Available online at www.sciencedirect.com
ScienceDirect
Lingua 117 (2007) 645–665
Lingua
www.elsevier.com/locate/lingua

Microvariation, variation, and the features of universal grammar

Mark Hale [a,1], Madelyn Kissock [b,*], Charles Reiss [a,2]

[a] *Department of Classics, Modern Languages and Linguistics, 1455 de Maisonneuve Boulevard West, Concordia University, Montréal, Que., Canada H3G 1M8*
[b] *Linguistics Department, 329 O'Dowd Hall, Oakland University, Rochester, MI 48309, USA*

Received 20 January 2005; received in revised form 30 March 2006; accepted 30 March 2006
Available online 27 June 2006

Abstract

Accounting for the tremendous diversity of acoustic realizations of what linguists label as the 'same' sounds (featurally and/or implicitly through use of IPA symbols) cross-linguistically is a challenging task. Since a certain amount of this diversity is directly reflected in the input data presented to acquirers, and since determining its source is a key component of phonological learning, it is crucial for any account of early phonological acquisition that we develop an understanding of the sources of this diversity. This paper presents an explicit account of types and sources of such diversity, which we will term 'microvariation,' and makes a critical distinction between microvariation – a product of various extra-grammatical mechanisms and properties – and phonology. Differences in acoustic realizations due to differences in phonological feature representations, which we will call 'variation', should, we maintain, be treated as entirely separate phenomena (although they clearly produce cross-linguistic diversity, as well). A close examination reveals that variation (feature-based distinctions) can be obscured by microvariation and that microvariation can be mistakenly taken as representing variation. Teasing these apart provides us with a clearer understanding of the nature of the grammar and its acquisition, as well as of its interaction with other systems.

Keywords: Phonological features; Acquisition; Universal grammar (UG)

* Corresponding author. Tel.: +1 248 370 2174; fax: +1 248 370 3144.
E-mail addresses: hale1@alcor.concordia.ca (M. Hale), kissock@oakland.edu (M. Kissock), reiss@alcor.concordia.ca (C. Reiss).
[1] Tel.: +1 514 848 2424x2758; fax: +1 514 848 8679.
[2] Tel.: +1 514 848 2424x2491; fax: +1 514 848 8679.

0024-3841/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.lingua.2006.03.009

1. Introduction

The advent of Optimality Theory has shifted the focus of most phonological theorizing from representational issues such as the nature and organization of features to the discovery of properties of the computational system that relate representations.[3] As such, much recent work has focussed primarily upon developing optimality-theoretic accounts for phenomena which had been treated within earlier, rule-based frameworks or which had not been discussed previously in any framework. In this paper, we would like to return to an area that has, in our view, been neglected for some time, although, explicitly or implicitly, it constitutes the base of all phonological theories—namely, the nature of the phonological features themselves.[4] As far as we know, today's lack of attention to phonological features is not the result of a general consensus that there are no outstanding research questions regarding features but rather is the natural course of events when new developments in other areas of phonology outshine topics which have already been discussed at some length and which, consequently, appear less new and intriguing. In this paper, we argue that we are not yet done with features and that there is good evidence from certain types of variation and 'microvariation,' that suggests that our current feature system is not rich enough to account for phonological systems and their acquisition.

The next section of the paper presents our background assumptions about what aspects of the cognitive system we believe can be grouped together and considered phonology (i.e., linguistic computation) versus phonetics (i.e., the acoustic or gestural score). The third section introduces a contrast between 'variation' and 'microvariation,' and presents an explicit categorization of apparent instances of microvariation. We conclude that section with a consideration of which types of apparent 'microvariation' might be relevant to the pursuit of phonological theory. In the fourth section, we focus our attention on 'Category D' microvariation—the type which we believe is of greatest significance for the pursuit of phonological theory. We present several previous accounts of the more problematic types of microvariation and present our own analysis of those cases. The final section of the paper concludes with a summary.

2. Background assumptions

Since our primary interest is in modelling aspects of the computational system (the grammar, especially phonology) which are the result (and thus the target) of phonological acquisition – in this case defining the primitives that operations are performed on – we begin by presenting our assumptions about the nature of this computational system. This includes a discussion of the division between phonology and phonetics and between linguistic and non-linguistic processing, as well as explicit definitions of the terminology we use. This is particularly important since there is not general agreement among researchers on these matters and there is considerable confusion introduced by lack of explicitness in terminology. The one point on which there *is* general agreement, at least among phonologists, is that 'phonology,' in its narrowest sense, consists of the mapping from underlying representation to surface representation. Keating (1988) explicitly addresses this issue and we follow her in both our use of terms and in our definition of 'phonology.' For Keating and for us, phonology involves only a feature-to-feature mapping and nothing else. Other researchers have defined phonology much more broadly and have extended it

[3] This regular alternating focus is the theme of Anderson's (1985) *Phonology in the Twentieth Century*: *Theories of Rules and Theories of Representations*.

[4] A notable exception to this general neglect is Steriade (2000).

well beyond the feature-to-feature mapping. Hammarberg (1976), for example, defines any aspect of pronunciation that involves cognition (for example, anticipatory co-articulation) as part of 'phonology.' For Hammarberg, then, a mapping of a set of features to a gestural score[5] or of an acoustic score to a set of features, because both require intention, are considered 'cognition,' and would be included in the definition of 'phonology.'[6]

We propose that mappings between dissimilar representational formats, such as from features to a gestural score, are performed by a pair of *transducers*. Transduction, in general, is a function which converts a representation in one representational 'alphabet' to a representation in a different representational alphabet.[7] Phonetic transduction is distinguished from phonological computation by the fact that it incorporates some type of *conversion* process—it changes one type of representation (featural, for example) into another type of representation (gestural score, for example). In our model, contra Hammarberg for example, phonological computations, unlike transduction, operate on only a single type of symbolic representation—features. Features are both the input to and the output of the phonology—phonological computations cannot convert features into other types of representations. Under any analysis, however, the incorporation of a transduction process of some type into the model of speech production seems inescapable, since there are no features actually present in the acoustic output of speech.

A further logical necessity, if phonological information is encoded by features, is the presence of two separate transducers—one for processing representations concerning perception and one for representations concerning articulation. Positing two transducers is suggested by the very different nature of perceptual versus articulatory processing. Both involve 'unidirectional' processing but the *direction* of processing is not the same in the two cases. Articulation demands transduction of features (the input) to some gestural score (the output), whereas audition requires transduction of a percept (the input) to features (the output).[8] In addition, we assume that there is an actual, physical difference between the mechanisms involved in perception and those involved in articulation and that dedicated transducers reflect this difference. We assume that these two transducers are innate and invariant—they are identical in all humans (barring some specific neurological impairment) and do not change over time or experience (i.e., they do not 'learn').[9] We will provide arguments for these particular claims in a later section.

