

# Vowel patterns in mind and sound

John Harris  
University College London

Geoff Lindsey  
City University, London

Language is primarily an auditory system of symbols. In so far as it is articulated it is also a motor system, but the motor aspect is clearly secondary to the auditory. In normal individuals the impulse to speech first takes effect in the sphere of auditory imagery and is then transmitted to the motor nerves that control the organs of speech. The motor processes and the accompanying motor feelings are not, however, the end, the final resting point. They are merely a means and a control leading to auditory perception in both speaker and hearer... Hence, the cycle of speech...begins and ends in the realm of sounds (Sapir 1921: 17-18).

## 1 Introduction

The sound aspect of a linguistic sign provides the information which enables listeners and speakers to access the sign's lexical and grammatical meaning.<sup>1</sup> The channel for this information is the speech signal, through which speakers transmit and monitor the information and listeners receive it. On the basis of this rather obvious point, it would be natural to conclude that phonological features — the code in terms of which the information is compiled — should be defined in terms of auditory imagery.

Yet for a generation the most influential brand of feature theory has been centred almost wholly on articulation. This is surprising; for, while articulations constitute a delivery system for linguistic information, they are not of themselves information-bearing. To echo Sapir: the perceptible form of a linguistic sign is essential and primary, the means by which it is produced secondary.

Consider an analogy from the realm of written signs. Countless disabled people have learned to read, despite lacking the mechanical wherewithal to write. Are such people illiterates? By no means. Only a bias towards production over perception, or some bogus egalitarianism in relation to these domains, would insist that they are. Who could learn to write without the ability to read?

Only a trained ape — an illiterate ape.

Even when the recent hegemony of articulatory features has been challenged, it has rarely been with anything like the wholehearted commitment to the auditory-acoustic that characterised Jakobsonian feature theory (Jakobson, Fant & Halle 1952). One response has been to make acoustic specifications parasitic on primarily articulatory features (see for example Clements & Hertz 1991). Another allows for acoustic features to coexist with an articulatory set (see Flemming 1995, Boersma 1998).

In work of this orientation, the auditory and the acoustic are often conflated, and for understandable reasons: they are more intimately linked with one another than either is with the articulatory. Just as optics deals with light processed by vision, so acoustics deals with sound processed by hearing. There is an acoustic-auditory chain consisting of waves in the air, disturbance in the middle ear, firings in the inner ear and cognitive activations. It is possible to draw a distinction between disturbance external to the human body and response internal to the human body, the latter being strictly auditory. But note that even this distinction gets tricky at the interface: is the vibrating air within the ear canal part of the human body or not?

The crux of the matter is that in speech the acoustic entails the auditory-perceptual, by definition. The informational or signifying potential of variations in atmospheric pressure is realised when they produce disturbances in the ear which trigger central responses. But the high degree of isomorphy between activity in the air and the ear-brain has no parallel in the relation between either of these domains and vocal-tract movements. Without air disturbance there is no ear disturbance; and without ear disturbance there is no information. But the informational force of air disturbances is quite independent of what produces them — be it a human vocal tract, a digital synthesiser or a budgerigar.

In this chapter, taking the representation of vowel quality as our illustrative focus, we set out reasons for rejecting feature definitions based on articulation or raw acoustics. The alternative view we present is somewhat old-fashioned, owing much to the tradition of Saussure, Sapir and Jakobson. It holds that the mental representation of speech sounds is constituted not of tongue heights, for instance — nor of formant heights, nor for that matter of basilar stimulation points. Rather it is constituted of information-bearing patterns which humans perceive in speech signals. The arguments on which this view is built are, we believe, fundamentally sound. However, they have hardly been heeded, still less countered, in the recent feature literature.

This conclusion, let us hasten to point out, does not lead us to deny that

phonology has been shaped in some measure by the physical nature of the vocal apparatus. Other shaping influences include gravity and atmospheric pressure. However, none of these belongs in mental representation. Nor does the conclusion entail totally expunging from phonology all considerations traditionally regarded as phonetic, the type of austere view sometimes advocated in the literature, most recently by Hale & Reiss (1999, this volume). We address this point in §2. While phonological categories qua mental objects cannot be the same as external acoustic events, we take the view that they must nevertheless stand in some non-arbitrary relation to them — hence the justification in referring to the categories, in the spirit of Saussure and Sapir, as auditory images.

We identify the set of categories constituted by sound images through the traditional phonological method of determining the manner in which sounds are organised into systems and natural classes. In §3, we briefly survey the salient patterns of organisation displayed by vowel systems, both in their maximal form and when subject to positional neutralisation. §4 suggests how the phonological categories which by hypothesis underpin these regularities can be construed as sound images, elementary patterns detectable in the acoustic signal. Our proposals rest on familiar phonetic findings but are novel in establishing the COMPOSITIONALITY of vowels: some vowels are literally made up of others. §5 considers and rejects various alternative proposals to explain vowel patterning in terms of vision, articulation, or raw acoustics. §6 concludes.

## **2 Phonology: module or epiphenomenon?**

Generations of academics and students have struggled over the conceptual distinction between phonetics and phonology. Of the two, phonetics probably yields more readily to succinct definition: ‘the study of the physical aspects of speech’ would be one attempt — though even this might exclude such phonetic domains of enquiry as psychoacoustics and the linguistic phonetics which presupposes an understanding of linguistic contrast.

In generative linguistic theory, phonology has classically been seen as a component of grammar, an independent cognitive module characterising what humans know about linguistic sound patterns. However, linguists have long been nagged by the obvious fact that much of the sound patterning in language reflects constraints imposed by the physical nature of the organs of speech and hearing. Many generalisations traditionally considered phonological, such as the

preference human language shows for voiceless over voiced obstruents, are clearly relatable to non-cognitive factors.

Which raises the fundamental question of whether there exists a core of immaculate phonology that can be exposed through phonetic cleansing. Answering this question in the affirmative holds an obvious appeal for professional phonologists, who frequently take the existence of a pure phonology as a matter of necessity. In the words of Hale & Reiss, 'there must be a core of formal properties (e.g., organization into syllables and feet, feature spreading processes) that are modality independent [i.e. equally applicable to sign language] and thus not based on phonetic substance. The goal of phonological theory should be to discover this formal core' (1999: 2).

