (1995). Vowel systems. Martin Maiden and Mair Parry (eds.) *The Dialects of Italy*. Routledge, London, 7-14.

McCarthy, John J. (2002). *A Thematic Guide to Optimality Theory*. Cambridge University Press, Cambridge.
McGregor, William (1990). *A functional grammar of Gooniyandi*. John Benjamins, Amsterdam.
Mithun, Marianne, and Hasan Basri (1986). The phonology of Selayarese. *Oceanic Linguistics* 25, 210-254.
Nelson, W.L. (1983). Physical principles for economies of skilled movement. *Biological Cybernetics* 46, 135-147.
Newman, Stanley (1944). *The Yokuts Language of California*. The Viking Fund Publications in Anthropology 2, The Viking Fund, New York.
Newton, Brian (1972). *The generative interpretation of dialect: A study of Modern Greek phonology*. Cambridge University Press, Cambridge.
Ní Chiosáin, Maire, and Jaye Padgett (2001). Markedness, segment realization, and locality in spreading. Linda Lombardi (ed.) *Segmental Phonology in Optimality Theory*, Cambridge University Press, Cambridge, 118-156.
Ohala, John J. (1980). The application of phonological universals in speech pathology. Norman Lass (ed.) *Speech and language: advances in basic research and practice,* vol. 3, 75-97. Academic Press, New York.
Ohala, John J. (1981). The listener as a source of sound change. *Papers from the parasession on language and behavior*, Chicago Linguistics Society, 178–203.
Ohala, John J. (1985). Around *Flat*. Victoria Fromkin (ed.) *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Academic Press, New York, 223-242.
Ohala, John J. (1990). The phonetics and phonology of aspects of assimilation. John Kingston and Mary E. Beckman (eds.) *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge University Press, Cambridge, 258–275.
Ohala, John J. (1993). The perceptual basis of some sound patterns. D. A. Connell and A. Arvaniti (eds.) *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge University Press, Cambridge, 87-94.
Olive, Joseph P., Alice Greenwood, and John Coleman (1993). *Acoustics of American English speech: a dynamic approach*. Springer-Verlag, New York.
Perkell, Joseph S. (1997). Articulatory processes. William J. Hardcastle and John Laver (eds.) *The Handbook of Phonetic Sciences*. Blackwell, Oxford, 333-370.
Pisoni, David B., and Jeffrey Tash (1974). Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics* 15, 285-290.
Plomp, Reinier (1975). Auditory analysis and timbre perception. Gunnar Fant and Michel Tatham (eds.) *Auditory Analysis and Perception of Speech.* Academic Press, New York, 7-22.
Podgorny, Peter, and W.R. Garner (1979). Reaction time as a measure of onter- and intraobject visual similarity: Letters of the alphabet. *Perception and Psychophysics* 26, 37-52.
Prince, Alan, and Paul Smolensky. (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. To appear: Cambridge, MA: MIT Press.
Puppel, Stanislaw, Jadwiga Nawrocka-Fisiak, and Halina Krasskowska (1977). *A Handbook of Polish Pronunciation for English Learners*. Panstwowe Wydawnictwo Naukowe, Warsaw.

Redford, M.A. & Diehl, R.L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America*, 106, 1555-1565

Robert-Ribes, Jordi, Jean-Luc Schwartz, Tahar Lallouache, and Pierre Escudier (1998). Complementarity and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise. *Journal of the Acoustical Society of America* 103, 3677-3689.

Shadle, Christine (1991). The effects of geometry on the source mechanisms of consonants. *Journal of Phonetics* 19, 409-424.

Shepard, Roger N. (1972). Psychological representation of speech sounds. Edward David and Peter Denes (eds.) *Human Communication: A Unified View*. McGraw-Hill, New York, 67-113.

Smeets, Rieks (1984). *Studies in West Circassian Phonology and Morphology*. Hakuchi Press, Leiden.

Smith, Richard, and Connie Smith (1971). Southern Barasano phonemics. *Linguistics* 75, 80-85.

Spajić, Siniša, Peter Ladefoged, and Peri Bhaskararao (1994). The rhotics of Toda. *UCLA Working Papers in Phonetics* 87, 35-66.

Steriade, Donca (1995). Neutralization and the expression of contrast. Ms, UCLA.

Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. *UCLA Working Papers in Linguistics* 3, 25-146.

Steriade, Donca (1999). Alternatives to syllable-based accounts of consonantal phonotactics. Osamu Fujimura, Brian D. Joseph, and Bohumil Palek (eds.) *Proceedings of LP '98: Item Order in Language and Speech*. Karolinum Press, Prague, 205-245.

Steriade, Donca (2000). Paradigm uniformity and the phonetics-phonology boundary. Michael Broe and Janet Pierrehumbert (eds.) *Papers in Laboratory Phonology 5: Acquisition and the Lexicon*. Cambridge University Press, Cambridge, 313-334.

Steriade, Donca (2001). Directional asymmetries in place assimilation. Elizabeth Hume and Keith Johnson (eds.) *The Role of Speech Perception in Phonology*. Academic Press, New York.

Steriade, Donca (forthcoming). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. To appear in Kristin Hanson and Sharon Inkelas (eds) *The nature of the word: essays in honor of Paul Kiparsky*. MIT Press, Cambridge.

Stevens, Kenneth N. (1985). Evidence for the role of acoustic boundaries in the perception of speech sounds. Victoria A. Fromkin (ed.) *Phonetic linguistics: essays on honor of Peter Ladefoged*. Academic Press, New York, 243-255.

Stevens, Kenneth N. (1998). *Acoustic Phonetics*. MIT Press, Cambridge.

Stevens, Kenneth N., and Sheila E. Blumstein (1975). Quantal aspects of consonant production and perception: a study of retroflex consonants. *Journal of Phonetics* 3, 215-233.

Stevens, Kenneth N., and Sheila E. Blumstein (1981). The search for invariant acoustic correlates of phonetic features. Peter D. Eimas and Joanne L. Miller (eds.) *Perspectives on the study of speech*. Lawrence Erlbaum, Hillsdale, 1-38.

Stevens, Kenneth N., S. Jay Keyser, and Haruko Kawasaki (1986). Toward a phonetic and phonological theory of redundant features. Joseph S. Perkell and Dennis H. Klatt

(eds) *Invariance and Variability in Speech Processes*. Lawrence Erlbaum, Hillsdale, 426-449.
Trubetzkoy, Nikolai S. (1939). *Grundzüge de Phonologie*. Travaux du Cercle Linguistique de Prague 7.
Tucker, A.N. (1962). The syllable in Luganda: a prosodic approach. *Journal of African Languages* 1, 122-166.
Uno Y., M. Kawato and R. Suzuki (1989). Formation and control of optimal trajectories in human multijoint arm movement: Minimum torque-change model. *Biological Cybernetics* 61, 89-101.
Van Bergem, Dick R. (1994). A model of coarticulatory effects on the schwa. *Speech Communication* 14, 143-162.
Vennemann, Theo (1972). Zur Silbenstruktut der deutchen Standardsprache. Theo Venneman (ed.) *Silben, Segmente, Akzente*. Niemeyer, Tübingen.
Westbury, John, and Patricia A. Keating (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics* 22, 145-66.
Wright, Richard (1996). *Consonant clusters and cue preservation in Tsou*. PhD dissertation, UCLA.
Wright, Richard (2001). Perceptual cues in contrast maintenance. Elizabeth Hume and Keith Johnson (eds.) *The Role of Speech Perception in Phonology*. Academic Press, New York, 251-277.