Our model assumes strict modularity—no component can see 'inside' another component. Only the output of one module may be fed to another module and then only in the case of particular modules. So, for example, the output of the phonology is the input to the articulatory transducer, but the articulatory transducer does not feed its output to the perceptual transducer, nor vice versa. Thus the two transducers operate independently of one another and have no interaction. Following Hale and Reiss (2000a,b), our model necessarily divorces (phonological)

[5] We borrow the term 'gestural score' from Browman and Goldstein (1990) and extend it to the acoustic domain.

[6] While Hammarberg defines phonology more broadly than we do, we believe that his insightful distinction between cognitive and non-cognitive processes is an important one in understanding the issues under discussion here.

[7] We borrow this terminology from Pylyshyn (1984). Our use of the term is related to his but not identical with it. Pylyshyn's primary concern is the more general area of computation and cognition.

[8] We do not intend by this that there are *only* two transducers. We assume that transduction is a complex process which involves many different transducers. However, for our purposes here, only the highest level of transduction is immediately relevant—that of features to gestural score or perceptual score to features. It is for this reason that we use the term 'perceptual transducer' because we are concerned with higher level processing, not processing at lower levels such as the inner ear.

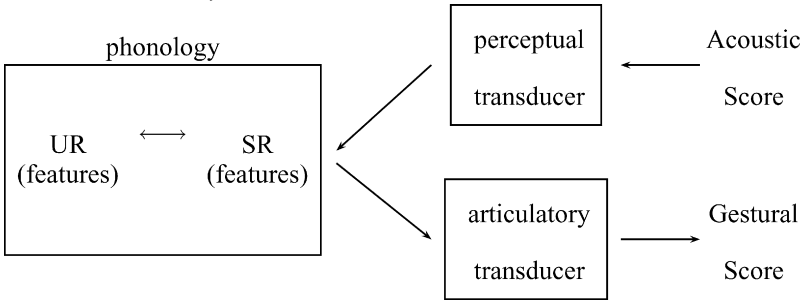
[9] This claim regarding innate, unchanging transduction is made solely for these two transducers, which take one type of symbolic representation and convert it to another type of symbolic representation. It is not a claim that motor skills, for example, are not learned or do not develop or mature over time—it is not a claim about motor skills at all.

features from both articulation and perception. Features are simply symbolic, 'substance-free,' primitives which are manipulated by the phonology and the transducers.[10] The very fact that two separate transducers are required – one for articulation and one for perception – forces the separation of features from any physical substance. Since what is mapped onto a single feature comes from two very different sources, this separation from the physical substance is a logical necessity—a single feature cannot, for example, be *both* derived from the muscle commands involved in raising the tongue body *and* from a neural impulse triggered by some portion of an acoustic wave (let alone be some actual property of the wave itself). Thus, we can consider the transduction process, too, as invariant in that the *relationship* or *mapping* between a particular feature bundle and a particular gestural score is a deterministic (and thus consistent) conversion process and, similarly, that the relationship or mapping of a particular perceptual input to a feature bundle is deterministic. Crucially, the features and their transduced output forms are different from one another.

Finally, we assume that phonological features, themselves, are universal in the sense of universal grammar (UG). A universal feature is not one that is found 'universally' but rather a feature, which is drawn from a universally available but finite inventory. We believe that the innateness of features follows directly from learnability arguments, among other things, and will present some specific arguments in support of this position when discussing Kingston and Diehl (1994) in a later section.[11] Any UG feature *may* be present in the actual mental representations of a particular instantiation of natural language but it is not clear that every feature *must* be present. Since the symbolic representations of natural language segments appear invariably to be feature *bundles*, it is actually the possible featural combinations that are responsible for the wide variety of sounds found in language, not the sheer number of features.[12]

The relevant mappings in our model of phonology and transduction are shown in (1) below.

(1) Phonology: UR ↔ SR
Transducer$_{perceptual}$: SR ← Acoustic Score
Transducer$_{articulatory}$: SR → Gestural Score



This model and the accompanying discussion have so far presented only what we believe to be computations and processes *specific to language*. As such, they represent only a portion of what

[10] The general proposal that phonological operations manipulate features goes back to Jakobson et al. (1963) and Chomsky and Halle (1968). Neither work proposed a totally 'substance-free' system, the former connecting features to acoustic properties and the latter to articulatory and acoustic properties.

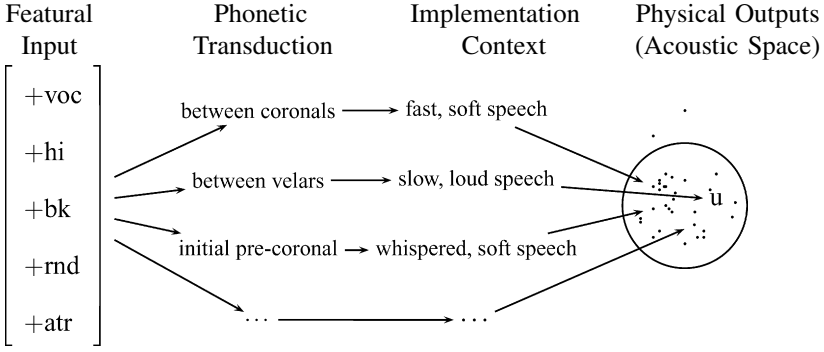
[11] For a view opposing the universality of features, see Pulleyblank (2001).

[12] We use the term 'bundles' loosely to represent groupings of features. We take no position here on the best way to represent these groupings.

makes up an individual's actual behavioral output or *performance*. We assume that between, for example, the gestural score and the point of actual physical output, there can be input or modification from many other *non-linguistic* facets of cognition that determine amplitude, speech rate, affect (e.g., tone of voice), and other situational effects. Crucially, none of these post-gestural-score additions contained in the physical output appears to be utilized by the linguistic computational or processing systems; therefore a sharp distinction between the two types of processing, linguistic versus non-linguistic, is indicated.[13]

While we believe that features are purely formal representations and that the transducers are invariant across speakers and deterministic in the way they perform conversions, we still predict that there will be differences in the measurable, physical instantiations of any *single* feature bundle. These differences will have at least one of three possible sources (and will probably be due to more than one): (1) the articulatory transducer, although mechanical, implements features and feature bundles in a context-sensitive manner[14]; (2) there are within- and cross-speaker differences in physical attributes (e.g., sub-glottal pressure, size and shape of oral cavity); and (3) there are external, physical forces which are implicated in actual production (e.g., external air density). As a result, the output for a particular featural representation will correspond to multiple, physical instantiations (on separate occasions) which nevertheless typically fall within some reasonably well-defined *acoustic space*.[15] We schematically represent this potential chain of events leading to physical output in (2) below.

(2) Transduction and the acoustic space.