However, Occam's razor works just as effectively on brains as it does on tongues and ears. Many phonological properties not amenable to functional-phonetic explanation are almost certainly shared with other domains. For instance, categoriality and arboreal structure (or, more generally, head-dependent relations) are shared with syntax. Moreover, it is even open to debate whether these are peculiar to universal grammar, since both can plausibly be attributed to cognition in general. As for Hale & Reiss's own examples, feet and syllables seem to be required by our cognitive musical faculties. And it is hard to think of a human behaviour to which some analogue of feature spreading would *not* be applicable (again music would do as an example). The search for phonology's formal core may turn out to be about as successful as the quest for universal solvent or the world's edge.

An alternative view, not widely endorsed by generativists, is that phonology is epiphenomenal, a well-defined area of study but not one that corresponds to any specific organ of body or mind. A research strategy founded on this view does not *preclude* the possibility that there exist uniquely phonological forms of knowledge or behaviour but places the burden of proof on the demonstration of their existence. Explanation of linguistic sound patterns (and, depending on one's definition, gesture patterns in sign language) is initially to be sought in domains such as cognitive psychology, aerodynamics, neurology and physiology, a research ethic maintained most prominently in the work of Ohala (see 1992 for references). Peculiarly phonological apparatus, on this view, should only be appealed to as an explanation of last resort.

Among the functionalist targets singled out for attack by Hale & Reiss (1999) are recent Optimality-theoretic treatments of positional vowel neutralisation. Two main characteristics of this phenomenon demand explanation. Why do subsystems of vowel contrast vary in size according to the phonological context

in which they occur? And why do reduced subsystems tend to mimic the maximal inventories of languages with simpler overall systems? A hermetically sealed phonological account would posit neutralisation rules or constraints for the first characteristic, attribute the second to universal markedness preferences, and leave things at that. Functionalist OT accounts go further by proposing that each of the constraints which deliver these effects directly embodies some pressure applied by the physics of speech. Beckman (1997), for example, proposes a set of positional faithfulness constraints which, when ranked high enough, protect certain privileged contexts, such as word- and stem-initial syllables, from constraints which favour neutralisation (see also Zoll 1998 for discussion and further references). The faithfulness constraints, she claims, are motivated by psycholinguistic evidence which demonstrates the perceptual salience of the relevant contexts. Flemming (1995) attributes the universal tendency towards unmarked patterning in both full and reduced systems to a preference for the maximal dispersion of contrasts in vowel space, compelled by phonological constraints which are grounded in auditory processing (cf. Lindblom 1986).

Critiques of this general functionalist approach centre on the question of why it should be necessary to duplicate in the grammar explanations of sound patterning which are independently required by general theories of sound change and language acquisition. On this point, the views of Ohala and Hale & Reiss largely converge (see also Hyman 1998a). Indeed, it might even be said that the duplication is inherently contradictory. For example, it is not immediately obvious how positional asymmetries in vowel distribution can simultaneously ‘arise from’ grammar-internal constraints (Beckman 1997: 7) and be motivated by grammar-external speech functioning. The constraints in question could have at best some intermediate place in a chain of causation.

What has been conspicuous by its recent absence from this debate is any serious discussion of the very nature of the linguistically significant categories that code vowel contrasts. Much of the relevant literature simply takes the categories for granted — essentially the articulatory features inherited from SPE or an acoustic set based on the parameters of machine spectrography. It is our contention that both of these approaches are fundamentally flawed. They not only fail to establish a satisfactory bridge to auditory perception but also inadequately describe the nature of vowel systems and processes.

We present an alternative, compositional model in which the qualities *a*, *i*, *u*, which anchor vowel systems and which show up preferentially as neutralisation reflexes, are simpler than other vowels — literally simpler, in terms of the

signal patterns that are detected by speakers. That is, vowel neutralisation can be shown to involve both a reduction in the complexity of phonological representations and a corresponding reduction in the amount of information that can be extracted from the signal.

We are quite prepared to accept that the compositional aspect of this account may ultimately prove amenable to explanation in terms of some combination of physiological, neurological and cognitive facts that are not specific to language. Indeed, decisive evidence may eventually emerge to debunk the generativist notion of a fully modular phonology altogether. However, neither of these developments could relieve working phonologists of their duty to model as precisely as possible the speech pattern phenomena that call for ever deeper explanation.

In short, a Hjelmslevian programme of the kind advocated by Hale & Reiss is both too exclusive and too inclusive. It is too exclusive to the extent that it banishes auditory imagery from the mental representation of phonology, and too inclusive to the extent that it presupposes, contra Occam, a modular core which cannot be the null hypothesis.

### **3 Vowel patterns in phonology**

It is a well-known fact that vowel systems across the world's languages show a clear predilection for triangular patterning built around the 'corner' vowels *a*, *i*, *u*. Recurrent extensions of this basic set include the canonical five-term system (the most favoured of all) and, through the addition of ATR contrasts, seven- and nine-term inventories (Crothers 1978, Maddieson 1984).

Typically of course, a language's maximal vowel inventory is not sustained in all phonological contexts. Positional neutralisation of vowel contrasts produces two general patterns, which may occur singly or in combination in a given language. One can be described as centrifugal: vowels belonging to a contracted subsystem disperse to the far corners of vowel space. Textbook examples are provided by Modern Greek, Russian and Tamil, where canonical five-term systems reduce to *a*, *i*, *u* in unstressed positions. The other pattern is centripetal, in which neutralisation reflexes are drawn into central areas of vowel space. Languages exhibiting this type of reduction typically do so in conjunction with some part of the centrifugal pattern. Romance is particularly rich in examples — Catalan, Neapolitan Italian (Bafile 1997), Portuguese and Romansch (Kamprath 1987), to name a few.

The particular version of the pattern that occurs in Catalan results in a seven-term peripheral system contracting to *i*, *u*, *ə*, producing stress-conditioned alternations such as those below (Palmada Félez 1991):

(1)	<i>prím</i>	‘slim’	<i>əprimár</i>	‘to slim’
	<i>sérp</i>	‘snake’	<i>sərpəntí</i>	‘winding’
	<i>pél</i>	‘hair’	<i>pəlút</i>	‘hairy’
	<i>gát</i>	‘cat’	<i>gətét</i>	‘kitten’
	<i>lúm</i>	‘light’	<i>luminós</i>	‘luminous’
	<i>gós</i>	‘dog’	<i>gusét</i>	‘puppy’
	<i>pórt</i>	‘port’	<i>purtuári</i>	‘of the port’

All of the languages just mentioned illustrate the manner in which the size and shape of vowel systems can vary as a function of stress. However, it would be far from the truth to assume that all cases of systemic contraction are conditioned in this way. This immediately renders problematic any attempt to explain vowel neutralisation in terms of diminished perceptual salience resulting from reduced loudness and duration, a point we return to below. In some cases, the context to which the maximal inventory is tied is defined in purely morphological terms. The telling evidence comes from languages in which stress either is orthogonal to neutralisation or is absent altogether. The full six-vowel system of Chumash, for example, is restricted to roots; in affixes, we find contraction to *a*, *i*, *u* (Applegate 1972). This distribution is quite independent of word stress, which typically falls on the penultimate syllable. Neutralisation in the total absence of stress is found, for example, in Punu and Ruund (both Bantu), where canonical five-term systems contract to *a*, *i*, *u* in suffixes (Hyman 1999: 239).

