Phonology **18** (2001) 7–44. Printed in the United Kingdom

Scalar and categorical phenomena in a unified model of phonetics and phonology[*]

Edward Flemming
Stanford University

It is often assumed that there is a sharp division between phonetic and phonological processes, but the two are often strikingly similar, as in the case of phonetic consonant–vowel coarticulation and phonological assimilation between consonants and vowels. Parallels of this kind are best accounted for if both types of phenomena are analysed within a unified framework, so similarities result from the fact that both phonetic and phonological processes are subject to the same constraints. A unified model of phonetics and phonology is developed and exemplified through the analysis of parallel phonetic and phonological assimilation processes. The model operates in terms of scalar phonetic representations to accommodate phonetic detail, but categorical phenomena can still be derived from the interaction of speech production constraints with constraints that motivate the formation of distinct categories of sounds for the purposes of linguistic contrast.

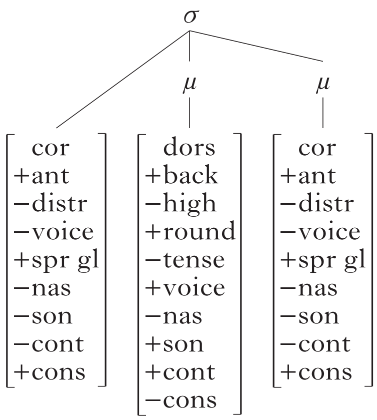
1 Introduction

Standard phonological representations provide only a very coarse-grained specification of the phonetic properties of a word or phrase. That is, representation in terms of standard phonological features and timing units corresponds approximately to the level of detail contain in a broad phonetic transcription, and the representation of time is in many ways even more limited. For example, consider the representation in (1) of the word ‘taught’ [tʰɔt].

---

[*] Many thanks to audiences at HOT ’97, SWOT ’97, the GLOW workshop on phonetics in phonology and the Stanford Phonology Workshop for comments on various portions of this work, and to Carlos Gussenhoven and two anonymous reviewers for invaluable feedback on the form and substance of this paper.

(1)



This representation does not tell us the precise duration of any of the segments, the duration of aspiration, the nature of the movement of the articulators from one segment to the next, the fact that /ɔ/ in this context is fronted compared to its realisation in /ɔ/ *awe*, the fact that the velum is slightly lowered (a characteristic of low vowels; Ohala 1971), etc.

It has often been assumed that many of these details of phonetic realisation are a consequence of universal principles. For example, Chomsky & Halle (1968: 295) state that 'phonetic transcriptions [i.e. the output of the grammar – EF] omit properties of the signal that are supplied by universal rules. These properties include…the transition between a vowel and an adjacent consonant, the adjustments in the vocal tract shape made in anticipation of subsequent motions, etc.' However this assumption turns out to be excessively optimistic – most of these aspects of phonetic realisation are subject to language-specific variation (Keating 1985).[1]

As discussed by Keating (1985), this finding presents us with the following dilemma. Since standard phonological representations omit much language-specific phonetic detail, we must either enrich phonological representations to include these details or posit an additional, language-specific phonetic component of grammar which supplies these details. Most phonologists who address this issue appear to have adopted this second option – the more conservative option from the point of view of phonology, since it allows the retention of standard, coarse-grained representations. This is also the model adopted by Keating herself (e.g. Keating 1984, 1990). That is, the grammar of sound is hypothesised to be divided into two components, phonetics and phonology, which are assumed to operate in terms of very different representations and rules or constraints.[2]

---

[1] For evidence of language-specific variation in the coarticulatory phenomena referred to by Chomsky & Halle, see for example Magen (1984), Manuel & Krakow (1984), Keating & Cohn (1988), Manuel (1990) and Flemming (1997).

[2] Further examples of work assuming a strict division between phonetics and phonology can be found in *Journal of Phonetics* **18** (1990) (a special issue on the

The differences between phonetics and phonology in these divided models can be illustrated from Pierrehumbert's (1980) analysis of English intonation. In this model, intonational melodies are represented phonologically as strings of High and Low tones, each marked as a pitch accent, phrase accent or boundary tone. These tones are subject to the usual kinds of phonological restrictions on distribution (eg. a phrase accent must be preceded by a pitch accent (1980: 13)) and rules (e.g. associating tones to syllables and phrase edges). The phonetic representation, on the other hand, is a fundamental frequency contour – i.e. F0 (in Hertz) as a function of time (in seconds). Tones are mapped into F0 targets by context-sensitive rules which involve arithmetic operations on scalar variables. For example, in a sequence of two High tones, the pitch (in Hz) of the second is scaled to the pitch of the first according to the relative prominence of the two tones, where prominence is a scalar value assigned to each accent (1980: 79). The complete F0 contour is then derived by interpolation between targets.

Much subsequent work adopts this basic outline of the nature of the phonetic component (e.g. Pierrehumbert & Beckman 1988, Cohn 1993, Keating 1990; cf. Zsiga 1997 for a slightly different model), although for segmental phenomena many more parameters are required in the phonetic representation (e.g. formant frequencies or articulator positions).

It is generally assumed that phonological principles do not apply in the phonetic component, so, for example, McCarthy (1986: 250ff) dismisses several apparent counterexamples to the Obligatory Contour Principle on the grounds that they are phonetic processes, and thus not expected to be subject to this phonological principle. Other examples of this line of reasoning are found in Padgett (1995: 47) and Steriade (1993a: 345).

In this paper I re-examine the case for distinguishing language-specific phonetics from phonology, and conclude that this move is unmotivated. It is feasible to account for phonetic and phonological phenomena within a unified framework, and such a model is better able to account for the many similarities between phonetics and phonology.

It is appropriate to distinguish components of grammar where the representations and principles operative in each component are fundamentally distinct, thus it is uncontroversial to distinguish phonology from syntax. It is difficult to justify a distinction between phonetics and phonology on these grounds. Phonetics and phonology are not obviously distinguished by the nature of the representations involved, or in terms of the phenomena they encompass. As far as representation is concerned, most of the primitives of phonological representation remain phonetically based, in the sense that features and timing units are provided with broadly phonetic definitions. This has the peculiar consequence that

---

phonetics–phonology interface), and in the *Papers in laboratory phonology* series: Kingston & Beckman (1990), Docherty & Ladd (1992), Keating (1994), Connell & Arvaniti (1995) and Broe & Pierrehumbert (2000).

sound is represented twice in grammar, once at a coarse level of detail in the phonology and then again at a finer grain in the phonetics. Perhaps more significant is the fact that there are also substantial similarities between many phenomena which are conventionally classified as phonetic and those which are conventionally classified as phonological; for example, coarticulation is similar in many respects to assimilation.

The aim of this paper is to explore the idea that these parallels are best accounted for by analysing both 'phonetic' and 'phonological' phenomena within a unified framework, so the similar properties of the two can be derived from the same constraints. Unifying phonetics and phonology does not imply a denial of the distinction between scalar and categorical phenomena. Rather, the proposal is to derive phonological categories using scalar phonetic representations (cf. Lindblom 1986, Kirchner 1997). This allows categorical and scalar phenomena (e.g. neutralising assimilation and coarticulation) to be derived within a single component, so the same constraints can apply to both, giving rise to the observed parallels between the two.

The organisation of the paper is as follows. Examples of parallels between phonetic and phonological phenomena are presented in §2. Then the next two sections outline a unified model of phonetics and phonology designed to account for the existence of such parallels. §3 describes a constraint-based approach to phonetics, exemplified by an analysis of aspects of consonant–vowel coarticulation. The analysis of such phenomena obviously requires phonetically detailed representations, so if phonetics and phonology are unified, phonological phenomena must also be analysed with respect to the same detailed representations. This issue is addressed in §4, where the model is developed to account for phonological processes involving categorical neutralisation. Finally, §5 addresses arguments against including these kinds of phonetic details in phonological representations.

2 Parallels between phonetic and phonological phenomena

In this section I will present some particularly clear examples of parallels between phenomena which are conventionally classified as 'phonetic' and those which are conventionally classified as 'phonological'. One of the problems with the existing distinction between phonetics and phonology is the lack of clear criteria for assigning a given phenomenon to one portion of the grammar or another. This can be observed in the frequent disagreements over the status of a particular phenomenon, such as postlexical assimilation of coronal stops in English (Hayes 1992, Nolan 1992) or Korean vowel devoicing (Docherty 1995, S.-A. Jun 1995). This situation can make it difficult to test phonological hypotheses because it can be unclear whether some data are phonological and thus relevant to the hypothesis, or phonetic, and thus irrelevant. As noted above, apparent counterexamples to theoretical proposals have occasionally been dismissed on precisely these grounds. The problem also arises here: it is difficult to undertake a general survey of the differences between putative phonetic and phonological processes when it is often open to question whether the theory assigns a given phenomenon to one component of grammar or the other. However, it is generally accepted that only phonological processes are categorically neutralising, so this criterion is used to establish the phonological status of a process in most of the examples below. That is, these are mostly cases of parallelism between neutralising processes and those that involve fine degrees of variation that cannot be differentiated in standard phonological representations. But it should be noted that the very existence of such uncertainty about the hypothesised dividing line between phonetics and phonology lends credence to the idea that the line does not exist.

2.1 Assimilation and coarticulation

The parallels between assimilation and coarticulation are obvious and often noted: both involve assimilation of one segment to a neighbouring segment.

2.1.1 *Contextual nasalisation of vowel.* For example, most languages, including English, show partial nasalisation of vowels adjacent to nasals. This phenomenon is typically regarded as phonetic, since nasalisation is partial, with the velum lowering through the vowel (e.g. Cohn 1993), and because it is thought to be an almost inevitable consequence of moving the velum into and out of the lowered position required to produce a nasal consonant. This 'phonetic' nasal assimilation is mirrored by phonological neutralising nasal assimilation in languages like Bengali (Ferguson & Chowdhury 1960) and Nupe (Hyman 1975), in which nasalised vowels generally contrast with oral vowels, but the contrast is neutralised adjacent to nasals, presumably through assimilation of oral vowels to the nasal. Data from Nupe are shown in (2).

(2) *Nupe* (Hyman 1975)

a.	ba	'to cut'	bã	'to break'
b.	*ma		mã	'to give birth'

2.1.2 *Fronting of vowels by coronal.* Fronting of vowels by coronals has been well documented in the phonological literature (e.g. Clements 1991, Hume 1992). An example of neutralising fronting conditioned by coronals is provided by Cantonese (Cheng 1991). Cantonese generally contrasts front and back rounded vowels (3a), but between coronals, only the front rounded vowels appear (3b). Following Cheng, this can be analysed as the result of neutralising fronting of back vowels in this context. Note that a single coronal does not have the same effect, whether initial (3c) or final (3a).

(3) *Cantonese*
  a. kʰyt ‘decide’ kʰut ‘bracket’
  b. tʰyt ‘to take off’ *tʰut
  c. tʰuk ‘bald head’

Coarticulatory fronting of back vowels in the context of coronals is a ubiquitous effect cross-linguistically, and is typically consigned to the phonetics. For example, in English the vowel /u/ is substantially fronted between coronals, e.g. in /tut/ ‘toot’, compared to its realisation in the absence of coronals, e.g. /ku/ ‘coo’, and an intermediate degree of fronting is observed in a word like /tu/ ‘two’, which contains a single coronal. Note also that the degree of fronting varies substantially between languages (see §3.1 below).