The above figure only schematizes articulation. However, the concept of multiple physical instantiations being related to a single featural representation is, of course, the same for perception.[16] In the case of perception, however, physically distinct inputs (from different occasions)

[13] As is well known, phonological acquisition routinely and successfully takes place in environments where there is great variability in the physical production of featurally identical speech sounds. For a summary of some of the research on this topic, see Jusczyk and Luce (2002).

[14] This includes whatever effects the implementation of one feature may have upon the implementation of another feature within a single feature bundle as well as co-articulation effects.

[15] We avoid using the term 'target' since it incorrectly suggests both that there exists a single, 'correct' physical target and that the transducer has frequent 'misses.' It is also true that acoustic spaces for distinct feature bundles may overlap depending upon a number of factors such as speech rate and coarticulation, among others. We are not claiming that the outputs of distinct feature bundles are mutually exclusive acoustic spaces.

[16] The phenomenon of mapping, multiple physically-distinct inputs to a single mental representation, which Pylyshyn (1984) labels 'equivalence classes,' appears to be true of many aspects of human cognition. The notion was discussed for phonology as early as Sapir (1933).

will be stripped of their context-dependent and idiosyncratic physical properties, and reduced to a single identical set of features. In both articulation and perception, the acoustic space is an idealization of some physically definable area. We hypothesize that only a change in features will produce any significant change in acoustic space. This is because the differences incurred through differences in context or physical attributes will remain, we assume, relatively constant independent of the make-up of the particular feature bundle. So, roughly, the particular features determine the *locus* of physical instantiations and non-featural attributes determine the cluster pattern around the locus. We turn now to a description of some of these acoustic spaces and the task of teasing apart the role of features, linguistic (i.e., featural) context, and non-linguistic contributions to the physical output.

3. 'Microvariation' vs. 'variation'

The fact that the physical (i.e., acoustic) properties of speech sounds do not map in a one-to-one manner to the perception of those same sounds, termed 'lack of invariance,' is a well known property of natural spoken language and one that has been much researched in the area of speech perception.[17] To develop an understanding of lack of invariance, we believe that it is crucial to distinguish between two distinct phenomena – *variation* and *microvariation* – which we will define as follows. *Variation* belongs to the realm of phonology and only involves differences in featural representations from one utterance to the next. Differences in featural representations may arise in two ways—a single UR may give rise to two or more SR's (a coronal stop and its flapped counterpart in English, for example) or two SR's may be featurally distinct not due to allophony, but rather because their URs are featurally distinct. *Microvariation*, on the other hand, we define as non-phonological, typically smaller-scale differences, introduced either by the transduction process, individual physical properties, or external physical events. Crucially, microvariation, under our definition, is never the result of a difference in featural representation. While this appears to be a fairly clear-cut distinction, we will see in what follows that there are both cases of microvariation 'masquerading' as variation and, conversely, variation masquerading as microvariation in the literature. These cases seem likely to reveal important aspects of phonology and, as such, will be the main topic of discussion in the next section.

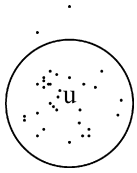
We have broken down types of microvariation into four categories labelled A, B, C, and D and further broken down category D into three subcategories. The categories are largely based on whether the microvariation is non-systematic or systematic, whether it is within-speaker or cross-speaker, and upon the source of the microvariation (external physical, individual physical, and so on).

3.1. Category A

Category A is distinguished by within-speaker, *non-systematic* microvariation. The schematic figure just illustrates a typical case of apparently random hits within an acoustic space (the [u] space here), the differing hits being due to small differences in the physical state of the speaker at particular points in time.[18]

[17] For a brief but interesting discussion of directions taken in such research, see Appelbaum (1996).

[18] By 'random', we intend merely that the physical factors involved (barometric pressure, amount of fluid in vocal tract, precise state of relevant tissues, etc.), as well as the interactions between these factors, are too complex to allow for a full formal modeling.



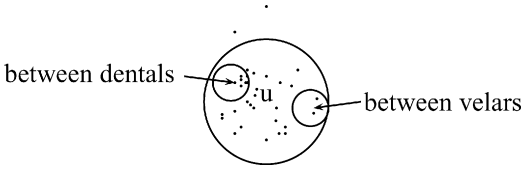
It should be noted that the effects of microvariation of the type outlined here and in Categories B and C are typically cumulative, and that all three types contribute to producing the 'acoustic space.'

3.2. Category B

This category covers cross-speaker microvariation which is *apparently systematic* because it is to be attributed to the (relatively constant) physical properties of the individual speakers. So, for example, different speakers have vocal folds of different thicknesses and lengths and oral cavities of slightly different shapes, all of which contribute toward the fact that they produce slightly different physical instantiations of a particular feature bundle. Stevens (1998:265) provides a nice illustration of this in the form of the contrast in spectral differences in high and low vowels produced by an adult female speaker and an adult male speaker.

3.3. Category C

Category C is defined by within-speaker microvariation in the form of *systematic* hits within the acoustic space due to the fact that the transducer must implement the same set of features in different syntagmatic contexts. The figure below shows a schematic example of the features of [u] being realized further forward when that [u] is between dentals than when it is between velars. Note that this is *not* due to the phonology (i.e., to a difference in features) any more than different realizations based on faster or slower speech rate is due to the phonology. It is the result of the transducer's 'implementation plan.'



Categories A, B, and C are all defined by different hits *within* a particular acoustic space. The cases described below in Category D, on the other hand, are defined by a significant difference in the acoustic space itself. Although we proposed earlier that such significant differences will normally be the product of an actual difference in featural representations, these cases are relatively complex and require a closer examination before making a final determination. It would be premature to label them as either cases *variation* or *microvariation* at this point, so we will avoid any initial categorization.

3.4. Category D

Category D is broken down into three subcategories. The common property throughout Category D is cross-speaker *systematic* differences which result in different acoustic spaces for what we take at the moment to be the same featural representations.

3.4.1. Category D1

In this subcategory, the acoustic space shows often dramatic, context-dependent variation because of *underspecification* in the Surface Representation (commonly known as 'phonetic underspecification').

As Keating (1988) showed, using the example of Russian [x] (no back feature in certain contexts) and English [h] (no oral features), the characteristic acoustic patterns for such sounds show gradual transition through the acoustic space of the feature or features in question (whereas fully specified SR's show more abrupt transitions). Cases of this type will be discussed in more detail below.

3.4.2. Category D2

These are cases of non-equivalent acoustic spaces, of which we have identified two subtypes.

(a) The acoustic spaces are overlapping, but not identical, as in the case of Danish [e] and Japanese [e].
(b) The acoustic space is broader for one speaker than for another speaker and the acoustic spaces are in a superset/subset relationship (based on area of hits) with the difference standing in seeming correlation with a difference in the relevant segmental inventories. The 'superset' acoustic space might be from a speaker with a three-vowel system and a relatively 'broad' hit area for [i] (for example), versus the 'subset' acoustic space being from a speaker with a larger vowel inventory and a correspondingly reduced acoustic space for hits for [i].