A variation on the reductive neutralisation theme is to be found in languages which combine it with assimilative neutralisation. In such cases, certain vowels only appear in a neutralisation site if they are harmonically supported by some other vowel occurring in a dominant nucleus. Pasiego Spanish provides an example in which stress is implicated (Penny 1969). Stress-free examples include Ibibio (Urua 1990) and the height-harmony pattern widely encountered in Bantu (Hyman 1998b, 1999). In each of these cases, as with non-harmonic contraction, it is mid vowels that are distributionally defective in the neutralisation contexts.

A relatively simple example of the Bantu pattern is provided by Chichewa

(Mtenje 1985): verb roots license a maximal five-vowel inventory (see (2)), while the basic vocalic content of suffix vowels is restricted to *a* (e.g. reciprocal *-an-*), *i* (e.g. applied *-il-*) and *u* (e.g. reversive *-ul-*). Mid vowels do occur in suffixes, but only as alternants of high vowels lowered under the harmonic influence of a mid root vowel (see (2)b).

(2)	Root	Root+applied	
(a)	<i>pind-a</i>	<i>pind-il-a</i>	‘bend’
	<i>put-a</i>	<i>put-il-a</i>	‘provoke’
	<i>bal-a</i>	<i>bal-il-a</i>	‘give birth’
(b)	<i>lomb-a</i>	<i>lomb-el-a</i>	‘write’
	<i>konz-a</i>	<i>konz-el-a</i>	‘correct’

Centrifugal patterns of vowel neutralisation have always been something of an embarrassment to orthodox SPE-style features. Standard specifications organised around [ $\pm$ high], [ $\pm$ low], [ $\pm$ back] and [ $\pm$ round] fail to tease out the fact that the corner vowels form a natural class vis-à-vis mid. This is one of the shortcomings that spurred the development of an alternative model of vowel contrast in which the vowels *a*, *i*, *u* are treated as the embodiment of independent segmental elements (Anderson & Jones 1974 and subsequent work by many researchers — see Harris & Lindsey 1995 for references). This is the approach we adopt here, employing the labels [A], [I] and [U] for the categories, as distinct from their respective pronunciations *a*, *i*, *u*. The following section is devoted to a discussion of the definition of these categories. Mid vowels are represented as compounds: *e* as [A, I], *o* as [A, U]. Thus mid vowels are more complex than corner vowels by virtue of being composed of them.

This model allows for centrifugal vowel reduction to be characterised quite simply: representationally complex vowels are barred from the neutralisation site.

Other types of vocalic process provide further support for compositionality. Perhaps the most graphic confirmation comes from vowel coalescence. In one recurrent pattern, two corner vowels collide through morphemic juxtaposition to yield a mid vowel, as in Zulu *na-inkosi* > *nekosi* ‘with the chief’, *na-umuntu* > *nomuntu* ‘with the person’. The phenomenon is straightforwardly represented as the compacting of two sequentially ordered elements into a single complex segment, [A]–[I] yielding [A, I] in the case of *a-i* > *e*. This

account compares favourably with one based on articulatory features, in which one set of specifications has to be rewritten by another: in the case of  $a-i > e$ , [-high, +low, +back]–[+high, –low, –back] is arbitrarily replaced by [-high, –low, –back].<sup>2</sup>

The AIU treatment of coalescence carries over directly to harmony. For example, height harmony consists in the spreading of [A] from a harmonic trigger to a vowel composed of [I] or [U] (cf. Goldsmith's (1985) analysis of the general Bantu pattern). Diphthongisation is simply the reverse of this effect: examples such as  $e > ai$  and  $o > au$  (cf. the history of English) manifest the breaking up of a complex vowel's components into a sequence of two simplex vowels.

To summarise: a range of phonological facts relating to the systemic, reductive and assimilatory behaviour of vowels supports the conclusion that certain vowels are complex in the sense that they are composed of other vowels.

## **4 Vowel patterns in the signal**

### **4.1 Elemental patterns**

Guided by the phonological patterning just reviewed, we will now propose definitions of the categories [A], [I] and [U] which will allow us to phonetically model the composition of complex  $e$  and  $o$  in terms of simplex  $a$ ,  $i$ ,  $u$ .

Before getting down to the specifics, let us make explicit a close parallel we are seeking to draw with compositionality in consonants. There is a clear sense in which consonantal lenition — like vowel reduction, often neutralising in effect — degrades phonetic information. Take for example the salient spectral discontinuities that provide cues to the phonological identity of a plosive in a VCV sequence, such as abrupt change in amplitude, noise burst, rapid formant transitions and  $F_0$  perturbation. Fewer of these cues are present in lenited reflexes, none of them in a vocalised reflex. Elsewhere we have shown how this phonetic impoverishment can be related to a categorial reduction in phonological representations (Lindsey & Harris 1990, Harris & Lindsey 1995, Harris, to appear).

When expressed in terms of standard feature classifications, it is not clear that phonetic-informational asymmetries of this order might also be found in vowel neutralisation. That is, there is no immediately obvious sense in which the outputs of vowel reduction could be said to be informationally more

impoverished than unreduced counterparts. For example, SPE-style tongue-body representations of mid peripheral *e* and centralised *ə* both require three feature values: [–high, –low, –back] versus [–high, –low, +back]. Specifications of this type grant equal informational status to all vowels.

Expression in terms of formant values also fails to suggest any informational asymmetries. As is evident from Table 1, vowels which the phonology tells us are simplex have the same amount of formant-value information as vowels which the phonology tells us are complex.

-----  
Table 1 about here  
-----

Quantal theory (Stevens 1972, 1989) might be turned to for an analysis which captures the special status of the corner vowels, specifically in terms of psychoacoustic salience related to formant convergence. However, this theory fails to account for the apparent compositionality relations which phonological patterning suggests — for example, the composition of *e* in terms of *i* and *a*. The quantal nature of the corner vowels derives from the salient manner in which F2 and F3 converge in *i* and F1 and F2 converge in both *a* and *u*. But this perspective actually obscures the compositionality of mid vowels: *e*, for example, cannot have an F2 which is simultaneously merged with F3 (as in *i*) and with F1 (as in *a*). The quantal characteristics of *i* and *a* are in fact both missing from *e*.