2.1.3 *Vowel assimilation.* Similarly, vowel assimilation is parallelled by vowel-to-vowel coarticulation. For example, in Basque the low vowel /a/ is raised to [e] following a high vowel. This gives rise to alternations in the form of the definite suffix /-a/ (de Rijk 1970):

(4) *Basque*
  sagar+a ‘apple (DEF)’ mutil+e ‘boy (DEF)’

Vowel-to-vowel coarticulation, found in all languages, also involves assimilation of one vowel to another across consonants, but involves fine degrees of partial assimilation in that vowels assimilate only partially in quality, and the effects may extend through only part of the duration of a segment (e.g. Öhman 1966), and thus is generally regarded as phonetic. For example, in an English sequence of a high vowel followed by a low vowel, e.g. [-ilɑ-] in the phrase ‘he lost’, the high vowel conditions partial raising and fronting of the low vowel, as in Basque, but the effect is relatively small, and diminishes over the duration of the vowel.

2.2 Closed-syllable vowel shortening

A rather different kind of phonological process which finds parallels at the phonetic level is closed-syllable vowel shortening. Many languages with vowel-length contrasts shorten long vowels in closed syllables, neutralising the contrast, e.g. Turkish (Clements & Keyser 1983: 59), Yawelmani Yokuts (Newman 1944), Egyptian Arabic (Broselow 1976). In Turkish this gives rise to the alternations shown in (5).

(5) *Turkish*
  zɑmɑːn+ɯ ‘time (ACC)’ zɑmɑn ‘time (NOM)’
  iːkɑːz+ɯ ‘warning (ACC)’ iːkɑz ‘warning (NOM)’

A similar pattern of vowel shortening in closed syllables is observed in many, if not most, languages (Maddieson 1985). However, the shortening effect is often relatively small, and in many cases cannot be represented in standard phonological terms because it cuts across a vowel-length distinction, e.g. in Finnish both long and short vowels undergo shortening in closed syllables without neutralising the distinction between them (Wiik 1965), so shortening cannot be represented as deletion of a mora, for example.

2.3 Retroflexion and palatal vowels

There is an articulatory incompatibility between retroflexion and a high front tongue-body position. This articulatory difficulty is resolved in a variety of ways across languages, some of which would be classified as phonological while others would be regarded as phonetic. The articulatory conflict arises because full retroflexion involves forming a constriction between the tongue tip and the hard palate, whereas a high front vowel or glide involves a constriction between the front of the tongue body and the hard palate. Obviously it is not possible for two parts of the tongue to form simultaneous constrictions at the palate, and moving from one articulation to the other, as in a sequence such as [iɖ], requires substantial tongue movement.

These difficult sequences are usually resolved by reducing retroflexion, or by lowering and/or retracting the tongue body. The second option is exemplified by some Dravidian languages (e.g. Kodagu) which retract front vowels to central before retroflexes (Zvelebil 1970). Historically, the central vowels were allophonic variants of front vowels, but subsequent developments have rendered the difference contrastive, so retraction is neutralising. This phonological process finds a ‘phonetic’ counterpart in Gugada (Platt 1972), where high front vowels are partially retracted and lowered before retroflexes (transcribed by Platt as [ɪə]). Similarly, in American English, there are generally substantial off-glides from front vowels into a following [ɹ] in words like ‘beer’.

Acoma takes the alternative route of eliminating retroflexion, where retroflexes neutralise with palato-alveolars before front vowels (Miller 1965). That is, retroflexion is lost because it is inconsistent with the palatalising effect of front vowels. This is clearly phonological since it neutralises consonant place distinctions. A similar pattern is observed in Molinos Mixtec (Hunter & Pike 1969).

The phonetic counterpart to this process is observed in Gujarati (Dave 1977) and Mantjiltjara (Marsh 1969), where retroflexion is reduced following [i]. This can be observed from palatograms in Dave (1977). In both cases contrasts with anterior coronals are preserved, so the variation cannot be represented in terms of standard features.

2.4 Vowel reduction

The parallels between ‘phonetic’ and ‘phonological’ vowel reduction in unstressed syllables are well known. In most languages with stress, vowels in unstressed syllables are closer to each other in the vowel space than when they occur in stressed syllables (e.g. Delattre 1969, Fourakis 1990, Koopmans-van Beinum 1980). In most cases this is treated as ‘low-level’ phonetic variation. Neutralising vowel reduction, as observed in Italian, Russian, etc., can be regarded as the extreme case of convergence of neighbouring vowels. The analysis suggested below is that both patterns of reduction are motivated by the same pressure to reduce unstressed vowels, probably related to their shorter duration (Lindblom 1963), with neutralisation tending to result where reduction would otherwise result in an insufficiently distinct contrast. Indeed, neutralising 'phonological' reduction is typically accompanied by 'phonetic' reduction. For example, in Brazilian Portuguese, the higher and lower mid vowels neutralise in unstressed syllables, but the high vowels are also somewhat lowered and the low vowel is somewhat raised relative to their realisations in stressed syllables (Nobre & Ingemann 1987).

2.5 Postnasal voicing

In many languages, stops are voiced after nasals (Hayes & Stivers 1996, Pater 1996), as in the data in (6) from Zoque (Wonderly 1951: 120).

(6)	/N+pama/	[mbama]	'my clothing'
	/N+tatah/	[ndatah]	'my father'
	/N+kaju/	[ŋgaju]	'my rooster'

This process is typically treated as phonological, although it usually applies in languages without stop voicing contrasts (like Zoque), and is thus non-neutralising. However, a number of Bantu languages do show neutralising, and therefore incontrovertibly phonological, voicing of stops after nasals, e.g. Kikuyu, Ki-Nande, Bukusu (Hyman 1998).[3] The data in (7) are from Kikuyu (Armstrong 1940: 41). Historically, voiced stops have lenited in most environments, so the neutralisation is actually between a voiceless stop and a voiced tap or fricative. However, postnasal voicing is involved, and the result is clearly neutralising, and thus phonological.

(7)		*1st sg perf*	*infinitive*	
	a.	ndɛ̄mɛ́ɛtɛ́	tɛ́m-à	'cut'
		ndōméetɛ́	tóm-à	'send'
		ndūtēetɛ́	ɾūt-ǎ	'teach, lead out'
		ndōoɣeétɛ́	ɾǒoɣ-ā	'jump'
	b.	ŋgɔ̄mɛ́ɛtɛ́	kɔ́m-à	'sleep, lie down'
		ŋgēɾéetɛ́	kéɾ-à	'cross (street, etc.)'
		ŋgōɾēetɛ́	ɣōɾ-ǎ	'buy'
		ŋgāēetɛ́	ɣāj-ǎ	'divide'

---

[3] Neutralising postnasal voicing is probably relatively rare because the release burst and Voice Onset Time provide more important cues to voicing than the presence of voicing during the stop closure (Raphael 1981). Postnasal position only makes it more difficult to realise a fully voiceless stop closure, and so does not affect the primary cues to stop voicing contrasts. I speculate that neutralising postnasal voicing is found in the Bantu languages mentioned because the nasal–stop clusters are actually prenasalised stops, and therefore presumably shorter than regular clusters. The reduced duration of stop closure might make velum lowering more liable to interfere with the production of a voiceless stop burst.

At the phonetic level, Hayes & Stivers (1996) show that English voiceless stops display greater duration of closure voicing after nasals than after other sonorants, a pattern that has been observed anecdotally in a number of other languages.

2.6 Stressed vowel lengthening

Substantial lengthening of stressed vowels is often treated as phonological (e.g. Ilokano, Pacific Yupik, Selayarese, Swedish, etc. – see Goldsmith 1990: 157ff, Hayes 1995: 82ff for overviews), e.g. as a consequence of constraints on foot structure (Hayes 1995). However lengthening of stressed vowels is also ubiquitous on a smaller scale that cannot easily be represented in moraic terms.[4]

3 Unified analyses of parallel phonetic and phonological phenomena

The extensive parallels between phonetic and phonological phenomena described in §2 are in need of explanation. They are certainly rather mysterious if phonetics and phonology are separate components of grammar, operating in terms of distinct representations and principles. For example, in such a model, nasal coarticulation (§2.1.1) might be analysed as interpolation between a raised velum target for the vowel and a lowered velum target for the nasal, whereas neutralising nasal assimilation would be analysed as the spread of [+nasal] from the nasal to the vowel. The mechanisms and their representations are formally very different, so it is not apparent how to establish a relationship between them.

In this section we will pursue a very straightforward account of the parallels between 'phonetic' and 'phonological' phenomena, according to which they arise because both are motivated by the same constraints. So, for example neutralising and coarticulatory fronting of vowels in the context of coronals both result from effort-avoidance constraints that disfavour the rapid movements required to move from a coronal to a back vowel. This implies a unified framework for the analysis of both 'phonetic' and 'phonological' phenomena.

The fact that some of the processes involve categorical neutralisation, whereas others do not, will be analysed as the result of interaction of the motivating constraints with constraints on non-phonetic aspects of phonological representations, particularly contrastive structure. For example, neutralising fronting between coronals in Cantonese results from the

---

[4] Hayes (1995) argues that substantial lengthening of stressed vowels occurs only in iambic languages, and is motivated by the iambic-trochaic law, whereas trochaic languages show less lengthening, and this is argued to be part of the phonetic implementation of stress – i.e. it has a different motivation. However, even given this analysis, the correlation between length and stress in both iambic and trochaic systems remains as a broader generalisation in need of explanation.

interaction of the effort-avoidance constraint which motivates fronting with constraints requiring that contrasts between vowels be sufficiently distinct, and a constraint preferring that contrasts be maintained. Where the effort involved in producing an adequately distinct contrast becomes too great, these constraints may outweigh contrast maintenance, and neutralisation results in preference to maintaining a poor contrast.

We will take as our central example the relationship between coarticulation and assimilation, in particular the fronting of vowels by coronals (§2.1.2 above). The remaining cases will be discussed in §4.3.3.

We will first illustrate a constraint-based analysis of phenomena involving phonetic detail through an outline of an analysis of consonant–vowel coarticulation in second formant frequency. Then we will show how this framework can be generalised to derive neutralising fronting as a demonstration that categorical neutralisation can be derived in a model that employs phonetically detailed representations.

3.1 Constraints on phonetic detail: consonant–vowel assimilation in F2

Constraint-based analyses of phonological phenomena are now widely familiar. However, the idea that parallel phonetic and phonological phenomena are motivated by the same constraint implies constraint-based analyses of phenomena involving phonetic detail, which is less familiar (but cf. Byrd 1996, Kirchner 1997). The workings of such an analysis will be demonstrated through a case study of assimilation in second formant frequency between consonants and vowels.

In broad terms, the constraint-based framework proposed for the analysis of phonetic detail is similar to optimality-theoretic phonology (Prince & Smolensky 1993), in that outputs are selected so as to best satisfy conflicting violable constraints. It differs in that the constraints must be formulated over representations incorporating scalar phonetic detail, and we will see that constraint conflict is resolved in terms of constraint weighting rather than strict dominance ranking of constraints.

In a CV or VC sequence the second formant frequency (F2) at the edge of the consonant varies depending on the adjacent vowel, and F2 in the vowel varies depending on the adjacent consonant. These patterns of variation can usefully be interpreted as showing that each consonant and vowel has an F2 target, but a dispreference for fast articulatory movements can result in failing to achieve these targets, especially when consecutive targets are far apart (Lindblom 1963). Hence the observed contextual variation in actual F2 values. This conception is illustrated in Fig. 1, showing a schematic F2 transition for a CV sequence. The consonant and vowel have F2 targets, L (for 'locus') and T, respectively, but these targets are systematically 'undershot' with the actual F2 values being displaced towards each other. The coarticulatory fronting of back vowels adjacent to

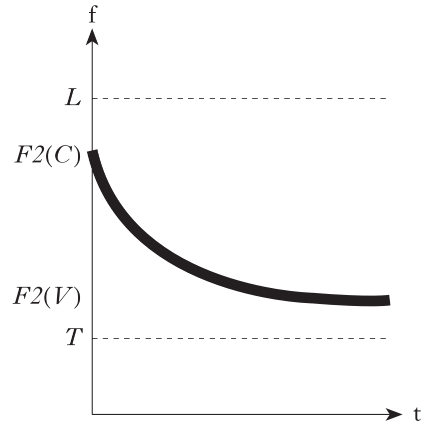


Figure 1

Schematic representation of a second formant transition from a stop to a vowel. *L* is the locus for the consonant, *F2*(*C*) is the actual value of F2 at the release of the consonant, *T* is the target for the vowel, *F2*(*V*) is the actual value of F2 at the vowel steady state.

coronals described in §2.1.2 is a particular instance of this general pattern: coronals generally have a high F2 target, so the F2 of an adjacent back vowel is raised, which is the acoustic correlate of fronting.