3.4.3. Category D3

Although the set relations here are similar to those of category D2(b) above, there is a crucial difference in 'motivation' for Category D3. In this case, the acoustic space is less broad for one speaker than for another speaker, but *without a contrast in inventory*. Maddieson and Wright (1995) report on the case of Amis—a three-vowel system which has quite 'narrow' acoustic spaces. The schematic illustrates the Amis case as contrasted with the quite broad acoustic spaces of many other three-vowel systems.

Amis [i]-acoustic space

'Typical' 3-vowel system [i] acoustic space

3.5. The linguistic irrelevance of categories A, B, and C

Categories A and B can both be traced to physiological properties and/or physical interference of various types. As such, they do not involve the phonology (no change in features) nor, in fact, any aspect of processing that is specific to language. We will not spend further time on them here.

Category C (coarticulation) can be accounted for by a combination of the effects of transduction and physiology/physical effects. While the phonology can systematically alter a featural representation in a specific (featural) context (e.g., [−nasal] to [+nasal] when adjacent to [+nasal]), recall that we consider these feature-based effects to be instances of *variation*, not of microvariation. Cases of the type outlined in category C are *not* the result of a change in featural representations and therefore are not variation, instead they are different physical realizations of

the *same* feature or set of features. Co-articulation (transitions from one articulatory gesture to another—from an oral vowel to a nasal consonant, for example) is contextually driven and will therefore be systematic, but it crucially does not involve modification of the featural representation by the phonology. We expect, then, to find some minimal nasalization on any vowel adjacent to a nasal consonant even if the featural representation for that vowel is no different from when the vowel is adjacent to a non-nasal consonant (unlike in English phonological vowel nasalization).

That systematic differences of this type may be ambiguous from an acquirer's standpoint between feature-based and non-feature-based sources is supported by a multitude of historical changes which instantiate some previously non-featural co-articulation effects as feature-based phonological processes—such as the case of English nasal vowels.[19] As with all of the other cases under discussion, co-articulation effects produce many different hits within an acoustic space. Since there is no defining line between acoustic spaces, such as where the space for nasalization of co-articulation stops and where that of feature-induced nasalization begins, it is up to the acquirer to deduce whether the overall evidence supports one or the other source. It is predicted that such ambiguity in the data may allow some acquirers to posit different phonological representations than those of their (adult) sources.

Note, again, that under our model and assumptions, any differences that arise from physiological properties of an individual speaker, physiological differences between speakers, and external, physical interference are *removed* from the data stream transmitted to the phonology (some, at least, of this stripped out information is sent elsewhere, to be interpreted by other cognitive systems). The transducer only provides symbolic, featural representations to the phonology—it has no way of telling the phonology that the person producing the utterance has a cold (nor could the phonology process that information even if it received it).

In the next section, we discuss the sources of the differences noted above, focussing on the (D) cases.

4. 'Category D' phenomena: a detailed consideration

As noted above, we have divided 'Category D' phenomena into several distinct subtypes. The theoretical analysis of the processes and systems which underlie these phenomena will be of particular interest to the development of phonological theory, so we will spend the rest of the paper discussing them.

4.1. 'Hidden' microvariation and variation

Keating (1988) convincingly argues for a distinction which draws a fine line between our cases of microvariation due to context discussed above as Category C and a separate type of contextually-driven difference (Category D1). Our cases of systematic microvariation due to context in C are the result of interactions arising from the mechanical realization of sequences of *fully specified* (along the relevant dimension) feature representations.

Keating's example is crucially different. In the case of Russian [x], where no context rules apply,[20] she argues that the output of the phonology is *not* a fully specified featural

[19] We would like to thank an anonymous reviewer for pointing out that such cases may also be ambiguous.

[20] Specifically, the SR for a form like /axi/, where a context rule has filled in the feature [-back] for the fricative, has a fully specified representation for [x]. A form like /ixa/, on the other hand, where no context rules apply, leaves the fricative underspecified for the feature [back].

representation—rather, it is a feature bundle underspecified for the feature [back]. (Note that this means it will *never* have any featural specification for [back] because the transducer does not 'fill in' features.) The result of this underspecification is that the physical output will just show a gradual transition from whatever the adjacent sounds' values for [back] are.

Further support for this type of output underspecification comes from the case of Marshallese vowels (see Hale, 2000), which we describe below. The Marshallese vowel system is quite striking. The 'surface' vowels are given below (where the 'tie' symbol, as in i͡u, represents a transition from one vowel to another, in this case, *i* to *u*).

(3)

i	ɯ	u	i͡ɯ	i͡u	ɯ͡i	ɯ͡u	u͡i	u͡ɯ
ɪ	ɤ	ʊ	ɪ͡ɤ	ɪ͡ʊ	ɤ͡ɪ	ɤ͡ʊ	ʊ͡ɪ	ʊ͡ɤ
e	ʌ	o	e͡ʌ	e͡o	ʌ͡e	ʌ͡o	o͡e	o͡ʌ
ɛ	ɐ	ɔ	ɛ͡ɐ	ɛ͡ɔ	ɐ͡ɛ	ɐ͡ɔ	ɔ͡ɛ	ɔ͡ɐ

As Choi (1992) demonstrates, there is a smooth transition between the back and round features of the segment to the left and the right of a 'tied' Marshallese vowel in every instance. As Bender (1968) showed, the most coherent phonological analysis of the Marshallese vowel inventory is one in which the vowels themselves bear no features along the dimensions back and round. That is, they differ from one another *only* along the height and ATR dimensions. We will use C^{j} to represent 'front' (i.e., palatalized) consonants, $C^{ɯ}$ to represent back non-round (i.e., velarized) consonants and C^{w} to represent back round (i.e., labialized) consonants. For the vowels not specified along the back or round dimensions, we introduce the symbols $[V_{HI}]$ for a [+high,−low,+ATR] underspecified vowel, $[V_{MID}]$ for a [−high,−low,+ATR] underspecified vowel, and $[V_{LO}]$ for an [−high,−low,−ATR] vowel, likewise underspecified.[21] Finally, we introduce the *necessary* distinction between the output of phonological computation (which we place between traditional square brackets) and the articulatory-acoustic output of the body (which we place between 'body' brackets). We can then represent schematically the treatment of the underspecified Marshallese vowel segments as follows[22]:

a. $C^{j}V_{HI}C^{w}$: /$t^{j}V_{HI}k^{w}$/ > [$t^{j}V_{HI}k^{w}$] > ‖t^{j}i͡uk^{w}‖ 'uncover an earth oven'

b. $C^{j}V_{MID}C^{ɯ}$: /$n^{j}V_{MID}t^{ɯ}$/ > [$n^{j}V_{MID}t^{ɯ}$] > ‖n^{j}e͡ʌ$t^{ɯ}$‖ 'squid'

c. $C^{j}V_{LO}C^{j}$: /$t^{j}V_{LO}t^{j}$/ > [$t^{j}V_{LO}t^{j}$] > ‖t^{j}ɛt^{j}‖ '*Lutjanus Flavipes*'

Note that this gives Marshallese what appears superficially to be a large and rather unique vowel inventory, whereas, in fact, for all grammatical (i.e., featural) purposes, the Marshallese inventory is quite small.