We now present an alternative view of vowels, according to which speaker-hearers extract three basic patterns from vocalic speech signals. We consider the internalised form of these patterns, that is the three basic auditory images, to be the elements which we notate as [A], [I], [U]. These three patterns will be shown to embody in a sense the quantal characteristics of the three corner vowels; but we go beyond quantal theory in suggesting that languages may use these patterns not only alone but also in combination.

These three elements are precisely analogous to those which speaker-hearers extract from obstruent speech signals, such as frication noise and abrupt amplitude-change (notated [h] and [ʔ] respectively), and which likewise can occur both alone and in combination (Harris & Lindsey 1995). This approach allows us to demonstrate that the consequences of vowel neutralisation are informationally parallel to those of consonant lenition.

Unlike SPE-style features, each pattern may occur alone in speech: the

elements are independently pronounceable, requiring nothing akin to the filling-in of redundant features (Harris & Lindsey 1995). The solo pronunciations of the categories [I], [A], [U] are the three corner vowels *i*, *a*, *u*.

The elementary auditory images can be notated either with the capital-letter symbols [A], [I], [U] or, more transparently, with iconic symbols or with the mnemonics mAss, dIp and rUmp (see Figure 1).

-----  
Figure 1 about here  
-----

In the case of [I], the internal mental category (or auditory image) corresponds to a pattern in external signals consisting of energy distributed to the top and bottom of the vowel spectrum (that is, approximately 0-2.5kHz for men and 0-3kHz for women), with a trough or dip in between (Figure 1a). In [A], the mental category/auditory image corresponds to a converse pattern in the external signal, a central mass with troughs at top and bottom of the vowel spectrum (Figure 1b). In [U], the internal category/auditory image corresponds to an external marked skewing of acoustic energy to the lower half of the spectrum (over and above the  $-6\text{dB}$  per octave downward slope characteristic of all vowels; Figure 1c).

It is now possible to define *e* as comprising two of these patterns, namely dIp and mAss (see Figure 2a). Acoustically, *e* shares with *i* the clear energy dip between F1 and F2. We model this as the presence of the elementary image, dIp. But in *e*, F1 and F2 are closer together than in *i*, such that the vowel's energy is massed towards a central spectral region, with troughs at top and bottom of the frequency range. We model this as the presence of an elementary mAss pattern in addition to the dIp pattern. Both auditory images are activated by this vowel.

-----  
Figure 2 about here  
-----

By the same token, *o* exhibits a rUmp pattern, since its energy is markedly skewed to the lower part of the frequency range. On the other hand, the peak energy is far enough above the bottom of the frequency range to constitute a mAss, with troughs above and below (see Figure 2b). Both auditory images are

activated.

Now consider the signal characteristics of centralised vowel quality, such as occurs in centripetal reduction. Acoustically, schwa has equally spaced formants; that is, it exhibits no merged-formant spectral peaks (see Figure 3). In terms of the key vowel patterns we have identified, schwa lacks all of the following: a mid-frequency dip in energy, a massing of its energy in the central spectral region, and a pronounced skewing of energy towards the rump of its spectrum. In other words, schwa is informationally empty. This is consistent with the notion in AIU theory that schwa is the phonetic expression of a nucleus devoid of segmental specification.

-----  
Figure 3 about here  
-----

We are now in a position to offer a unified treatment of both centrifugal vowel neutralisation (manifested in peripheral raising and lowering) and centripetal reduction (centralisation): both simultaneously involve a loss of internal categorial and external signal information.

#### **4.2 Internal–external isomorphy**

Let us now revisit the question of how the elements or auditory images proposed here differ from SPE-type features.

Over several decades' use, square brackets and plus/minus signs have taken on a kind of inherent authority, as if the physical expression of SPE-type features could be taken for granted. Let us dispel this assumption at once: acoustic cues to SPE-type features have not been well established. Formulae like '[+anterior]' remain what they always were — notational labels for conjectural generalisations over bodies of observed speech behaviour.

Like SPE-type features, elements are conjectural generalisations over data. But there are at least three differences. Firstly, we believe elements capture generalisations more elegantly, such as the vowel neutralisation facts outlined in §3. Secondly, elements may occur in isolation; that is, each is independently pronounceable without any need of redundancy fill-in machinery. Thirdly, elements allow far greater isomorphy between the conjectured internal objects and the external phenomena which cue them.

It is a reasonable assumption that the informational content of sounds is

directly encoded in phonological representations. (In fact, there is a good case for saying that this is the whole point of phonological representations.<sup>3</sup>) In terms of the patterns defined here and motivated on the basis of phonological evidence reviewed in §3, corner vowels can be said to carry less information than mid vowels. It makes sense to conclude that the proposed patterns correspond directly to phonological categories. Unlike orthodox feature theory, the model requires no adaptor mechanism to translate between a set of auditory-acoustic terms and a non-matching set of phonological terms. The isomorphism between the internalised vowel categories and external patterns in the signal that the AIU model establishes is unprecedented in the annals of recent feature history.

We feel justified in using our element terminology ambiguously, to refer (i) to the inner objects which, among other things, make up lexical addresses and (ii) to characteristics of external acoustic signals which cue the inner objects. A useful analogy can be drawn with astronomical constellations. What is Orion? Orion is an ambiguously used term referring both to a mental concept and to a grouping of physical stars — a physical object, in so far as it is perceived as such. We represent Orion on the page either with the word **Orion** or with a stylized graphic of points and lines somewhat resembling an optical or photographic impression of the physical sky. Our elements are analogous. Each is an ambiguously used term referring both to a mental object and to a patterning of acoustic energy — a physical object, in so far as it is perceived as such. We represent it on the page either with a capital letter or with an icon somewhat resembling an aural or spectrographic impression of the physical energy.

Thinking back to an SPE-type formula such as ‘[+anterior]’ now highlights the difference between such objects and our elements. It has always been assumed that SPE features have both mental and non-mental referents. To be sure, the non-mental referent of, say, [+anterior] could be represented on the page by a stylized graphic, presumably a sketch of the front end of the mouth. But the front end of the mouth, we suggest, has no signifying value whatever.

## **5 Alternative approaches to vowel patterning**

**5.0** In this section, we consider and reject three alternatives to the account of vowel patterning presented in the previous section, one based on vision (§5.1), one on articulatory features (§5.2), and one on formant features (§5.3).