The nature of this partial assimilation between consonants and vowels has been described in fairly precise quantitative terms. We will review these descriptions, then outline a constraint-based model which can derive the observed patterns.[5]

3.1.1 *Consonant undershoot*. A large number of studies have shown that, for obstruent consonants, F2 at consonant release or closure, *F2*(*C*), follows F2 in the middle of the vowel, *F2*(*V*), i.e. *F2*(*C*) is higher when *F2*(*V*) is higher, and *vice versa*. This is illustrated for the English voiced velar /g/ in Fig. 2. The measurements are from one speaker, reading /gVt/ syllables, where V is each of /i ɪ eɪ æ ʌ ɑ ɔ u/.

This relationship between *F2*(*C*) and *F2*(*V*) for a given consonant has consistently been found to be highly linear (i.e. the points in plots like Fig. 2 fall close to a straight line) (Lindblom 1963, Sussman *et al*. 1993, etc.), so it can be described by a simple equation of the form shown in (8) (Klatt 1987), usually called a 'locus equation':

(8) $F2(C) = k_1(F2(V) - L) + L$

---

[5] For the sake of precision, the empirical generalisations and models discussed below will be formulated in mathematical notation. However, all formulas are restated in the text, so it should be possible to follow the discussion without paying close attention to the actual formulas.

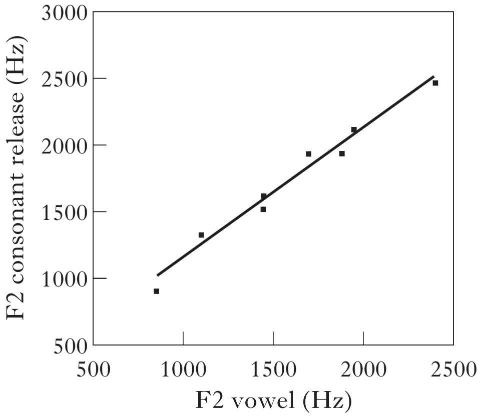


Figure 2

Plot of F2 measured at consonant release against F2 at the steady state, or stationary point, of the vowel, with regression line ($r^2 = 0{\cdot}97$).

where L is the target F2, or F2 'locus', for the consonant, and k_1 depends on the consonant and the style and rate of speech, and determines the slope of the line in a plot like Fig. 2. The interpretation of (8) is that there is a target F2 value, or locus, for a given consonant type, but the actual F2 at the consonant deviates towards the F2 in the adjacent vowel by a proportion of the difference between the consonant locus and the F2 in the vowel. That proportion is specified by k_1 – the larger k_1 is, the greater the degree of assimilation to the vowel. This parameter varies from consonant to consonant, e.g. it is higher for /b/ and /g/ than for /d/, and according to the style and rate of speech – it is generally a little higher in faster, more casual speech (Duez 1989).

3.1.2 *Vowel undershoot*. $F2(V)$ in turn assimilates to F2 of adjacent consonants. Modelling studies by Lindblom (1963) and Broad & Clermont (1987) have found support for the following relationship between the two:[6]

(9) $F2(V) = k_2(F2(C) - T) + T$

where T is the F2 target for the vowel, and k_2 depends on the consonant type, vowel duration and style of speech (Moon & Lindblom 1994). As with the locus equation, this implies that there is a target F2 for the vowel, T, but the actual F2 of the vowel 'undershoots' this target, deviating towards the F2 of the consonant by a proportion k_2 of the difference between the vowel target and the consonant F2.

---

[6] Broad & Clermont actually propose that vowel undershoot is proportional to $L - T$, rather than $F2(C) - T$. But given that the deviation of $F2(C)$ from L is proportional to $F2(V) - L$, this also implies the proportionality shown in (9).

3.2 An optimisation model

The observed pattern of partial assimilation between adjacent consonants and vowels can be viewed as a compromise between achieving the F2 targets for the consonant and vowel, and a preference to avoid fast movement between the two (hence a preference to minimise the difference between the two). This analysis is formalised here as a model in which $F2(V)$ and $F2(C)$ for a given CV sequence are selected by optimisation, i.e. so as to minimise violation of two basic constraints:

(10) a. Don't deviate from targets.
  b. Minimise articulator velocity (effort).

The second constraint, (10b), is assumed to be related to effort minimisation, on the reasonable assumption that faster movements involve greater effort, other things being equal.[7]

More specifically, these constraints are formalised as constraints on $F2(V)$ and $F2(C)$, shown in (11).

(11)	*Constraint*	*Cost of violation*
Ident(C)	$F2(C) = L$	$w_c(F2(C) - L)^2$
Ident(V)	$F2(V) = T$	$w_v(F2(V) - T)^2$
MinimiseEffort	$F2(C) = F2(V)$	$w_e(F2(C) - F2(V))^2$

The targets, L and T, are fixed for each consonant and vowel type, and w_c, w_v, w_e are positive weights.

The constraints Ident(C) and Ident(V) implement the requirement that realisations not deviate from targets – they require that the actual F2 at consonant and vowel equal the target value for that segment (these constraints are analogous to OT 'faithfulness' constraints, hence their names).[8] The MinimiseEffort constraint is formalised as a preference that there be no movement between consonant and vowel, i.e. zero velocity.

---

[7] This is not intended to imply that velocity is equivalent to effort, only that producing faster movements involves greater effort, other things being equal. This follows from common approximations of effort as related to force (e.g. Lindblom 1990b). There might also be a cost relating to articulatory precision, but this factor shouldn't be relevant in comparing stops.

[8] A reviewer points out that it is not clear that speech perception involves the extraction of formant frequencies, which would make formant-based constraints questionable. However there is good reason to think that formant frequencies provide a good approximation to the auditory dimensions of vowel quality, even if they are not the actual dimensions. For example, Plomp (1975) concludes that models of vowel dissimilarity based on differences between one-third octave band spectra correspond very closely to models based on formant frequencies. Bladon & Lindblom (1981) also tested a model of perceptual differences between vowels based on the whole spectrum, and while they obtained good results, they also concluded that it failed to give sufficient importance to the frequencies of formant peaks (see also Klatt 1979).

The main reason for doubting that formant frequencies play an important role in human speech processing is not that they provide a poor characterisation of human vowel perception, rather it is the difficulty of extracting them reliably.

Obviously these constraints conflict – if the consonant and vowel targets differ, then satisfying the IDENT constraints entails violating MINIMISEEFFORT, and *vice versa*. The idea is that the selected values of $F2(C)$ and $F2(V)$ should be those that best satisfy the constraints, as in OT. However, in standard OT, constraint conflict is resolved by ranking the constraints, with the higher-ranked constraint prevailing in cases of conflict. Constraint ranking is not appropriate in this context, because none of the constraints is completely dominant: if the IDENT constraints were dominant then targets would always be achieved; if MINIMISEEFFORT were dominant then one target would be completely ignored. To model the observed variation in consonant and vowel F2, it is necessary to allow violation of the constraints to trade off – that is, each constraint can be violated to a greater or lesser degree, and the best $F2(C)$ and $F2(V)$ are those that violate the three constraints least overall.

For the IDENT constraints, greater deviations from target constitute greater violations of the constraints. This is quantified as a 'cost' incurred for violating the constraint, calculated according to the formula in the third column of (11) – in the case of these constraints, the cost of violation is the square of the deviation from the target.

Similarly, violation of MINIMISEEFFORT is worse the larger the movement between consonant and vowel. Specifically, the cost of violation depends on the square of the change in F2 between consonant and vowel. This formulation involves a number of simplifications. First, effort cost should properly depend on articulatory movements. Change in formant frequency is used as an index of distance moved to keep all constraints in the acoustic domain. Second, velocity also depends on the duration of the transition – for present purposes we assume a fixed duration, so velocity depends only on the distance moved. Finally, different articulators presumably require different amounts of effort to move at a given speed. This variation will be modelled as variation in the effort weight factor w_e.

To determine how well candidate F2 values satisfy the combined constraints, we simply sum the costs imposed by the individual constraints. The weight of each constraint determines it relative importance in the overall evaluation. So the best values of $F2(C)$ and $F2(V)$ are those which incur the least total cost. In standard optimisation terms, $F2(C)$ and $F2(V)$ in a CV syllable are selected so as to minimise the overall cost function shown in (12).

(12) $$cost = w_c(F2(C)-L)^2 + w_v(F2(V)-T)^2 + w_e(F2(C)-F2(V))^2$$

So, if $L = 1700$ Hz and $T = 1000$ Hz, the costs of sample pairs of $F2(C)$ and $F2(V)$ are as shown in Table I. It can be seen that fully satisfying either the IDENT constraints or MINIMISEEFFORT yields high total costs, so the optimal solution will be a compromise between these requirements, with some violation of each constraint.

A fuller picture of the evaluation of candidate values for $F2(C)$ and $F2(V)$ can be obtained by plotting cost as a function of $F2(C)$ and $F2(V)$,

F2(*C*)	*F2*(*V*)	Ident(C)	Ident(V)	MinEffort	*total cost*
1700	1000	0	0	490,000	490,000
1500	1200	40,000	40,000	90,000	170,000
1350	1350	122,500	122,500	0	245,000

Table I

Evaluation of sample values for *F2*(*C*) and *F2*(*V*), with $L = 1700$ Hz, $T = 1000$ Hz, and all weights set to 1.

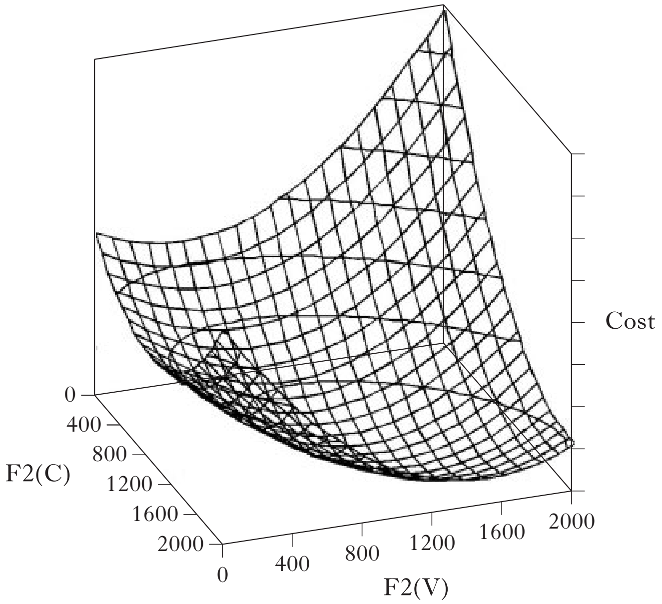


Figure 3

Cost plotted against *F2*(*C*) and *F2*(*V*), with $L = 1700$ Hz, $T = 1000$ Hz, and all weights set to 1. The minimum is located at *F2*(*V*) = 1233 Hz, *F2*(*C*) = 1467 Hz.

as in Fig. 3. It can be seen that the cost function forms a bowl-shaped surface, with optimal solution at the bottom of the bowl. This cost function is sufficiently simple to allow us to derive general expressions specifying the F2 values that minimise cost for a given set of targets and constraint weights. The minimum is located at the bottom of the bowl, where the slope along each dimension is zero. The location of this minimum can thus be found by differentiating the cost function with respect to *F2*(*C*) and *F2*(*V*) to yield two equations specifying the gradient along each dimension, then solving to find where both gradients are equal to zero. This procedure shows that the observed patterns of consonant and vowel undershoot follow from the optimisation model.