[21] We will not concern ourselves with the fourth height contrast in this paper.

[22] The reader can see that if one places each of the vowels between each of the types of consonant, the 'surface' vowel set cited above emerges—for details, see Hale (2000).

The case of /$t^jV_{LO}t^j$/ is particularly interesting. This vowel will show apparent steady-state realization in the [ɛ] space, much like English [ɛ]. If one considers only measured acoustic space in one's analysis of variation or microvariation, then the Marshallese [ɛ] and English [ɛ] appear to be identical and thus neither variation nor microvariation is present. Further analysis, however, reveals that this identity is purely superficial and that it actually obscures a significant difference between Marshallese [ɛ] and English [ɛ]. The front and non-round properties of Marshallese [ɛ] are entirely a product of the transducer implementing a feature bundle which is underspecified for back and round in a particular context (between palatalized consonants). English [ɛ], on the other hand, is the result of a transducer implementing a feature bundle that includes specific values for back and round.

If we look within Marshallese itself, we see that underspecification of this type can be classified as a source of *microvariation*. A high vowel in Marshallese has precisely the same featural specification regardless of whether it surfaces as [i], [ɯ], [u], [i͡ɯ], [i͡u], [ɯ͡i], [ɯ͡u], [u͡i], or [u͡ɯ]. The transducer's contextually driven implementation of this feature bundle is what determines the realization. It is because this is a product of transduction rather than a difference in features that this case falls under our definition of microvariation. Interestingly, if we look, instead, at the contrast between Marshallese and English vowels, we see that the apparent identity which holds between, for example, Marshallese [ɛ] and English [ɛ] is nevertheless actually a case of *variation*—it results from a *difference* in featural representations. Note that such phenomena give rise to serious problems for any neo-behaviorist approaches to phonology (i.e., those that only allow relationships between observed surface forms to be considered), since from the performance perspective the two vowels appear to be the same.

4.2. Accounting for non-equivalent acoustic spaces

The mismatch in acoustic spaces outlined in Categories D2 and D3 (overlapping, non-identical acoustic space and two types of superset/subset cases) have the properties of being *systematic*, *experience-driven*, and *speaker-independent* (meaning simply that they are not speaker-specific but instead seem to cluster around speakers who have shared some commonality of input during acquisition). At the same time, none of them appears to be context-driven. All of this suggests that the source does not lie in individual or cross-speaker physiology or in random, external, physical interference.

Several theories have been proposed to account for differences of this type. Many of these either explicitly or implicitly include some notion of 'expanding to fill the available space.' This idea has typically been appealed to in order to account for cases such as the D2(b) type where, for example, the acoustic space for vowels in a three-way contrast system is typically much broader than the acoustic space for vowels when the vowel inventory is larger (roughly five or more vowels assuming symmetric distribution). Stevens (1998:573–574) notes this especially with respect to vowels in context:

> The degree to which a vowel undergoes modification because of the influence of an adjacent segment depends on several factors. One contributing factor is the presence in the language of a contrasting vowel that differs minimally from the target vowel, so that the formant frequencies of this vowel are close to those of the target vowel. The allowed influence of context on the target vowel can be greater if the adjacent vowel is distant, but must be smaller if the shift in the formant pattern due to context can lead to misinterpretation of the vowel.

> Thus it is expected that the contextual modifications of vowels in a given language are different depending on the inventory of vowels in the language.

This description could, arguably, unify *both* cases discussed under D2 above. If the differing acoustic spaces for Danish [e] and Japanese [e] were actually dependent upon the (differing) acoustic spaces of adjacent vowels (Danish [i] versus Japanese [i], for example), then the notion of 'available space' could describe the differences between Danish [e] and Japanese [e]. Numerous problems seem to remain, however. First, the idea of expanding to fill available space predicts that languages which have the same number and type of vowels should have the same acoustic spaces for corresponding vowels. This is contradicted by the actual evidence, however. As discussed in Vaux (2002), Greek and Spanish both have seven-vowel systems which are typically analyzed as phonologically identical but the acoustic spaces for corresponding vowels in Greek and Spanish are different. A second problem for the available space idea is that there exist cases like Amis (Category D3), described in Maddieson and Wright (1995). Amis has only a three-vowel system but the three vowels do *not* fill the available space. Third, there is a circularity problem in defining 'available space' based only on other segments' space. That is, there is no 'starting point'—every vowel space will have some other vowel space dependent upon it. Finally, if all these problems were not sufficient evidence of the lack of relevance of the notion, we point out that there appear to be no cases in which a language actually makes use of maximal differentiation within the available space, even though maximal differentiation is implicitly or explicitly at the root of the available space hypothesis. For example, a vowel inventory whose members were maximally different would have a creaky [i], a plain [u], a nasal [a], and a long [e]. Note also that diachronic vowel mergers are not predicted under this hypothesis and yet they often occur.

In any event, neither this fairly general idea which is dependent upon the notion 'available space', nor other hypotheses regarding the same sorts of phenomena (Perceptual Magnet Effect (Kuhl, 1991); Quantal Theory (Stevens, 1972); Dispersion Theory (Liljencrants and Lindblom, 1972)) offer an actual explanation of *how* these acoustic spaces are associated with featural representations and how such an association might be acquired.[23] Scobbie (2001) approaches the problem from a production standpoint but also concludes that 'inventory-based' models fail to address a number of relevant phenomena.

Kingston and Diehl (1994), on the other hand, propose that both language-specific phonetic variation and phonetic variation due to context (both cases of microvariation under our definition) are the result of a learning process.

> We believe that phonetic implementation has acquired the appearance of an automatic process because it is so thoroughly overlearned. To acquire the fluency that makes phonetic implementation appear automatic, speakers must learn the appropriate control regimes for articulating each allophone of each phoneme. Similarly, listeners must learn to recognize the patterns of acoustic properties that characterize each allophone, as well as how the allophones correspond to phonemes. (pp. 420–421)

The Kingston and Diehl model, while it does address the issue from a learnability standpoint, lays a virtually impossible burden upon the learner, in our opinion.[24] The general process of

[23] For a thorough discussion and critique of such theories, especially Quantal Theory and Dispersion Theory, see Vaux (2002).