## 5.1 Vision

One way in which *a*, *i*, *u* might be considered simpler than *e* and *o* is in terms of lip shape, schematised in Figure 4. Of the various labial configurations associated with the five canonical vowels, the three associated with *a* (box shape), *i* (slit shape) and *u* (round shape) can reasonably be considered maximally distinct. Plausibly, these vowels involve relatively simple processing in visual perception and, despite the apparently greater displacement of the lips from a rest position, articulation. Informally, the postures are respectively ‘as open as possible’, ‘as spread as possible’, and ‘as rounded as possible’. Since *e* and *o* are intermediate between the three extremes of *a*, *i*, *u*, they may involve subtler and hence more complex visual processing.

-----  
Figure 4 about here  
-----

One reason for at least considering the possibility that vision influences the patterns of vowel neutralisation is that speech-reading by eye is known to play a role in speech perception, as demonstrated for example by the McGurk effect (McGurk & MacDonald 1976). The interaction of vision and audition is further demonstrated by more recent research showing that, when subjects see speech without hearing it, their auditory cortex is nonetheless activated (see for example Calvert *et al.* 1997).

Against this must be weighed the consideration that speech is entirely intelligible without visual input and that the congenitally blind acquire normal phonology. These facts demonstrate that vision could play no more than a subsidiary role in shaping the vowel patterns under discussion here, at best enhancing the primacy of *a*, *i*, *u*.<sup>4</sup>

## 5.2 Articulation

Researchers have sought to explain consonantal lenition in unstressed syllables as resulting from target undershoot, on the grounds that unstressed syllables are characterised by shorter duration and less extreme articulatory movements than stressed counterparts (see de Jong 1998 for recent discussion and references). Extended to vowel neutralisation, this approach might be considered consistent with the occurrence of centripetal patterns of reduction: the trajectories

followed by the tongue in the production of centralised reflexes are shorter than those followed in peripheral vowels (Fourakis 1990). But it makes completely the wrong prediction about centrifugal vowel neutralisation. Given the more extreme dorsal manoeuvres involved in the articulation of the corner vowels, it is mid vowels that would be expected to show up as the preferred peripheral reflexes of neutralisation, precisely the opposite of what we find.

In any event, this account fails to explain why the same neutralisation effects can occur quite independently of stress prominence, as noted in §3 above.

This specific failing of the target-undershoot account is symptomatic of a more general inability of articulation-oriented approaches to model vowel neutralisation adequately. One problem, already alluded to in §4, centres on the failure of articulatory features to express the informational asymmetries between neutralised and unneutralised vocalic reflexes.

The case for persisting with an articulation-based approach to vowel categorisation has been made with renewed vigour in the recent literature, especially with regard to the use of scalar height features. If accepted, these must then be implicated in the specification of both centrifugal and centripetal vowel neutralisation. Earlier advocacy of such features (e.g. Ladefoged 1971) was widely resisted, at least as phonological classifiers, on the grounds that they fail to capture the non-continuous behaviour of vowel-height contrasts in such phenomena as chain shifts, harmony and, as we have already seen here, positional neutralisation itself.

The upturn in the fortunes of scalar vowel height has coincided with the recent proposal that what counts as phonologically distinctive in a particular grammar emerges as a result of phonetically fine-grained features being coarsely chunked by ranked constraints (Flemming 1995, Hayes 1996, Kirchner 1997). Kirchner (1997) assumes the existence of a unitary analogue dimension of tongue height that can be finely digitised for the purpose of feature classification. There is, however, no reason to take the articulatory reality of this dimension for granted.

Scalar vowel height categorisations are founded on the notion of ‘the highest point of tongue’. In fact, enough is known about vowel production for us to conclude that this is no more than a descriptively convenient fiction. Vowel quality is determined by the overall volume and geometry of the vocal tract, themselves determined by a combination of factors — the positioning of the tongue by the extrinsic muscles, the shaping of the tongue by the intrinsic muscles, the positioning of the lower mandible, overall lip shape, the shape of the pharynx and so forth (see Perkell 1997 for a literature summary). The

mutually influencing nature of these dimensions is confirmed by the fact that there can be considerable variation in the way a particular quality is executed, not just across languages and speakers but also within the speech of a single speaker (see Lindblom *et al.* 1979 and, for further references, Lieberman & Blumstein 1988: 162 ff.). Nowhere in this overall scheme of vowel production is there evidence of an independent dimension of tongue height functioning as a unit of central control in speech.

In exploiting the concept of vowel height, phonologists have been guilty of the same misuse of pseudo-articulatory labels as is perpetuated in practical phonetics by the Cardinal Vowel system. Vowel ‘heights’ are to be taken no more literally than the ‘slenderness’ of vowels or the ‘darkness’ of laterals. Employing such notions as descriptive conveniences should not lull us into the misapprehension that these are real categories in phonology-to-speech mapping.

Much more fundamental, however, are the grounds for doubting whether an articulatory approach is the appropriate way to model not just vowel patterning but phonological categorisation in general. The reasons are hardly a secret, although they have been largely ignored in the phonological literature since SPE made the switch away from Jakobsonian acoustic features. The main problem, as noted by Saussure, Sapir and indeed Jakobson (1968) himself, arises from the fundamental asymmetry between speech production and auditory perception, by virtue of which the former is parasitic upon the latter.

It is well established that speech perception precedes speech production throughout the language acquisition of normally hearing infants. Moreover, the congenitally deaf do not learn to produce speech normally, while those born with even the severest obstacles to speech production will, in the absence of cognitive deficit, acquire normal speech perceptual ability. Further, when individuals become deaf after the acquisition of phonology, their speech production suffers immediately. So does that of hearing adults under conditions of distorted auditory feedback.

The conclusion towards which such facts inexorably point is that articulation has as its *raison d’être* the production of pre-existing auditory targets. In spite of this, among phonologists the notion has predominated that the phonetic definitions of phonological entities are primarily or exclusively articulatory. This view reaches its apotheosis in the work of Bromberger & Halle (1988, this volume), who claim that phonological forms represent articulatory intentions and that phonological derivation is an aspect of speech production itself.

The fact remains that the first and necessary experience of humans for the acquisition of phonology is the hearing of speech sounds.<sup>5</sup> Articulatory competence is secondary — developmentally and epistemologically.

### 5.3 Acoustics

Of course, acoustic or auditory feature definition has not been completely neglected in recent phonological theory. In phonetically driven constraint-based theory, for example, there have been moves to rehabilitate auditory-acoustic features, albeit only as co-occupants of the categorial roster with articulatory features — the validity of which continue to be taken for granted (see Flemming 1995, Boersma 1998).