The location of the cost minimum along the $F2(C)$ dimension is given by the expression in (13), which has the form of a locus equation, as in (8) above, with k_1 replaced by a function of the constraint weights:

$$(13)\ \ F2(C) = \frac{w_e}{w_c + w_e}(F2(V) - L) + L$$

Taking the minimum along the $F2(V)$ dimension derives vowel undershoot, as in (9) above, with k_2 replaced by a function of the constraint weights:

$$(14)\ \ F2(V) = \frac{w_e}{w_v + w_e}(F2(C) - T) + T$$

So the model can derive the observed patterns of mutual assimilation in F2 using simple, output-oriented constraints.

The optimal values for $F2(C)$ and $F2(V)$, obtained by substituting (13) into (14), are:

$$(15)\ \ \text{a. } F2(C) = u_c(L - T) + L \quad \text{where} \quad u_c = \frac{w_e w_v}{w_e w_c + w_v w_c + w_e w_v}$$

$$\text{b. } F2(V) = u_v(L - T) + T \quad \text{where} \quad u_v = \frac{w_e w_c}{w_e w_c + w_v w_c + w_e w_v}$$

That is, $F2(C)$ and $F2(V)$ undershoot their respective targets by a proportion of the distance between consonant locus and vowel target. The proportion depends on the relative weights of the terms of the cost function. In effect, the interval between L and T is divided into three parts: consonant undershoot, vowel undershoot and transition in proportions $w_e w_v : w_e w_c : w_v w_c$. So the more heavily weighted MINIMISE EFFORT is, the more important it is to have a small transition, and thus more undershoot of vowel and consonant targets results. This shortfall is distributed between consonant and vowel according to the relative weights of IDENT(C) and IDENT(V).

More specifically, this model can be used to derive the coarticulatory fronting of vowels by coronals, discussed in §2.1.2 above. Many coronals have a relatively high F2 locus while back vowels have a low F2, so MINIMISEEFFORT motivates raising the F2 of a back vowel (i.e. fronting it) in order to reduce the size of the transition between them. A simple generalisation of the model also has the consequence that the greatest fronting effect will be observed between two coronals: to extend the model to CVC syllables, we can simply assume IDENT(C) applies to each consonant, so we sum the costs of undershoot for each consonant, and MINIMISEEFFORT applies to both the CV and VC transitions, so these costs are summed also. Then in the case of a coronal–vowel–coronal syllable such as [tut], each coronal will exert a fronting influence on the back

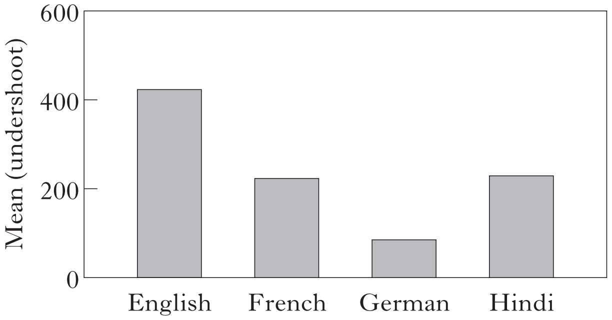


Figure 4

/u/ undershoot between coronals in four languages (in Hz).

vowel, resulting in more vowel fronting than would be conditioned by either coronal alone. This model of CVCs is almost certainly oversimplified, primarily because it neglects the difference in vowel duration between a vowel in a CVC syllable and the same vowel in a CV syllable. However this will lead the model to underpredict the increase in fronting resulting from adding a following coronal, because a shorter vowel implies faster transitions, so achieving a low F2 between coronals requires greater movement velocity, and hence greater effort cost, than in an open syllable.

Given that some phonologists have suggested that this type of coarticulatory effect is a consequence of 'universal rules', it is important to reiterate that consonant–vowel coarticulation varies cross-linguistically (Cohn & Keating 1988). This is certainly true of fronting of the back vowel /u/ between coronals, as shown by Fig. 4, which shows the amount of undershoot observed in this context in each of four languages. The data are from an unpublished study by myself and Hee-Sun Kim. The target F2 for /u/ in each language was estimated by measuring F2 in a context where /u/ was not influenced by neighbouring consonants – either in isolation or adjacent to a laryngeal such as [h]. Undershoot is then measured as the difference between the lowest F2 of /u/ in /tut/ and the target value for that language. Each column represents the mean undershoot from four speakers. Significant variation between the languages can be observed, with English showing very substantial undershoot, whereas German shows very little. In terms of the present model, this variation can be analysed in terms of differences in constraint weights, e.g. English assigns IDENT(V) a low weight compared to German.

4 Neutralisation

We now turn to the analysis of neutralising processes as paradigm cases of phonological phenomena and show that they can be derived by adding constraints on non-phonetic aspects of phonological representation to the model developed so far. That is, the analysis of neutralisation phenomena does not require separate coarse-grained representations, and can be motivated by the same constraints that motivate 'phonetic' phenomena like consonant–vowel coarticulation. Thus we can maintain the simple account of phonetics–phonology parallels according to which they result from the same constraints shaping both kinds of processes.

Most work that explicitly addresses the criteria for distinguishing phonetic and phonological processes adopts some variant of the idea that phonology is 'categorical' whereas phonetics is 'gradient' or 'quantitative' (e.g. Keating 1990, Pierrehumbert 1990, Cohn 1993). However, the nature of this opposition has not been precisely elucidated. There is a simple interpretation of this distinction according to which phonology employs discrete categories whereas phonetics operates in terms of continuous scales. But as Pierrehumbert (1990) observes, this alone cannot serve as the basis for distinguishing phonetics and phonology, because, by proliferating categories, a categorical representation can approximate a continuous one to an arbitrary degree of precision, as digitisation of speech signals demonstrates.

However, there is one clear sense of 'categorical process' that needs to be addressed in developing a model that unifies phonetic and phonology: neutralising processes are categorical in the sense that they collapse distinctive categories into a single category. For example, in §2.1.2 above, Cantonese vowel fronting is presented as the 'phonological' counterpart to coarticulatory fronting of vowels adjacent to coronals. The Cantonese process is categorically neutralising: /u/ and /y/ distinguish words in other contexts, but this contrast is neutralised to the front rounded vowel between coronals. Coarticulatory fronting conditioned by coronals in languages like English is not categorical in this sense – no contrasts are neutralised, the realisation of the contrasts is simply shifted somewhat. This seems to form part of the basis for the intuition that vowel fronting by coronals is categorical phonology in Cantonese, but gradient phonetics in English. Certainly, the analysis of coarticulatory fronting developed above does not seem completely adequate for the Cantonese data. As already noted, it does predict that the environment between coronals is an environment where back vowels will be fronted, and that the fronting effect will be stronger than adjacent to a single coronal. But, as it stands, the analysis seems to make neutralisation a rather improbable event since it would require front and back rounded vowels to be mapped onto precisely the same value of F2.

4.1 Constraints on contrasts

The existence of neutralisation does show that languages contain phonetic categories, but these are language-specific categories, i.e. the contrasting sounds of that language. The key to analysing neutralisation is to develop a proper analysis of the language-specific selection of these contrastive sound categories, e.g. contrasting vowels, from the space of possibilities implied by scalar representations. Following Flemming (1995, 1996), it is hypothesised that the selection of contrasts is subject to three constraints:

(16) a. Maximise the number of contrasts (in any given context).
     b. Maximise the distinctiveness of contrasts.
     c. Minimise effort.

These constraints are intended to represent goals which a system of contrasts should meet to best serve efficient communication. Maximising the number of contrasting sounds in any given context makes it possible to differentiate words rapidly. The distinctiveness of contrasts should be maximised so it is easy for the listener to distinguish words, and minimising effort allows speakers to communicate without undue expenditure of effort. These last two constraints are familiar from the work of Passy (1890), Martinet (1952, 1955) and Lindblom (1986, 1990), among others.

These constraints conflict with each other: given that the space of sounds which humans can produce is limited, fitting more contrasts into that space implies that those contrasts cannot be as distinct as a smaller set. Similarly, if effortful sounds are avoided, that excludes certain regions of the space of possible sounds, so contrasts will have too be more closely packed (and thus less distinct) than they would be if distributed through the entire space. Thus the selection of a set of contrasts involves balancing these three requirements, and different languages can reach different compromises, resulting in cross-linguistic variation in systems of contrasts.

The analysis of neutralisation proceeds from the observation that the effort involved in realising a particular sound depends on the context in which it appears, because it is likely that much of that effort is required to move from and to preceding and following sounds, as in the model of §3.2, where effort is based solely on articulator velocity. As a result, the best set of vowel contrasts in isolation, where no movement is required, may well be different from the best set for the context between coronals. Effort minimisation might lead to the selection of an inventory of vowels where the back vowels are somewhat fronted in this context, as is observed in English. However, a contrast between front and back rounded vowels, [y] *vs.* [u], becomes problematic if the back vowel is fronted too much, because that renders [u] too similar to [y].

So the contrast between [u] and [y] could be satisfactory in favourable contexts, but unsatisfactory, and therefore not selected, in the unfavourable context between coronals. This yields precisely the pattern of neutralisation observed in Cantonese. That is, neutralisation is simply a situation in which a contrast type is selected in some contexts but not in others, and according to the model of contrast selection just sketched, contrasts will be eliminated first in contexts where it is more difficult to make them distinct (cf. Steriade 1995, 1997).

According to this analysis, neutralising fronting between coronals is motivated by exactly the same constraint that motivates coarticulatory fronting in the same context, namely MINIMISEEFFORT, thus providing a direct account of the parallels between the two.[9]

4.2 A simplified analysis of neutralising vowel fronting

Development of comprehensive analyses incorporating constraints on contrasts can become rather complex, but the important point here is to demonstrate the general principle that neutralisation can be derived using phonetically detailed representations, and using the same constraints that govern phonetic detail, as in coarticulation. To this end, we will first consider a highly simplified analysis that illustrates that principle, then discuss how it can be developed.

The case we will consider is a simple analogue of Cantonese: the problem is to select an inventory of vowels that can appear between coronals. To keep the analysis as simple as possible, we will consider only high rounded vowels, so the problem is to choose between two candidate shapes of inventory, /y u/, with a contrast, or /y/, without. Further, we will specify the realisations of the vowels in terms of F2 only.

The constraints are shown in (17). The first two constraints are carried over unchanged from the analysis of CV coarticulation: IDENT(C) requires that F2 at a consonant be equal to the locus for that consonant, and MINIMISEEFFORT prefers that there be no change in F2 between consonant and vowel. But instead of an IDENT constraint requiring that the vowel F2 be equal to its target, we have the constraint MINDIST $= \Delta$, which requires that the [u] differ from the nearest contrasting vowel ([y]) by some minimum distance on the F2 dimension, Δ. The cost of violating this

---

[9] A reviewer raises the question whether the functionally based constraints proposed here are part of language users' mental grammars, or whether they operate externally to individual mental grammars, as part of the theory of sound change, for example (as suggested by Ohala 1995 and Hyman 1998). The particular implication of this question for the relationship between phonetics and phonology is that if the constraints argued here to be common to phonetics and phonology are not part of mental grammars, then we must separately consider the status of the phonetics–phonology divide in the psycholinguistics domain.