[24] Note that the idea that speakers must 'learn' to recognize patterns is counter-predicted by a large body of experimental work on infant perception which points to innate perceptual categorization of speech sounds.

assigning abstract features to acoustic or articulatory properties does not appear especially daunting until one examines in detail exactly what such a learning task would entail. At the least, if phonological features are purely symbolic arguments with only a *representational* not a *content* relationship to articulatory or acoustic properties, a learner will have no guidance in assigning features to either type of physical event. The result will be that a mapping of Feature A to Articulatory Gesture Y for one speaker will only by chance be identical to a mapping in another speaker's system. Such a difference turns out to be vacuous in terms of outputs of the two speakers, but it will entail dramatic differences in cognitive representations across speakers. There are, however, two far more serious problems with this approach. First, if a speaker must learn mappings between particular phonological features and particular aspects of, for example, the acoustic signal, the number of features will need to be dramatically increased. The 'lack of invariance' aspect of natural language suggests, in fact, that a speaker might require an infinite number of features to accommodate the micro-differences in acoustic signals because there are an infinite number of contexts-fast/slow, loud/soft, whispery/strident, etc. A second, related problem is that there is no mechanism in this model which will allow the acquirer to group hits within an acoustic space into a single featural representation. As a result, we would have no explanation for the systematic cross-speaker identification of slightly different acoustic signals as 'the same'—the phenomenon of categorical perception being a prime example of this. All of these also argue in support of an invariant transduction process, as proposed earlier in the paper.

Other accounts of the 'acoustic space' phenomena based on non-generative assumptions hypothesize that there is statistical learning based on probability distributions, see, for example, Pierrehumbert (2001). Such accounts are so divergent from ours in their underlying assumptions about the linguistic system that we cannot easily draw comparisons but, on the surface, they seem to share the drawbacks of, for example, the Kingston and Diehl (1994) approach.[25]

We believe, contra Kingston and Diehl inter alia, that learning must be based instead on innate knowledge of some (relevant) set of primitives (see also Fodor, 1976; Jackendoff, 1990; Pylyshyn, 1984; Hale and Reiss, 2003). One cannot *learn* the post-transduction articulatory correlates of specific phonological features (for one receives no evidence for the nature of such a linkage), nor can one *learn* what mental features correspond to a given output of the perceptual transduction process (for the same reason). We still need to explain, though, how, for example, an acquirer of Danish comes to have an acoustic space for [e] which is different from the acoustic space of a Japanese acquirer's [e]. Certain evidence from acquisition seems relevant to consider at this point.

4.3. Evidence from acquisition

Two acquisition phenomena, in particular, appear to be relevant to this issue of 'language-specific' acoustic spaces of the types outlined in Categories D2 and D3. These are[26]:

[25] Similar criticisms hold of the 'phonetics in phonology' ideas as exemplified in the recent volume edited by Gussenhoven and Kager (2001). For example, a really long [iiiiiiiii] can be recognized the first time it is heard as an /i/, although one of its physical, statistical values (duration) is outside of the range of previously encountered [i]s. Based on durational statistics, [iiiiiiiii] might be taken to be a sentence, since it has the durational statistical properties of a sentence.

[26] The logical contradiction indicated by the two points below is discussed in Hale and Kissock (1997).

- Loss of discrimination capacity for non-native contrasts around 10–12 months (see Rivera-Gaxiola et al., 2005; Best and McRoberts, 2003; Polka and Werker, 1994; Werker and Tees, 1984, inter alia).
- Bilingual second language acquisition (BSLA) of contrasts not present in L_1 (see Mayberry and Lock, 2003; Flege et al., 2003; Flege, 1995, 1991; Yamada, 1995; Mack, 1989, inter alia).[27]

The connections between these acquisition phenomena and the cases in Categories D2 and D3 are that they all appear to show *experience-driven* differences or changes. In the loss of discrimination ability case, *failure* to experience certain input seems to result in a loss of ability to discriminate such input. Crucially, as convincingly demonstrated by Werker and her associates, this loss of ability to discriminate is *not* a loss in 'raw' auditory ability (i.e., not a pure hearing loss of some type). Instead, the loss of ability appears to be a side effect of constructing a linguistic system based on the particulars of the linguistic input in the learner's environment. In the bilingual second language acquisition case, exactly the opposite is true. Upon receiving new and different input (from the L_2), the learner acquires a *new* (or, more accurately, a renewed) ability for discrimination. This results in a single individual producing and perceiving two apparently distinct sets of language-specific contrasts, one according to the L_1 and one according to the L_2. In the D2 and D3 Category cases, something has caused acquirers to 'zero in' on different acoustic spaces for what we currently count as the 'same' featural representations. What is it, then, that all these acquirers have 'learned'? Is it something about the phonology (the feature-to-feature mapping) or something external to the grammar?

4.4. A featural account

If we are correct in arguing (contra Kingston and Diehl) that the transducer cannot learn from experience, then experience-driven differences must have their origin in either the phonology or in non-linguistic perceptual processing. Given our model, it follows logically that, for articulation, for example, differences in acoustic space (our D2 and D3 types) must result from transducers receiving different inputs from the speakers' phonologies (cf. our earlier discussion of the predicted output of featural differences versus differences that are the result of non-linguistic input or modification). Since the input to the articulatory transducer is sets of features, it follows that the explanation for the D2 and D3 types of variation must be featural (i.e., phonological). If we consider the differences from a perceptual standpoint, the difference is, again, simply a difference in features—a speaker of one language has had linguistic input which (potentially) corresponds to different sets of features than a speaker of another language. This is,

[27] We acknowledge that opinions are, in fact, divided on the existence of true bilingualism. Flege (and others who support the position that an L2 can be acquired in a native-like manner) emphasizes that native-like L2 acquisition in both perception and production seems closely related to age of learning and is best achieved below the ages of 5–7. Data cited in Flege (1991) (from experimental results of Flege and Eefting, 1987a) showed that Spanish/English bilingual children performed similarly to monolingual English children in a VOT perception experiment. Additional experiments on the production of L2 sounds in Flege (1991) showed that early L2 learners would attain native-like production, as well as perception. Flege (1995:238) claims that "... without accurate perceptual 'targets' to guide the sensorimotor learning of L2 sounds, production of the L2 sounds will be inaccurate." While we do not necessarily agree that there are targets that guide sensorimotor learning of sounds, it does seem likely that without an ability to perceive a sound, the likelihood of producing anything close to that sound in any sort of consistent way will be negligible. Because of this, the L2 production experiments seem relevant in addition to the L2 perception experiments. For a more current discussion of the opposing view along with experimental results supporting that position – that there is no true bilingualism – see Peperkamp and Dupoux (2002).

of course, just the typical case—inventories of segments (which are only various combinations of features) are rarely, if ever, identical from one language to the next.