Which returns us to the proposal of Beckman (1997), Flemming (1995) and others that positional vowel neutralisation can be explained in terms of functional constraints originating in auditory processing. At least in languages where amplitude is one of the physical correlates of stress, unstressed syllables are less audible than stressed, especially in quiet speech or under conditions of background noise. Because of this, it would be reasonable to expect unstressed syllables to carry less functional load and thus exhibit fewer contrasts. However, like the articulatory-strength account touched on above, this fails to explain why neutralisation also occurs independently of stress prominence. Moreover, even amongst languages in which vowel reduction is stress-conditioned there are some in which amplitude is apparently not one of the physical correlates involved; examples include Tamil and Malayalam (Mohanam 1986: 112).

The specific ‘auditory’ features which Flemming (1995) employs for the representation of vowel quality are in fact scalar categorisations of individual formant values. Like SPE articulatory features, these fail to give expression to the informational asymmetries between reduced and unreduced vowels. Compare, for example, how a mid peripheral and a centralised vowel are characterised in terms of F1/F2 features of the type proposed by Flemming:

(3) (a) <i>e</i>	(b) <i>ə</i>
$\begin{bmatrix} \text{-lowest F1} \\ \text{-low F1} \\ \text{-high F1} \\ \text{-highest F1} \\ \text{-lowest F2} \\ \text{-low F2} \\ \text{+high F2} \\ \text{+highest F2} \\ \text{:} \end{bmatrix}$	$\begin{bmatrix} \text{-lowest F1} \\ \text{-low F1} \\ \text{+high F1} \\ \text{-highest F1} \\ \text{-lowest F2} \\ \text{-low F2} \\ \text{-high F2} \\ \text{-highest F2} \\ \text{:} \end{bmatrix}$

If every vowel is defined in terms of the frequency of individual formants, no vowel can be said to bear any more or less information than another.<sup>6</sup>

Several decades of acoustic-phonetic research have generated valuable

numeric data detailing typical resonant frequencies in the speech of men, women and children. But it would be misguided to assume that truly auditory representations in phonology should contain such data in relatively undigested form. The human listener is not equipped with a formant-tracking device. Vowel quality is perceived on the basis of gestalt spectral patterns rather than the precise centre-frequencies of formants (Lindblom 1986) — a point with which the element proposal in §4 is in obvious accord.

## 6 Conclusion

It is astonishing that mainstream generative interest in the way humans phonologically parse speech signals became sidelined once the Jakobson, Fant & Halle (1952) feature set had been usurped by the articulatory set ushered in by SPE. With certain notable exceptions, recent phonological theory and speech-recognition technology have been virtually estranged, most phonologists retreating from acoustics and auditory perception to the reassurance of simplistic articulatory or pseudo-articulatory labels.

The enduring popularity of articulation-based features is hard to explain. It would probably be only mildly unfair to put it down to the stereotypical image of a speech chain which commences with an articulatory act or to the accident of pedagogical and technological history whereby elementary courses on phonetics begin with vocal-tract anatomy rather than with spectrography or the auditory system.

Persisting with articulatory features can only hinder progress in understanding the sound facet of linguistic signs, surely the very essence of phonology. To repeat a point made earlier: linguistic information is projected by means of articulations but is not embodied in them. The information is constituted by sound patterns, which are extracted from the speech signal by listeners and used as targets by talkers.

The vowel patterns mAss, dIp and rUmp are motivated firstly by phonological investigation which identifies cross-linguistic regularities in vowel systems and processes. They are proposed as inner, mental objects — auditory images. Although our compositional analysis allows an unprecedented isomorphy between the internal and external aspects of speech, it should be emphasised that the elements are not ontologically independent physical phenomena inhering in sound waves. Rather, they constitute an aspect of the specifically human way with sound. We may reasonably assume that an organism different from ourselves would not parse human speech in like manner — just as the signals used by crickets, toads and whales are semiologically opaque to us, when perceptible at all.

How does the compositional parsing of vowels arise? We remain

open-minded about the explanatory factors that future research will identify, whether internal and/or external and, if internal, whether central or peripheral. It might be that mAss, dIp and rUmp are components of universal grammar — though the very isomorphy between these elements and the external environment would rule this possibility out if universal grammar is viewed as a radically internal device that has evolved in isolation from the outside world (see the chapters by Burton-Roberts and Carr in this volume). Alternatively, it might be that elements form part of some more generalised genetic endowment, or that they are posited de novo by learners through some interaction between external signals and general receptive capabilities.

Whether there turns out to be an immaculate phonological core or not, ‘phonological theory’ can always serve as a name for the enterprise which seeks to identify and explain patterning in speech sound.

## References

- Applegate, R. B. (1972). Ineseño Chumash grammar. PhD dissertation, UC Berkeley.
- Anderson, J. M. & C. Jones (1974). Three theses concerning phonological representations. *Journal of Linguistics* 10. 1-26.
- Bafile, L. (1997). L'innalzamento vocalico in napoletano: un caso di interazione fra fonologia e morfologia. In L. Agostiniani (ed.), *Atti del III Convegno Internazionale della Società Internazionale di Linguistica e Filologia Italiana*, 1-22. Napoli: Edizioni Scientifiche Italiane.
- Beckman, J. N. (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14. 1-46.
- Boersma, P. (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. LOT dissertations 11. The Hague: Holland Academic Graphics.
- Bromberger, S. & M. Halle (1988). Why phonology is different. *Linguistic Inquiry* 20. 51-70.
- Calvert, G. A., E. T. Bullmore, M. J. Brammer, R. Campbell, S. C. R. Williams, P. K. McGuire, P. W. R. Woodruff, S. D. Iversen & A. S. David (1997). Activation of the auditory cortex during silent lipreading. *Science* 276. 593-596.
- Clements, G. N. & S. R. Hertz (1991). Nonlinear phonology and acoustic interpretation. *Proceedings of the XIIth International Congress of Phonetic Sciences*, Vol 1/5, 364-373. Provence: Université de Provence.
- Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (eds.), *Universals of human language, vol. 2: phonology*, 93-57. Stanford: Stanford University Press.
- Flemming, E. (1995). Auditory representations in phonology. PhD dissertation, UCLA.
- Fourakis, M. (1990). Tempo, stress, and vowel reduction in American English. *Journal of the Acoustical Society of America* 90. 1816-1827.
- Goldsmith, John A. (1985). Vowel harmony in Khalkha Mongolian, Yaka, Finnish and Hungarian. *Phonology* 2. 251-274.
- Hale, M. & C. Reiss (1999). Substance abuse and dysfunctionality: current trends in phonology. To appear in *Linguistic Inquiry*.
- Harris, J. (to appear). Release the captive coda: the foot as a domain of phonetic interpretation. *Laboratory Phonology* 6.
- Harris, J. & G. Lindsey (1995). The elements of phonological representation. In J. Durand & F. Katamba (eds.), 34-79. *Frontiers of phonology: atoms, structures, derivations*. Harlow, Essex: Longman.
- Hayes, B. (1996). Phonetically driven phonology: the role of Optimality Theory and inductive grounding. *Proceedings of the 1966 Milwaukee Conference on Formalism and Functionalism in Linguistics*.
- Hyman, L. M. (1998a). The limits of phonetic determinism in phonology: \*NC revisited. Ms, UC Berkeley.
- Hyman, L. M. (1998b). Positional prominence and the 'positional trough' in Yaka. *Phonology* 15. 41-75.