Given that our current understanding of phonological processing is extremely limited, this question must remain wide open. It is certainly plausible that functional constraints have their ultimate basis external to speakers. The most plausible account of how phonologies have come to optimally balance functional constraints is an evolutionary one – that is phonologies have become adapted for communication through an evolutionary process of adaptation through selection (Lindblom 1989, 1990b, Kirby 1999, Nettle 1999, Haspelmath, forthcoming). But even if functional constraints originated in this way, it does not necessarily mean that these constraints are not also represented in mental grammars. For example, Kirby (1999) argues that language universals that have emerged through linguistic evolution could become innate through natural selection in favour of faster language learning. That is, if all languages share some property, then learners who are innately constrained to assume this property will learn extant languages more easily (cf. Briscoe 2000). It is also possible that language users learn languages in terms of functional constraints without innate knowledge specific to language – an efficient mental representation of phonology might reflect the factors that shaped it. More definitive statements will have to await further empirical work in psycholinguistics.

constraint is the square of the shortfall of the difference between the vowels, $F2(y) - F2(u)$. So, following the model of contrast just outlined above, the vowel is constrained to be distinct from contrasting vowels rather than being constrained to achieve a specified target. This represents an improvement over Ident constraints which require that a target be specified for each segment type. It is obviously preferable for these targets to be themselves derived from constraints – i.e. the constraints on contrast.[10] Thus Ident(C) should be regarded as a stand-in for the interacting constraints that yield L as $F2(C)$ in a minimum effort context.

(17)	*Constraint*	*Cost of violation*
Ident(C)	$F2(C_n) = L_n$	$w_{cn}(F2(C_n) - L_n)^2$
MinimiseEffort	$F2(C_n) = F2(V)$	$w_e(F2(C_n) - F2(V))^2$
MinDist = Δ	$\lvert F2(y) - F2(u) \rvert \geqslant \Delta$	$w_v(\lvert F2(y) - F2(u) \rvert - \Delta)^2$ for $\lvert F2(y) - F2(u) \rvert < \Delta$
MaximiseContrasts		$-w_n$

For simplicity we will assume that $F2(y)$ is fixed, so there are only four parameters to be set: whether the contrast is maintained or not, and, if it is, the F2 of [u], $F2(u)$ and the F2 at C1 and C2.

The candidates are inventories of contrasting syllables. An inventory is evaluated by summing the cost of the syllables, then subtracting the benefit (i.e. a negative cost) accrued by maintaining more contrasts, and the lowest cost inventory is selected. If the benefit associated with MaximiseContrasts is a fixed amount per contrast, then in effect this procedure requires that the cost associated with each contrasting form be less than the benefit associated with adding that contrast. Since we are assuming a fixed realisation for /tyt/, this form incurs a fixed cost, so we need consider only the trade-off between the added cost of realising contrasting /tut/ and the benefit of having an additional contrast. So in effect, we are asking whether it is worth maintaining a contrast given the effort involved and dinstinctiveness achieved. If the effort and distinctiveness costs exceed the benefit of maintaining a contrast then it is better to neutralise.

Finally, MinDist applies only to contrasting vowels, so this constraint is not applicable if the contrast is neutralised.

As pointed out above, the context between coronals is problematic for the /y/–/u/ contrast: coronals have a high F2 locus, so the combined effect of Ident(C) and MinimiseEffort is to exert a fronting influence on a back vowel, which is required by MinDist = Δ to have a low F2 to remain distinct from /y/. Neutralising the contrast between /y/ and /u/

---

[10] The MinDist constraint also represents an implementation of the proposal by Manuel & Krakow (1984) and Manuel (1990) that coarticulatory variation is more restricted in crowded vowel inventories. Given the same constraint weights, there will be less vowel undershoot if the contrasting vowel is closer, because the MinDist constraint will impose higher costs.

eliminates the cost imposed by realising this form, but neutralisation itself carries a cost, since it reduces the number of available contrasts. So neutralisation becomes optimal only when the cost of realising /tut/ is greater than the benefit of maintaining a contrast, $-w_n$.

For example, the values and weights shown in the first column of Table II result in /u/ in /tut/ being realised with an F2 of 1367 Hz, and the total cost of constraint violations is 201,667. Maintaining the contrast in this form will be optimal if the benefit of maintaining a contrast, w_n, is greater than 201,667, otherwise a lower total cost will be achieved by neutralising the contrast. In Table II, w_n is set at 200,000, so the optimal outcome is neutralisation of the contrast in this environment. Column two shows the cost associated with the open syllable /tu/, given the same weights as in the first column. This cost is much lower than for /tut/, so maintaining a contrast is clearly optimal. The third column illustrates the consequences of varying the constraint weights: w_c is slightly reduced and w_v is correspondingly increased – i.e. the importance of vowel distinctiveness is slightly increased relative to the importance of achieving the consonant target. This results in a lower cost of maintaining the contrast, so this becomes the optimal outcome.

	/tut/	/tu/	/tut/
$F2(y)$	2000 Hz	2000 Hz	2000 Hz
L_t	2100 Hz	2100 Hz	2100 Hz
Δ	1000 Hz	1000 Hz	1000 Hz
w_c	0·25	0·25	0·2
w_e	0·25	0·25	0·25
w_v	0·5	0·5	0·55
cost of maintaining contrast	201,667	117,097	191,511
benefit per contrast (w_n)	200,000	200,000	200,000

Table II

Model parameters and associated minimum cost of maintaining a contrast between /u/ and /y/ (see text for details).

While highly simplified, the model demonstrates how categorical neutralisation can be derived with phonetically detailed representations, and with the same constraints that are used to derive 'phonetic' coarticulation effects. This is achieved through the interaction of phonetic constraints with constraints on contrasts, which are inherently categorical. Vowel fronting is motivated by MINIMISEEFFORT, just as in English vowel fronting in the context of coronals. The categorical behaviour results from a maximum cut-off for the cost that will be expended to maintain a contrast, set by the MAXIMISECONTRASTS constraint. If the cut-off is not exceeded, then 'allophonic' coarticulatory fronting results, and there is no neutralisation. In this way, contrasting phonetic categories and scalar contextual variation in those categories are derived within the same model, subject to the same constraints. This gives rise to the observed parallels between scalar and neutralising processes which provided the starting point for this investigation.

Before we move on to consider what is involved in scaling up this simplified analysis, there is a little more to be said about the circumstances under which neutralising vowel fronting is expected, based on the model. The first condition is that the language in question have a contrast whose distinctiveness will become small if back vowels are fronted – in the case of Cantonese this is the contrast between front and back rounded vowels. This is a situation in which considerable violation of MINIMISEEFFORT is required to avoid excessive violation of MINDIST due to vowel fronting, so the cost of maintaining the contrast can easily become high, making neutralisation preferable. Fronting conditioned by coronals is unlikely to have this effect in a language like English, because there are no front rounded vowels.

With regard to the contexts that condition fronting, this obviously depends on the F2 locus of the neighbouring consonants: a consonant will only exert a fronting influence on a vowel if the locus of the consonant is higher than the target for the vowel. As already noted, coronal stops typically have a relatively high F2 locus, e.g. 2104 Hz for /d/ *vs.* 1138 Hz for /b/, based on English data from Fowler (1994), and thus will condition fronting of back vowels. But according to many estimates, velar /g/ has an even higher locus (2709 Hz, based on data from Fowler 1994), and velars do not condition neutralising fronting (although they do have a coarticulatory fronting influence). According to the model this can occur only if the weight w_c is low relative to the other weights. In this circumstance, deviation from the locus (i.e. violation of IDENT(C)) does not incur much cost, so the consonant will assimilate to the vowel to reduce effort rather than fronting vowel. Empirically, velars do have low w_c relative to w_e, as can be seen from the steep slope of locus equations for velars, compared to a relatively shallow slope for coronals. This may indicate that properties of the release burst are more important than the precise frequency of F2 in distinguishing velars from non-velars.[11]

A more comprehensive analysis of Cantonese vowel fronting involves deriving the selection of complete vowel inventories for various contexts. The type of constraints required to derive vowel inventories, disregarding context, have been explored by Lindblom (1986) and ten Bosch *et al.* (1987). Combining the kinds of constraints they develop with the MINIMISEEFFORT and MAXIMISECONTRASTS constraints proposed here is relatively straightforward. The resulting models are less tractable mathematically, but the basic principle that categorical neutralisation results

---

[11] Klatt (1987) and Sussman *et al.* (1991) argue that velars actually have two F2 loci – a high locus before front vowels, and a low locus before back vowels. This would obviously exclude any fronting effect.

from selecting different sizes of inventories in different contexts remains the same.

4.3 Further issues

In this section we will consider further issues raised by the model developed so far, and compare it to existing optimality-theoretic models.

4.3.1 *Speech rate.* Coarticulatory fronting of vowels and neutralising fronting, as in Cantonese, do not differ only in the categorical nature of the neutralisation. Coarticulatory effects also vary substantially with speech rate, whereas Cantonese vowel fronting does not. Back vowels in the context of coronals tend to be more fronted when vowel duration is reduced (Lindblom 1963), presumably because faster movement is required to avoid undershoot. Cantonese vowel fronting, on the other hand, applies regardless of speech rate. This independence from speech rate appears to be typical of neutralisation processes. The model developed so far does not account for this difference.

The refinement required here is to provide a proper account of variability in phonetic realisation. That is, we have assumed that the selection of contrasting sound categories involves identifying fixed phonetic realisations of those categories. This is obviously a simplification: The realisation of a form can vary depending on factors such as speech rate. Furthermore it is clear that this is an essential property of language: languages should be adaptable to the variety of circumstances in which they are used, and that includes being spoken at a variety of rates. So in identifying an optimal set of contrasting sounds, it is not sufficient to consider a single, fixed realisation for each. Rather the constraints on contrasts must evaluate a range of realisations, specifically, at least a range of rate-dependent realisations. That is, a phonetics-phonology does not simply specify a set of contrasting phonetic forms, it is a system for realising a set of contrasting forms under a range of conditions (including variable speech rate). So it is not sufficient to have a single set of realisations which represent an optimal balance between distinctiveness, effort and number of contrasts. This balance must be achieved across a range of speech rates. The situation is analogous to automobile engine design: one of the desiderata for an engine is that it should be fuel-efficient. But it is not sufficient for an engine to run efficiently at a steady 55 miles per hour; it must be efficient across a range of speeds.

So rather than selecting fixed phonetic forms, a phonetics-phonology must select for each form a phonetic realisation as a function of speech rate. Consequently the evaluation of distinctiveness and effort costs must be modified to be defined over these variable realisations. To take the case of simplified Cantonese, this implies that rather than selecting a single realisation for /tut/, the model should select rate-dependent realisations for /tut/. To a first approximation, varying rate should only affect the effort involved in producing a given formant movement: a given magnitude of transition should require more effort if the transition duration is shorter. As a result, the optimal realisation will involve greater undershoot at faster rates (cf. Lindblom 1963), and a greater overall cost.

One approach to evaluating the performance of the system with speech-rate variation is to take the average effort and distinctiveness costs over a range of transition durations. One could also imagine that this average might be weighted to emphasise usual rates of speech. In simplified Cantonese, the cost of /tut/ will be greater than the cost of /tu/ (as in Table II above) at all rates of speech, because reduced duration increases effort costs in both forms. So /tut/ will cost more than /tu/ under any averaging scheme. Consequently the specifics of evaluation across rates do not have any immediate implications for the case at hand.