It seems clear that the D2(b) case – e.g., a three-vowel system with a relatively large acoustic space for [i] versus a five-vowel system with a relatively narrow one – can be connected directly to the underspecification (D1) cases discussed in detail above. Having a richer underlying vowel inventory requires that lexical items be more fully specified, while having a smaller vowel inventory allows these lexical items to be stored with greater underspecification.[28] We say 'allows' because, as the important Amis (D3) case shows, such underspecification is not *required* in the presence of a restricted inventory. As the synchronic data indicate, the actual evidence regarding the acoustic space available to the Amis acquirer differs from that available to the learner of a three-vowel system with broader acoustic spaces and, given that the acquirer assumes that the acoustic spaces are an invariant by-product of unmodifiable transducers, he or she can only conclude that the vowels in the target language are highly specified. Only *positive evidence* will lead an acquirer to collapse a more specified featural representation into a less specified one (the Subset Principle at work in phonological acquisition, see Hale and Reiss (2003) for discussion). In cases like the Amis one, no such positive evidence presents itself. The acquirer gets *only* tokens that are consistent with the features of [i], for example, not tokens consistent with both [i] and [e] (whose height features could, in a three-vowel system, later be removed once Lexicon Optimization shows them to be irrelevant).

Studies in cross-linguistic differences in VOT production and perception present a picture very similar to the five-vowel versus three-vowel case above, where the three vowels occupy a large acoustic space. Adult categorical perception reflects the distinctions present in the native language, specifically whether there exists a two or three category distinction—voiced/voicless or voiced/voiceless/voiceless aspirated, respectively (Lisker and Abramson, 1970). Infant categorical perception reflects a three-category division (Eimas et al., 1971; Aslin et al., 1981, inter alia). The reduction from three categories to two categories for acquirers who are in a Spanish-language environment, for example, mirrors the transition from an initial hypothesis that there is a featurally-represented height distinction between tokens of [i] and [e] to one where there is no featural distinction between [i] and [e]—the case discussed above. In the case of a reduction in number of VOT distinctions, as in the enlargement of the vowel space, positive evidence (no lexical contrast based on whether a token has long or short lag, for example) allows the acquirer to collapse long and short lag into a single category that just contrasts with voiced by removing the featural distinction between long and short lag representations.[29]

This leaves only the D2(a) cases—non-subset but intersecting (i.e., non-nested) acoustic spaces such as found in the case of Danish [e] and Japanese [e]. Underspecification alone cannot, as far as we can see, account for these cases, for if one of the vowels were underspecified but

[28] For example, the acquirer exposed to a five-vowel [i, e, a, o, u] system will be forced to store at least a [−back] feature to separate [i] and [e] from the other vowels and a [+high] feature to separate [i] from [e] whereas in a broad-space three-vowel [i, a, u] system, the acquirer need only store back features for [i] as all hits in the [−back] space are taken as tokens of [i] regardless of height.

[29] It seems once again that this reorganization does not actually involve 'loss' of discrimination. As Werker and Tees (1984) found for the dental/retroflex and other contrasts, it is still possible for subjects to access the original categories of VOT (Aslin et al., 1981) under certain experimental conditions. Again, this is predicted if the 'reorganization' of categories exists only at the phonological level and not at lower levels of perceptual processing. A full discussion of the experimental results on VOT phenomena and their relationship to a theory of phonological representations is definitely required to complete this picture but is too lengthy and complex to be included here. Such a discussion would also address the question of boundary shifts that are sometimes said to accompany such VOT reorganization.

otherwise featurally non-distinct from the other, we would find, as we have seen, a superset-subset acoustic space relationship. Since the explanation for all other cases of variation of this type has been featural, and given the learning theoretic complexities of positing a 'learning' transducer, we must assume that the difference between Danish [e] and Japanese [e] is featural as well.

We would like to note that at least some of these cases are almost certainly nothing more than artifacts of our use of the IPA or of our use of the existing feature system, or both. A cursory look at a comparison of vowels (plots of first two formants) in Japanese and Danish in Ladefoged (2001) shows Japanese [e] and Danish [ɛ] occupying the same acoustic space. In general, tokens that occupy the same acoustic space should have the same IPA label.[30] It is, however, common practice to label vowels and sometimes consonants based on the language-specific inventory of sounds such that identification is relative rather than absolute (i.e., acoustic-based). In addition, the featural properties of some sounds are often underdetermined because they, too, are being used relative to the inventory of some particular language. Features for [r] are notorious in this regard, often, for any single language, not distinguishing trills from taps, taps from approximants, velarization, and so on. The commonly held notion, inherited from Structuralism, has been that a necessary and sufficient featural description needs only to capture the contrasts *within* an inventory. This is a seriously flawed approach if one of the tasks is to determine the properties of a feature inventory given by UG, with which every acquirer is endowed, and which is supported by the experimental evidence on infant perception of sounds not present in the ambient target language.[31]

Even after discarding spurious cases of the type described above, we expect that there will be cases that require a thorough examination and explanation. For these remaining cases, experimental sources of support for our hypothesis are going to be found in only a very limited domain, as far as we can see—in the perception of bilinguals whose L1 and L2 differ along the relevant vowel parameters. Using bilingual subjects will control for a number of speaker-specific, extra-grammatical factors which can make token comparison difficult. Obtaining production data (clusters of tokens in a target area) and the results of perceptual identification tasks which focus on both the intersecting and non-intersecting areas (for vowels such as Danish [e] and Japanese [e]) should provide information about the correctness of our hypothesis.

We summarize this section with an explicit description of how differences in acoustic space as well as acquisition of those differences can be accounted for in our model. UG provides an innate set of features to be used in phonological computation and (some subset of) those features serve as the input and output to an articulatory and perceptual transducer, respectively. Transduction of features into gestural score and of acoustic score into features is deterministic and invariant. 'Learning' is the process of building phonological representations *based on the featural input from the perceptual transducer* and determining which computational processes will operate in any particular instantiation of a phonology.[32]

We assume, furthermore, that there is an 'endpoint' to the learning process such that modification of the phonology based on input will cease after a certain time. (The lexicon, on the other hand, has no 'cap' on learning—new lexical items may be acquired at any time.) Attaining

[30] This is not to assert that these two *particular* sounds are undifferentiated acoustically. We use it here only to make the point about use of IPA labels.

[31] Note that the perception and categorization of vowels seems in general to be less straightforward than that of consonants.

[32] We take no position here on whether this involves reranking constraints in a constraint-based system or determining the specific application of rules, since this aspect of the phonology is not relevant to the immediate discussion.

adult-like phonological representations, itself, is a two-part process. The first part of the process involves storing initial phonological representations (for lexical items acquired at the earliest stages of acquisition) with fully specified featural representations (for an extensive discussion of why fully-specified representations are necessary at the initial stage, see Hale and Reiss, 1998, 2003).[33] The second part is lexicon optimization, in which the stored representations in the lexicon are subjected to a review for predictable versus idiosyncratic information. Predictable information is then stripped from the representations and is, instead, added dynamically by the computational system. The end result is a set of URs (featural representations for lexical items) whose characteristics are determined by (1) limitations of UG (what constitutes possible features); (2) the (featural) input that the acquirer received and stored; and (3) the process of Lexicon Optimization. We expect, then, that (2), featural input to the acquirer, will be key in determining how similar one individual's phonology will be to another individual's phonology (properties of UG being shared by all individuals and results of Lexicon Optimization falling out from stored input). All of the experience-driven cases we have discussed above are predicted under this model. We go through these in detail below, following an order which groups the most similar cases together.