- Hyman, L. M. (1999). 'The historical interpretation of vowel harmony in Bantu.' In J. M. Hombert & L. M. Hyman (eds.), *Recent advances in Bantu historical linguistics*, 235-295. Stanford, CA: CSLI.
- Jakobson, R. (1968). *Child language, aphasia and phonological universals*. Translated by A. Keiler. The Hague: Mouton.
- Jakobson, R., G. Fant & M. Halle (1952). *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.
- Jong, K. de (1998). Stress-related variation in the articulation of coda alveolar stops: flapping revisited. *Journal of Phonetics* 26. 283-310.
- Kamprath, C. K. (1987). Suprasegmental structure in a Raeto-Romansch dialect: a case study in metrical and lexical phonology. PhD dissertation, University of Texas at Austin.
- Kirchner, R. M. (1997). Contrastiveness and faithfulness. *Phonology* 14. 83-111.
- Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago, IL: University of Chicago Press.
- Lieberman, P. & S. E. Blumstein (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (eds.), *Experimental phonology*, 13-44. Orlando: Academic Press.
- Lindblom, B., J. Lubker & T. Gay (1979). Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming by predictive simulation. *Journal of Phonetics* 7. 147-162.
- Lindsey, G. & J. Harris (1990). Phonetic interpretation in generative grammar. *UCL Working Papers in Linguistics* 2. 355-69.
- McGurk, H. & J. MacDonald (1976). Hearing lips and seeing voices. *Nature* 264. 746-748.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
- Mtenje, A. A. (1985). Arguments for an autosegmental analysis of Chichewa vowel harmony. *Lingua* 66. 21-52.
- Mohanan, K. P. (1986). *The theory of Lexical Phonology*. Dordrecht: Reidel.
- Ohala, J. J. (1992). Bibliography. *Language and Speech* 35. 5-13.
- Palmada Félez, B. (1991). La fonologia del català I els principis actius. Doctoral dissertation, Universitat Autònoma de Barcelona.
- Penny, R. J. (1969). Vowel-harmony in the speech of Montes de Pas (Santander). *Orbis* 18. 148-166.
- Perkell, J. S. (1997). Articulatory processes. In W. J. Hardcastle & J. Laver (eds.), *Handbook of phonetic sciences*, 333-370. Oxford: Blackwell.
- Sapir, E. (1921). *Language*. New York: Harcourt, Brace & World.
- Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In P. B. Denes & E. E. David Jnr. (eds.), *Human Communication: a unified view*, 51-66. New York: McGraw Hill.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics* 17. 3-46.
- Urua, E. E. (1990). Aspects of Ibibio phonology and morphology. PhD dissertation, University of Ibadan.
- Zoll, C. (1998). Positional asymmetries and licensing. Ms, MIT. ROA-282-0998.

	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
<i>i</i>	300	2200	3000
<i>e</i>	500	1800	2700
<i>a</i>	700	1200	2500
<i>o</i>	500	900	2400
<i>u</i>	300	600	2300
<i>ə</i>	400	1400	2400

Table 1. Ball-park formant frequencies of six vowels (adult male).