The crucial point is that while effort and distinctiveness are rate-dependent, the number of contrasts is not. The requirement that the number of contrasts available in any given context be maximised is hypothesised to follow from the need to differentiate words. So this constraint counts contrasts which can differentiate words – i.e. potential lexical contrasts. But if a contrast can distinguish words at any rate of speech, then these words must be distinguished in the lexicon. So if this contrast is then neutralised at higher rates of speech, that amounts to realising a lexical contrast with zero distinctiveness rather than eliminating a potentially problematic contrast. To avoid a contrast becoming indistinct (or very effortful) at fast rates of speech, it has to be neutralised at all rates of speech – i.e. it must not be able to distinguish words in the lexicon at all. That is, the price of speaking fast is that some contrasts become indistinct, but this cost is increased rather than reduced by 'neutralising' the contrast at faster rates. It will generally be better to realise the contrast as far as possible given the constraints imposed by speech rate, although effective neutralisation can arise as the end-point of extreme reduction.

So the average effort and distinctiveness costs of maintaining the /tut/–/tyt/ contrast are compared to the rate-independent benefit of maintaining the additional contrast. Loosely speaking, the 'decision' whether to maintain a contrast depends on the rate-varying cost of maintaining the contrast, but is not a decision that can be made on a rate-dependent basis, because it is a decision about what kind of words to allow in the lexicon. So the neutralisation of the /tut/–/tyt/ contrast in Cantonese occurs in spite of the fact that it could presumably be adequately realised in slow, careful speech, because it would not represent a satisfactory trade-off between effort and distinctiveness over a sufficient range of speech rates. Conversely, contrasts can be maintained even if they are indistinct at very high rates of speech as long as they are distinct over a sufficient range of less fast speech rates (e.g. the contrast between *will* and *wool* in English).[12]

---

[12] Steriade (1997) proposes a rather different solution to the problem of explaining why neutralisation is substantially rate-independent, namely that the effort and

4.3.2 *Constraint weighting* vs. *strict constraint dominance*. The proposed analysis gives a central role to the conflict between effort and distinctiveness, and is in this respect very similar to much recent work in phonology (e.g. Flemming 1995, J. Jun 1995, Silverman 1997, Steriade 1997, Boersma 1998, Kirchner 1998, Hayes 1999, etc). However, while these analyses are formulated in terms of Optimality Theory, the present analysis is formulated in superficially quite different terms. It is worth considering the extent of the real differences between the present model and standard OT, and the motivations for them.

There are in fact substantial similarities between the present framework and OT: both operate in terms of conflicting, violable constraints which apply to output forms. In both cases, the outputs are selected so as to best satisfy these conflicting constraints. The constraints may look different from familiar OT constraints, but, as I have tried to emphasise by giving them familiar names, they are essentially the same. An OT constraint can be regarded as a function from phonological forms to marks of violation (Prince & Smolensky 1993: 68f), and this is also the role of the constraints proposed here. They map phonetic/phonological forms onto marks of violation, expressed as real numbers (e.g. the square of the deviation from a target), and thus are simply non-binary constraints (Prince & Smolensky 1993: 72f). So the only difference in the nature of the constraints lies in the fact that many of the constraints here are formulated to refer to much finer details of realisation than is common in phonological analyses.

The real difference between the two models lies in the modes of constraint interaction, i.e. how the relative harmony with respect to a full set of constraints is determined from the evaluations of a candidate with respect to individual constraints. In the model proposed here, the overall evaluation of a candidate is expressed as a single number (its 'cost'), which is the weighted sum of the costs assigned by each constraint. In OT, on the other hand, constraint interaction is governed by ordering the constraints in a strict dominance hierarchy.

Two basic properties of the interaction between effort and distinctiveness motivate constraint weighting. First, effort and distinctiveness are scalar quantities, and they trade off against each other. That is, neither minimisation of effort nor distinctiveness strictly dominates the other: if minimisation of effort dominated then we would see no movement; if distinctiveness dominated then we would see only maximally distinct contrasts. Instead we see compromise between the two – a back vowel in the context of a coronal will be less distinct from front vowels in order to avoid excessive effort, but both distinctiveness and minimisation of effort

---

distinctiveness of contrasts are evaluated with respect to a standard rate and register of speech (cf. also Kirchner 1998: 299f). Actual realisations are then constrained to be as similar as possible to this canonical form. This approach is appealing in that it simplifies evaluation, but it is unclear why only one speech rate should be relevant to determining the quality of a contrast if it must actually be used at various speech rates.

are violated to some degree. Similarly, the distinctiveness of contrasts is traded off against each other: in the same coronal–back vowel sequence, not only is the back vowel fronted, assimilating to the coronal, but the coronal is somewhat velarised, assimilating to the back vowel. This three-way trade-off between vowel distinctiveness, consonant distinctiveness and effort is represented directly in the analysis of mutual assimilation in CV sequences presented in §3.2 (although IDENT constraints stand in for distinctiveness constraints to simplify the analysis).

Compromise between two constraints can be modelled with strict constraint dominance by decomposing each constraint into a set of ranked sub-constraints which can then be interleaved in the constraint ranking. This is a familiar strategy, adopted by Prince & Smolensky (1993) in their analysis of the compromise in syllabification between the requirements that nuclei be maximally sonorous and that syllable margins be minimally sonorous. However, where scalar compromise between constraints is conceptually central to an analysis, the need for constraint decomposition raises the concern that there is a lack of fit between the analysis and the framework in which it is formulated. More importantly, analysing the trade-off between two continuous-valued parameters, such as consonant and vowel undershoot, through constraint decomposition would require a great many sub-constraints (essentially quantising the F2 dimension; cf. Kirchner 1997, Boersma 1998), and a very particular ranking of these sub-constraints to derive the observed linear relationships between the two.

The second property of the interaction between effort and distinctiveness that is better modelled with weighted constraints is the existence of additive effects. That is, better satisfaction of two constraints can motivate greater violation of a third than either constraint could motivate alone. For example better vowel distinctiveness together with better consonant distinctiveness can make up for expending more effort. This type of interaction arises in weighted constraint systems: the sum of the costs of violating two lower-weighted constraints may add up to more than the cost of violating a higher-weighted constraint. This property is essential in accounting for the fact that there is more vowel fronting between coronals than in the context of a single coronal: the effort of making the transitions to the two coronals is summed, and outweighs vowel distinctiveness requirements to a greater extent than a single coronal would.

There is a device for modelling additive effects within the framework of strict constraint domination, namely local conjunction (Smolensky 1995). Local conjunction takes two constraints and forms a conjoined constraint which is violated if both of the base constraints are violated within some local domain. This conjoined constraint can be ranked higher than constraints which outrank both base constraints, thus in effect allowing violations of two lower-ranked constraints to outweigh violation of a higher-ranked constraint. Again, conjoined constraints would have to be proliferated to account for all of the acceptable trade-offs in the model of formant transitions.

The devices of constraint decomposition and local conjunction narrow the differences between a system that operates in terms of strict constraint dominance and one that sums numerically weighted constraint violations. So the fact that phonologists working with OT have found it necessary to 'subvert' strict constraint dominance with these devices makes it more plausible that phonology in general could be analysed in terms of weighted constraints. However, it is possible that the analysis of phenomena such as stress systems, which do not appear to revolve around the trade-off between effort avoidance and distinctiveness, will still motivate strict dominance relations between some constraints.[13]

4.3.3 *Constraints on duration.* The unified analysis of consonant–vowel coarticulation and assimilation of vowels to consonants identifies the source of the parallels between the two as being the constraints common to both: EFFORTMINIMISATION and the MINDIST/IDENT constraints. A review of the other examples of phonetics–phonology parallelism listed in §2 suggests that many of them could be analysed along similar lines. It is a standard idea that EFFORTMINIMISATION is a driving force behind coarticulation and assimilation, and the same has been argued for postnasal voicing by Hayes & Stivers (1996). The conflict between retroflexion and palatalisation is essentially similar also: substantial movement is required to move from full retroflexion to a high front tongue-body constriction, and effort can be reduced by assimilating the vowel to the consonant or *vice versa*, at the cost of reduced distinctiveness.

Closed-syllable vowel shortening (CSVS) and stressed vowel lengthening are somewhat different, in that MINIMISEEFFORT does not appear to be relevant; rather, it is constraints on duration that are at issue. So this section presents a brief sketch of an analysis of CSVS, in the interests of exemplifying a broader range of constraint types within the unified model.

As noted in §2.2 above, vowels are shorter in closed syllables than in open syllables in most languages (Maddieson 1985). In many cases, as in English, the effect is relatively small, and regarded as a matter of phonetic realisation. However, there are also many languages in which CSVS is neutralising, i.e. a contrast between long and short vowels is neutralised in closed syllables, with only short vowels appearing in that environment.

CSVS is probably a form of duration compensation, i.e. vowels shorten in closed syllables in order to keep syllable duration relatively constant in the face of the additional coda consonant. Evidence that this analysis is reasonable for English, at least, comes from the observation that duration compensation within the syllable is fairly general, in the sense that longer codas usually correlate with shorter vowels, and *vice versa*. For example, labial stops are longer than non-labial stops (Byrd 1993), and vowels are shorter before the labial stops (Peterson & Lehiste 1960). Similarly, voiceless stops are longer than voiced stops, and vowels are shorter before

---

[13] Borning *et al.* (1992) discuss (non-linguistic) examples of constraint systems which mix weighting and dominance relations.

voiceless stops. Vowels are also shorter before consonant clusters than before singleton consonants (see Munhall *et al.* 1992 for a review). Consonants also tend to be shorter after longer vowels (Munhall *et al.* 1992).

Compensation is usually less than complete. Complete compensation would result in a fixed duration for all syllables, but syllables actually get longer when more or longer segments are added. However, the individual segments generally get shorter when more or longer segments are added. This general pattern of partial compensation can be analysed in terms of a constraint-based model of duration assignment in which segment durations are selected so as to best satisfy the constraints in (19). Each segment type seg_i has a target duration, $target(seg_i)$. The first constraint penalises deviation from these segmental targets, with the cost of violation being equal to the square of the deviation from the target. Compensation results from the fact that there is also a target for the duration of a syllable, $target(\sigma)$, and the second constraint penalises deviation from this target, where the duration of the syllable, $dur(\sigma)$, is the sum of the durations of the segments in that syllable.

(19)	*Constraint*	*Cost of violation*
C-DURATION	$dur(seg_i) = target(seg_i)$	$w_{seg}(dur(seg_i) - target(seg_i))^2$
σ-DURATION	$dur(\sigma) = target(\sigma)$	$w_{\sigma}(dur(\sigma) - target(\sigma))^2$

These constraints conflict when the sum of the targets for the segments differs from $target(\sigma)$. In this case, minimising violation costs results in a compromise between the requirements of the segments and the syllable. So when a CVC syllable [tap] is compared to a CV syllable [ta], the CVC syllable will be somewhat longer overall, but the [t] and [a] in the CVC syllable will be shorter than their counterparts in the CV syllable to partially compensate for the duration of the coda consonant.[14]

This analysis can be extended to account for neutralising CSVS in much the same way as the analysis of coarticulation was extended to account for neutralising assimilation. CSVS is still motivated by the σ-DURATION constraint, but we add consideration of constraints on the vowel-length contrasts, as shown in (20). MinDist $= \Delta$, which requires that contrasting long and short vowels differ by some amount Δ in duration, conflicts with σ-DURATION in closed syllables, where it exerts pressure to shorten the vowel to avoid lengthening the syllable beyond its target length. The cost of realising a vowel-length contrast in the closed-syllable context must then be balanced against the benefit of having the additional contrast (MaximiseContrasts). As before, where the relative weights of the syllable-target and segment-target constraints make the

---

[14] This model is identical in effect to one proposed in Fujimura (1987), although they are formulated in different terms.

cost of realising the vowel-length contrast greater than the benefit of maintaining it, neutralisation becomes optimal.

(20)

	Constraint	*Cost of violation*
C-duration	$dur(seg_i) = target(seg_i)$	$w_{seg}(dur(seg_i) - target(seg_i))^2$
MinDist = Δ	$dur(V{:}) - dur(V) \geqslant \Delta$	$w_v(dur(V{:}) - dur(V) - \Delta)^2$ for $dur(V{:}) - dur(V) < \Delta$
σ-duration	$dur(\sigma) = target(\sigma)$	$w_\sigma(dur(\sigma) - target(\sigma))^2$
MaximiseContrasts		$-w_n$

Stressed vowel lengthening (§2.6) presumably involves additional constraints relating stress to duration, or perhaps rhythmic constraints of the kind discussed by Hayes (1995) under the rubric of the 'Iambic-Trochaic Law'.

Finally, the analysis of vowel reduction brings together duration, effort and distinctiveness constraints. The model of vowel undershoot presented in §3.1.2 above is based on one developed by Lindblom (1963) to account for vowel reduction resulting from reduced duration. Thus non-neutralising vowel reduction can be analysed by developing the model of formant transitions to properly account for the way in which effort cost increases with reduced vowel duration – that is, with shorter vowel durations, moving the same distance from consonant to vowel requires a faster, more effortful movement. Consequently, greater undershoot is liable to result. So again, the conflict between MinimiseEffort and distinctiveness constraints forms the core of the analysis. The additional twist is that undershoot is increased in unstressed syllables due to their shorter duration, which implicates an additional constraint regulating the relationship between stress and vowel duration, as in stressed vowel lengthening.

Again, neutralising vowel reduction arises where shortening of unstressed vowels makes it too costly to realise distinct vowel contrasts.

5 Phonetic detail in phonology and the typology of contrasts

Parallelism between phonetic and phonological phenomena provides a *prima facie* case for the unity of the two components, and the preceding sections outline a programmatic unified model that allows a direct and economical account of parallelism according to which parallel 'phonetic' and 'phonological' phenomena are motivated by the same constraints. In this section we will address the arguments that are intended to show that phonetics and phonology are separate, contrary to the conclusion drawn here.

The separation of phonetics and phonology more often figures as an assumption rather than a topic of research, so the literature contains few explicit arguments in favour of such a model. However, there seem to be two principal bases for separation that figure in discussions of the topic. The first is the claim that there is an observable difference between phonological and phonetic processes, with the former being categorical whereas the latter are gradient. This issue was addressed in detail in §3.2 above, where it was shown that categorical behaviour is not inconsistent with detailed representations as long as language-specific category structure is represented together with the detailed phonetic representation of those categories.

The second argument for distinguishing phonetics from phonology is more specifically an argument that the amount of phonetic detail in phonological representations must be severely restricted in order to avoid overpredicting the range of possible contrasts. If this is correct then phonological representations obviously must be distinct from phonetic representations.

The argument proceeds from the observation that the range of attested linguistic contrasts is much smaller than the range of phonetic differences. For example, tiny differences in stop duration do not form the basis for contrasts, but such differences are certainly possible phonetically. It is proposed that we should account for this observation by restricting representational possibilities so that only attested contrasts can be represented in phonology. This type of reasoning is clearly implicated in the following comment on laryngeal features (Keating 1984: 289):

> [Halle & Stevens] (and *SPE*) don't simply have the wrong features in these instances; they will *always* have *too many* features because they want to describe exactly how individual sounds are articulated. While we want the phonological features to have some phonetic basis, we also want to distinguish possible contrasts from possible differences.

The same idea is expressed in McCarthy's (1994: 191) statement that 'an adequate theory of phonological distinctive features must...be able to describe all *and only* the distinctions made by the sound systems of any of the world's languages' (emphasis added).

The assumption behind this approach to the typology of contrasts is that any phonologically representable difference should be a possible contrast. This assumption is far from necessary (cf. Anderson 1985: 122f), and is not even natural in the context of Optimality Theory. As Prince & Smolensky (1993: ch. 9) show, the contrastive status of sounds is a consequence of constraint ranking, so restrictions on possible contrasts can be addressed within the theory of constraints. The simplest illustration of this fact is the observation that if there is no faithfulness constraint referring to a particular feature, then that feature will never form the basis of a contrast, so the size of the feature set has no necessary consequences for the typology of possible contrasts (Kirchner 1997), but this is not the analysis adopted here.

The model of contrast selection proposed as part of the analysis of neutralisation (§4.1) provides a more satisfactory basis for developing the idea that contrasts are subject to explicit constraints. The constraint most relevant to an explanation for the observation that the range of attested contrasts is much smaller than the range of phonetic differences is the requirement that the distinctiveness of contrasts be maximised. It is reasonable to suppose that not only are more distinct contrasts preferred, but that some phonetic differences are simply too small to form a useful contrast. Thus a small difference in stop duration will always be rejected as insufficiently distinct to support an adequate contrast.

Furthermore, as observed by Steriade (1993b), phonologists do not consistently adhere to the principle that differences which are never contrastive should be excluded from phonological representations. Steriade notes that syllabification and prosodic structure are generally supposed to be universally non-contrastive, but form an important part of standard phonological representations. Non-contrastiveness is stipulated by excluding syllable and prosodic structure from underlying representations, but by the same token, a feature could be excluded from underlying representation to account for its failure to realise minimal contrasts. It is not hard to see why phonologists have allowed universally non-contrastive distinctions into their representations: even if it is assumed that phonological representations have a role to play in accounting for restrictions on contrast, this cannot be the only function that they serve. They must also provide the basis for the formulation of phonological rules or constraints. From this point of view, phonological representations should include all properties relevant to the analysis of phonological generalisations, and it is on this basis that syllable structure has been included in spite of its lack of contrastive potential.

It is not necessary to extend the domain of phonology in the way advocated here to see that a minimal feature set motivated by considerations of contrast is inadequate for the analysis of phonological generalisations. A simple case involves the laryngeal features. Lombardi (1995) proposes three laryngeal features, [voice], [aspirated] and [glottalised], in part because positing more features would allow the representation of non-contrastive differences. She explicitly notes that while Ladefoged (1971) distinguishes eleven classes of sounds on the basis of laryngeal properties, no language distinguishes all of these, and concludes that some of this detail must be excluded from phonological representations. For example, glottalised consonants, as found in many dialects of English, are collapsed together with ejectives as [glottalised] since no language contrasts these possibilities. However, this is problematic when we consider the phonological patterning of these sounds: ejectives are attested in both onset and coda, but are most commonly found in onset position, and often neutralise to plain stops in coda position (Steriade 1997),[15] as in Klamath (Barker 1964), Shapsug (Smeets 1984), Peruvian Aymara (MacEachern 1997: 46) and Maidu (Shipley 1963). Non-ejective glottalised stops, on

---

[15] Steriade argues that the relevant positional difference is actually 'pre-sonorant' *vs.* 'non-pre-sonorant' rather than 'onset' *vs.* 'coda'.

the other hand, are rarely found in onset position, and seem to be commonly restricted to coda position, e.g. English and Cantonese (Bauer & Benedict 1997). It is hard to formulate an account of these generalisations if both classes of sounds are represented as [glottalised] – the facts about ejectives suggest that the coda is a marked environment for [glottalised] stops, whereas the facts about glottalised stops suggest the reverse.

So the representational theory of contrast leaves us with the dilemma of allowing the representation of a difference which is never contrastive, or failing to provide the formulation of cross-linguistic generalisations about the distribution of different classes of sounds. Given a constraint-based approach to contrast, the dilemma is avoided since phonological representations are not subject to inconsistent requirements. A similar argument for the necessity of including universally non-contrastive distinctions in phonological representations has been made with respect to stop releases by McCawley (1967) and Steriade (1993b, 1994).

In conclusion, generalisations about phonological contrasts present no impediment to enriching phonological representations in the way argued for here.

6 Conclusions

Although phonetics and phonology are usually assumed to constitute separate components of grammar, operating in terms of very different representations and rules or constraints, actual phonetic and phonological phenomena are not as diverse as this model would suggest. As we have seen, there are many cases in which phonetic and phonological phenomena closely parallel each other.

I have argued that the existence of these parallels is best analysed as resulting from the phenomena having the same motivating constraints, and have outlined a unified framework for phonetics and phonology that allows for such an analysis. This framework is still programmatic, but demonstrates that it is feasible to derive variation in fine phonetic detail and categorical effects such as neutralisation with a single representation and set of constraints. This is made possible by representing the relationships of contrast alongside the detailed phonetic representations of the contrasting sounds. That is, the distinction between categorical and scalar phenomena is not rejected, rather it is treated as a distinction to be derived rather than stipulated.

The proposed model of phonetics and phonology is similar to optimality-theoretic phonology in that outputs are selected so as to best satisfy conflicting, violable constraints. However, the constraints considered here (particularly implementations of minimisation of effort and maximisation of distinctiveness) trade off against each other in an additive fashion, implying that these interactions are better modelled in a weighted constraint system rather than one which exclusively employs strict constraint dominance, as is the case for Optimality Theory.

REFERENCES

Anderson, Stephen R. (1985). *Phonology in the twentieth century*. Chicago: University of Chicago Press.

Armstrong, Lilias E. (1940). *The phonetic and tonal structure of Kikuyu*. London: Dawsons of Pall Mall.

Barker, M. A. R. (1964). *Klamath grammar*. Berkeley & Los Angeles: University of California Press.

Bauer, Robert S. & Paul K. Benedict (1997). *Modern Cantonese phonology*. New York: Mouton de Gruyter.

Bladon, R. A. W. & Björn Lindblom (1981). Modeling the judgment of vowel quality differences. *JASA* **69**. 1414–1422.

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

Borning, Alan, Bjorn Freeman-Benson & Molly Wilson (1992). Constraint hierarchies. *Lisp and Symbolic Computation* **5**. 223–270.

Bosch, L. F. M. ten, L. J. Bonder & L. C. W. Pols (1987). Static and dynamic structure of vowel systems. In *Proceedings of the 11th International Congress of Phonetic Sciences*. Vol. 1. Tallinn: Academy of Sciences of the Estonian S.S.R. 235–238.

Briscoe, Ted (2000). Grammatical acquisition: inductive bias and coevolution of language and the language acquisition device. *Lg* **76**. 245–296.

Broad, David & Frantz Clermont (1987). A methodology for modeling vowel formant contours in CVC context. *JASA* **81**. 155–165.

Broe, Michael & Janet Pierrehumbert (eds.) (2000) *Papers in laboratory phonology V: acquisition and the lexicon*. Cambridge: Cambridge University Press.

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

Byrd, Dani (1993). 54000 stops. *UCLA Working Papers in Phonetics* **83**. 97–115.

Byrd, Dani (1996). A phase window framework for articulatory timing. *Phonology* **13**. 139–169.

Channon, Robert & Linda Shockey (eds.) (1987). *In honor of Ilse Lehiste*. Dordrecht: Foris.

Cheng, Lisa (1991). Feature geometry of vowels and co-occurrence restrictions in Cantonese. *WCCFL* **9**. 107–124.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, George N. (1991). Place of articulation in consonants and vowels: a unified theory. *Working Papers of the Cornell Phonetics Laboratory* **5**. 77–123.

Clements, George N. & S. J. Keyser (1983). *CV phonology: a generative theory of the syllable*. Cambridge, Mass.: MIT Press.

Cohn, Abigail (1993). Nasalisation in English: phonology or phonetics? *Phonology* **10**. 43–81.

Connell, Bruce & Amalia Arvaniti (eds.) (1995). *Phonology and phonetic evidence: papers in laboratory phonology IV*. Cambridge: Cambridge University Press.

Darnell, Michael, Edith Moravscik, Frederick Newmeyer, Michael Noonan & Kathleen Wheatly (eds.) (1999). *Functionalism and formalism in linguistics*. Vol. 1: *General papers*. Amsterdam: Benjamins.