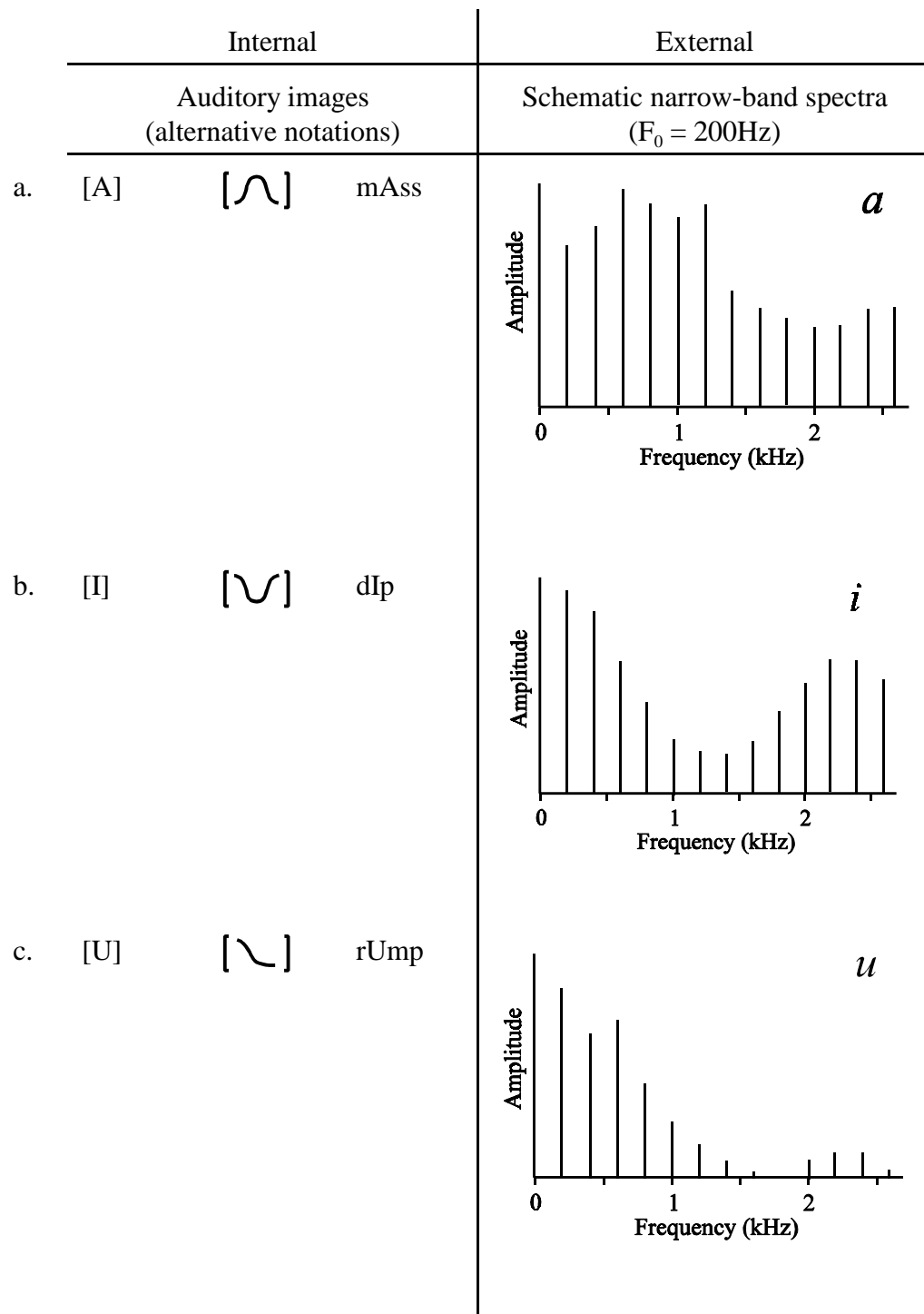


Figure 1. The three basic elementary patterns (auditory images) for vowels, and schematic spectra for three vowels, each of which exhibits only one elementary pattern.


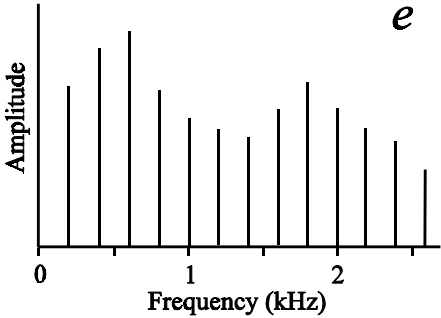

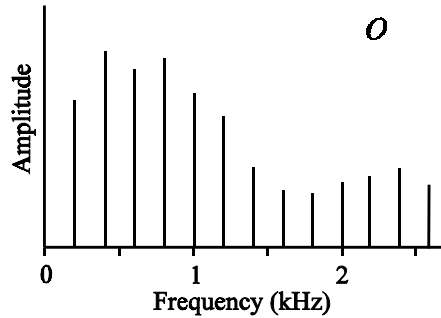
Internal			External
Auditory images (alternative notations)			Schematic narrow-band spectra ( $F_0 = 200\text{Hz}$ )
a.	[I, A]		 <p style="text-align: right;"><i>e</i></p>
b.	[U, A]		 <p style="text-align: right;"><i>o</i></p>

Figure 2. Composite elementary patterns (auditory images) and schematic spectra for two vowels, each of which exhibits two elementary patterns: (a) *e*, (b) *o*.

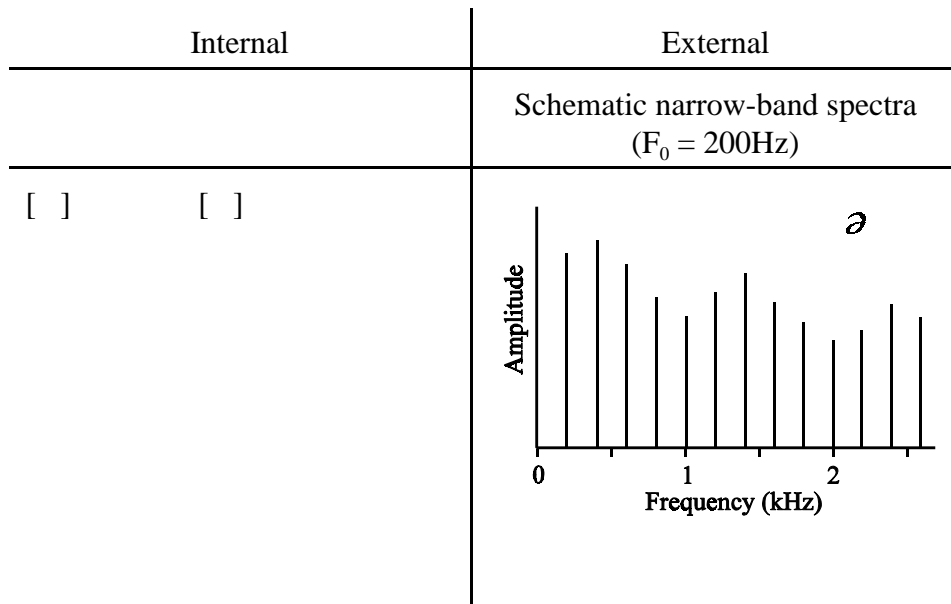


Figure 3. Categorical non-representation and schematic spectrum of a schwa vowel exhibiting none of the three elementary patterns.

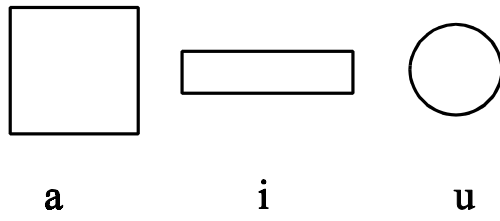


Figure 4. Schematic lip shapes.

## Notes

- 1 We are indebted to Neil Smith, Bencie Woll and the editors for their valuable comments on earlier drafts.
- 2 As an alternative articulatory account, not directly expressible by means of SPE-type features, it might seem appealing to think of coalescence as reflecting a compromise between the extreme dorsal gestures involved in the production of opposing corner vowels. The implementation of this idea is problematic, based as it is on an impressionistic notion of tongue height that has no foundation in articulatory reality. We expand on this point below.
- 3 This position is of course at variance with Hale & Reiss's (1999) view of phonology as being modality-independent (see also van der Hulst, this volume). However, taking phonology to be specifically targeted on speech does not prevent us from subsuming it under a grander study of all the modalities associated with language (including deaf sign and writing) — a branch of linguistic semiology which would concern itself with the informational content of all relevant perceptible physical media.
- 4 Visual information is of course the very stuff of sign language. The elements of sign phonology are internalised visual patterns.
- 5 Obviously, the hearing of speech sounds is not a prerequisite for the acquisition of sign language phonology, which presumably requires visual exposure to sign patterns. Regardless of modality, perception precedes production.
- 6 This is not to say that the perceptually salient convergence of formants cannot be expressed by means of features such as those in (3). It can, specifically by conjoining pairs of constraints which refer to particular formant values (Flemming 1995: ch 3). However, any pair of formant specifications can be potentially linked in this way, thereby missing the point that it is only certain combinations that contribute to the informational profile of vowels.