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, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics* **7**. 295–325.

Docherty, Gerald J. (1995). On a gestural account of lenis stop voicing in Korean: comments on Jun. In Connell & Arvaniti (1995). 254–264.

Docherty, Gerald J. & D. Robert Ladd (eds.) (1992). *Papers in laboratory phonology II: gesture, segment, prosody*. Cambridge: Cambridge University Press.

Duez, Danielle (1989). Second formant locus-nucleus patterns in spontaneous speech: some preliminary results on French. *PERILUS* **10**. 109–114.

Ferguson, Charles A. & Munier Chowdhury (1960). The phonemes of Bengali. *Lg* **36**. 22–59.

Flemming, Edward (1995). *Auditory representations in phonology*. PhD dissertation, UCLA.

Flemming, Edward (1996). Evidence for constraints on contrast: the dispersion theory of contrast. *UCLA Working Papers in Phonology* **1**. 86–106.

Flemming, Edward (1997). Phonetic optimization: compromise in speech production. In V. Miglio & B. Morén (eds.) *University of Maryland Working Papers in Linguistics* **5**. *Selected Phonology Papers from the Hopkins Optimality Theory Workshop 1997/University of Maryland Mayfest 1997*. 72–91.

Fourakis, Marios (1990). Tempo, stress and vowel reduction in American English. *JASA* **90**. 1816–1827.

Fowler, C. A. (1994). Invariants, specifiers, cues: an investigation of locus equations as information for place of articulation. *Perception and Psychophysics* **55**. 597–610.

Fromkin, Victoria A. (ed.) (1985). *Phonetic linguistics: essays in honor of Peter Ladefoged*. New York: Academic Press.

Fujimura, Osamu (1987). A linear model of speech timing. In Channon & Shockey (1987). 109–123.

Goldsmith, John (1990). *Autosegmental and metrical phonology*. Oxford & Cambridge, Mass.: Blackwell.

Haspelmath, Martin (forthcoming). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft*.

Hayes, Bruce (1992). Commentary on F. Nolan, 'The descriptive role of segments: evidence from assimilation'. In Docherty & Ladd (1992). 280–286.

Hayes, Bruce (1995). *Metrical stress theory: principles and case studies*. Chicago: Chicago University Press.

Hayes, Bruce (1999). Phonetically driven phonology: the role of Optimality Theory and inductive grounding. In Darnell *et al.* (1999). 243–285.

Hayes, Bruce & Tanya Stivers (1996). The phonetics of post-nasal voicing. Ms, UCLA.

Hume, Elizabeth (1992). *Front vowels, coronal consonants and their interaction in non-linear phonology*. PhD dissertation, Cornell University.

Hunter, Georgia G. & Eunice V. Pike (1969). The phonology and tone sandhi of Molinos Mixtec. *Linguistics* **47**. 24–40.

Hyman, Larry M. (1975). Nasal states and nasal processes. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.) *Nasálfest: papers from a symposium on nasals and nasalization*. Stanford: Language Universals Project, Stanford University. 249–264.

Hyman, Larry M. (1998). The limits of phonetic determinism in phonology: *NC revisited. Ms, UCB.

Jun, Jongho (1995). *Perceptual and articulatory factors in place assimilation: an Optimality-Theoretic approach*. PhD dissertation, UCLA.

Jun, Sun-Ah (1995). Asymmetrical prosodic effects on the laryngeal gesture in Korean. In Connell & Arvaniti (1995). 235–253.

Keating, Patricia A. (1984). Phonetic and phonological representation of stop consonant voicing. *Lg* **60**. 286–319.

Keating, Patricia A. (1985). Universal phonetics and the organization of grammars. In Fromkin (1985). 115–132.

Keating, Patricia A. (1990). Phonetic representations in a generative grammar. *JPh* **18**. 321–334.

Keating, Patricia A. (ed.) (1994). *Phonological structure and phonetic form: papers in laboratory phonology III*. Cambridge: Cambridge University Press.

Keating, Patricia A. & Abigail Cohn (1988). Cross-language effects of vowels on consonant onsets. *JASA* **84**. Suppl. 1. S84.

Kingston, John & Mary E. Beckman (eds.) (1990). *Papers in laboratory phonology I: between the grammar and physics of speech*. Cambridge: Cambridge University Press.

Kirby, Simon (1999). *Function, selection, and innateness*. Oxford: Oxford University Press.

Kirchner, Robert (1997). Contrastiveness and faithfulness. *Phonology* **14**. 83–111.

Kirchner, Robert (1998). *An effort-based approach to consonant lenition*. PhD dissertation, UCLA.

Klatt, Dennis H. (1979). Perceptual comparisons among a set of vowels similar to /æ/: some differences between psychophysical distance and phonetic distance. *JASA* **66**. Suppl. 1. S86.

Klatt, Dennis H. (1987). Review of text-to-speech conversion for English. *JASA* **82**. 737–793.

Koopmans-van Beinum, F. J. (1980). *Vowel contrast reduction: an acoustic and perceptual study of Dutch vowels in various speech conditions*. Amsterdam: Academische Pers.

Ladefoged, Peter (1971). *Preliminaries to linguistic phonetics*. Chicago: Chicago University Press.

Lindblom, Björn (1963). Spectrographic study of vowel reduction. *JASA* **35**. 1773–1781.

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

Lindblom, Björn (1989). Phonetic invariance and the adaptive nature of speech. In Ben A. G. Elsendoorn & Herman Bouma (eds.) *Working models of human perception*. San Diego: Academic Press. 139–174.

Lindblom, Björn (1990a). Phonetic content in phonology. *PERILUS* **11**. 101–118.

Lindblom, Björn (1990b). Models of phonetic variation and selection. *PERILUS* **11**. 65–100.

Lombardi, Linda (1995). Laryngeal neutralization and syllable wellformedness. *NLLT* **13**. 39–74.

McCarthy, John J. (1986). OCP effects: gemination and antigemination. *LI* **17**. 207–263.

McCarthy, John (1994). The phonetics and phonology of Semitic pharyngeals. In Keating (1994). 191–233.

McCawley, James D. (1967). Le rôle d'un système de traits phonologiques dans un théorie du langage. *Langages* **8**. 112–123. Reprinted as 'The role of a phonological feature system in a theory of language' in Valerie Makkai (ed.) (1972). *Phonological theory*. New York: Holt, Rinehart & Winston 522–528.

MacEachern, Margaret R. (1997). *Laryngeal cooccurrence restrictions*. PhD dissertation, UCLA.

Maddieson, Ian (1985). Phonetic cues to syllabification. In Fromkin (1985). 203–221.

Magen, Harriet (1984). Vowel-to-vowel coarticulation in English and Japanese. *JASA* **75**. Suppl. 1. S41.

Manuel, Sharon Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *JASA* **88**. 1286–1298.

Manuel, Sharon Y. & Rena A. Krakow (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Reports on Speech Research* **SR-77/78**. 69–78.

Marsh, James (1969). Mantjiltjara phonology. *Oceanic Linguistics* **8**. 131–152.

Martinet, André (1952). Function, structure, and sound change. *Word* **8**. 1–32.

Martinet, André (1955). *Économie des changements phonétiques*. Berne: Francke.

Miller, Wick R. (1965). *Acoma grammar and texts*. Berkeley & Los Angeles: University of California Press.

Moon, Seung-Jae & Björn Lindblom (1994). Interaction between duration, context, and speaking style in English stressed vowels. *JASA* **96**. 40–55.

Munhall, Kevin, Carol Fowler, Sarah Hawkins & Elliot Saltzman (1992). 'Compensatory shortening' in monosyllables of spoken English. *JPh* **20**. 225–239.

Nettle, Daniel (1999). Functionalism and its difficulties in biology and linguistics. In Darnell *et al*. (1999). 445–467.

Newman, Stanley (1944). *Yokuts language of California*. New York: Viking Fund.

Nobre, Maria A. & Frances Ingemann (1987). Oral vowel reduction in Brazilian Portuguese. In Channon & Shockey (1987). 195–206.

Nolan, Francis (1992). The descriptive role of segments: evidence from assimilation. In Docherty & Ladd (1992). 261–289.

Ohala, John J. (1971). Monitoring soft palate movements in speech. *Project on Linguistic Analysis, University of California, Berkeley* **13**. J01–J015.

Ohala, John J. (1995). The relation between phonetics and phonology. In William J. Hardcastle & John Laver (eds.) *The handbook of phonetic sciences*. Oxford: Blackwell. 674–694.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *JASA* **39**. 151–168.

Padgett, Jaye (1995). *Stricture in feature geometry*. Stanford: CSLI.

Passy, Paul (1890). *Étude sur les changements phonétiques et leur caractères généraux*. Paris: Librairie Firmin-Didot.

Pater, Joe (1996). *NC. *NELS* **26**. 227–239.

Peterson, G. E. & I. Lehiste (1960). Duration of syllable nuclei in English. *JASA* **32**. 693–703.

Pierrehumbert, Janet B. (1980). *The phonetics and phonology of English intonation*. PhD dissertation, MIT. Distributed 1987, Indiana University Linguistics Club.

Pierrehumbert, Janet B. (1990). Phonological and phonetic representation. *JPh* **18**. 375–394.

Pierrehumbert, Janet B. & Mary E. Beckman (1988). *Japanese tone structure*. Cambridge, Mass.: MIT Press.

Platt, J. T. (1972). *An outline grammar of the Gugada dialect, South Australia*. Canberra: Australian Institute of Aboriginal Studies.

Plomp, Reinier (1975). Auditory analysis and timbre perception. In Gunnar Fant & Marcel A. A. Tatham (eds.) *Auditory analysis and perception of speech*. New York: Academic Press. 7–22.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. To appear, Cambridge, Mass.: MIT Press.

Raphael, Lawrence (1981). Durations and contexts as cues to word-final cognate opposition in English. *Phonetica* **38**. 126–147.

Rijk, Rudolf de (1970). Vowel interaction in Bizcayan Basque. *Fontes Linguae Vasconum* **2**. 149–167.

Shipley, William (1963). *Maidu texts and dictionary*. Berkeley & Los Angeles: University of California Press.

Silverman, Daniel (1997). *Phasing and recoverability*. New York: Garland.

Smeets, R. (1984). *Studies in West Circassian phonology and morphology*. Leiden: Hakuchi Press.

Smolensky, Paul (1995). On the internal structure of the constraint component *Con* of UG. Handout of paper presented at UCLA.

Steriade, Donca (1993a). Orality and markedness. *BLS* **19**. 334–347.

Steriade, Donca (1993b). Closure, release, and nasal contours. In Marie Huffman & Rena Krakow (eds.) *Nasals, nasalization, and the velum*. Orlando: Academic Press. 401–470.

Steriade, Donca (1994). Complex onsets as single segments: the Mazateco pattern. In Jennifer Cole & Charles Kisseberth (eds.) *Perspectives in phonology*. Stanford: CSLI. 203–293.

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

Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. Ms, UCLA.

Sussman, H. M., K. A. Hoemeke & F. S. Ahmed (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *JASA* **94**. 1256–1268.

Sussman, H. M., H. A. McCaffrey & S. A. Matthews (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *JASA* **90**. 1309–1325.

Wiik, Kalevi (1965). *Finnish and English vowels*. Turku: Turun Yliopisto.

Wonderly, W. L. (1951). Zoque II: phonemes and morphophonemes. *IJAL* **17**. 105–123.

Zsiga, Elizabeth C. (1997). Features, gestures, and Igbo vowels: an approach to the phonology–phonetics interface. *Lg* **73**. 227–274.

Zvelebil, Kamil (1970). *Comparative Dravidian phonology*. The Hague: Mouton.