The cases of underspecification in D1 (Russian, Marshallese) and the D2(b) case (broader acoustic space for an [i] in a three-vowel system than for an [i] in a five-vowel system) are the direct result of a *lack* of some feature or features in the input to the phonology from the transducer. As the transducer is deterministic, this means that the input to the transducer itself (i.e., the acoustic score) did not contain information that could be converted into a feature or features. From the point of view of an acquirer, there is simply no information which would allow them to set up a final featural representation with [back] for Russian [x] or [back] and [round] for any Marshallese vowel. We say 'final' featural representation here because of the interesting cases of surface similarity between, for example, Marshallese [ɛ] and English [ɛ] mentioned in the discussion of Marshallese in section 4.1. By our own analysis, the Marshallese acquirer will set up an *initial* representation for [ɛ] which is fully specified and includes all the same features that an English acquirer will have initially for [ɛ] (assuming the same acoustic score). Crucially, however, the process of Lexicon Optimization will cause the acquirer to strip out [back] and [round] from [ɛ] in Marshallese, given the lack of evidence for [back] and [round] in other vowels and the fact that the presence of those two features can be completely attributed to the surrounding consonants. These [ɛ]-type cases, then, are different from the [i͡ɯ] type cases only in the route by which they become underspecified, not in their final representations as underspecified. The [i͡ɯ] type vowels, in contrast, simply have nothing in the acoustic score which could be converted by the transducer into the features [back] and [round].[34] From this somewhat more complicated picture, it is easy to see the origin of the differences between the [i] vowel spaces in the three versus five-vowel systems case. In the three-vowel system, an acquirer will get input for [i] which will overlap significantly with the [e] space, for example, and cause him/her to set up initial representations which distinguish such tokens. However, Lexicon Optimization will cause this acquirer to collapse initial representations of [e] with [i] by removing specification for the distinguishing feature [+/−high]. This single, underspecified

[33] Note that this means that a given UR will be featurally identical to its SR at the earliest stages of phonological learning.

[34] As far as we are aware, the Marshallese 'tied' vowels, such as [i͡ɯ], are not part of the UG-feature generated vowel possibilities. They are not found contrastively either within or across languages. Their peculiar physical (acoustic) features appear to be completely dependent upon context.

featural representation will naturally produce hits in a broader acoustic space because the transducer is receiving less specific information. The five-vowel system acquirer, on the other hand, will not be able to collapse his/her initial representations of [i] with [e] through Lexicon Optimization (their featural difference being critical in differentiating lexical representations that must be kept distinct). For a discussion of the learnability problems posed by such underspecified systems in an optimality-theoretic framework, see Hale and Kissock (in press).

The Amis case of a 'narrow' acoustic space in a three-vowel system can now be contrasted immediately with the 'broad'-space three-vowel system just discussed. Crucially, the Amis acquirer *doesn't get the same input*. To use the same example, all of the input for a non-low front vowel that the Amis acquirer receives falls into the space denoted by [+high] rather than the space denoted by *both* [+high] and [−high, −low]. Every initial featural representation is exactly for [i] and no positive evidence (input to the transducer) exists for the Amis acquirer that he/she should set up any representation for [e]. Lexicon Optimization does not play a role in the relevant final featural specification in this case, it's simply that the acquirer gets a different type of evidence for representations.

The bilingual second language acquisition case is characterized by an acquirer developing *new* contrasts—contrasts not present in the L1. Note that new contrastive sounds are not the only new aspects of a linguistic system that the acquirer develops. The bilingual acquirer exhibits new developments in every aspect of the grammar—syntax, morphology, *and* phonology (different constraint ranking or rules). In fact, there seems every reason to believe that the acquirer has a completely new, second grammar—that of the L2. By definition, bilingual acquisition of an L2 means that the acquirer is in no way different from another individual for whom the L2 was the 'L1.' Our usual characterization of knowledge of language is that the individual has a grammar (with all the relevant features and operations), so there is no reason to suppose this characterization should be valid in some cases (only of L1) and not in others (bilingual L2). From the point of view of acquiring new contrasts, then, this is just another case of normal acquisition where featural input and lexicon optimization determine the final featural representations. The loss at 10–12 months of discrimation ability is a more complex case but one which still reduces to a difference in featural representations in our model. As mentioned earlier, experiments suggest that the loss is particular to the linguistic system or, at the very least, is not what could be termed a 'hearing loss.' This supports the argument that it is a feature-based loss.[35]

Finally, we consider the Danish [e] versus Japanese [e] case, where the acoustic spaces are not in a superset/subset relationship and cannot be accounted for by underspecification. Based on our model, only a difference in featural representation will account for the difference in acoustic space between the Danish and Japanese vowels. Currently available models of the universal feature inventory are insufficiently rich to handle such a distinction, which is typically attributed to language-specific implementation rules. We think it highly likely that we have not yet discovered the final and correct characterization of UG-provided features and that only detailed consideration of the full complexities of phonological acquisition and cross-linguistic phonological variation will allow for the discovery of the full feature set.[36] We are presently engaged in this research and hope to report on relevant results in the near future.

[35] Numerous empirical and experimental design issues arise in the literature regarding this topic. For a critical review of Maye and Gerkin (2000) and Werker et al. (2002) and other widely cited literature on the topic, see Kissock (2002).

[36] Interestingly, Steriade (2000) comes to a similar conclusion even though the questions she is asking and the approach she takes are fundamentally different.

5. Conclusion

In conclusion, we have argued that:

- There are several distinct sources for microvariation – in our definition, non-phonological differences introduced in the transduction process, or by individual physical properties or external physical properties – in the speech output of humans.
- The microvariation that we find is a function of non-grammatical processes (the ambient environment, the anatomical traits of the individual speaking, the current physical state of those anatomical traits (e.g. moist versus dry), as well as the precise nature of the transduction process for a particular feature in a particular context). [Categories A, B, and C above.]
- Variation – by our definition, differences due to differences in featural representations – may be mistaken for microvariation and vice versa.
- In a seeming paradox, cross-linguistic differences may involve virtual identity of output (the vowel of Marshallese /$t^{j}V_{LO}t^{j}$/ is representationally distinct from, but physically virtually the same as, English [ɛ]), while still constituting cases of variation. Here, the variation is at the representational (and thus grammatically-relevant) level.
- One of the challenges before us is to determine precisely when we are dealing with microvariation and when we are dealing with variation (including cases of apparent non-variation). We believe such a research agenda will help us develop a better understanding of the nature of the grammar and of how the grammar interacts with other systems.
- Many aspects of apparently systematic, apparently phonetic implementation (phenomena usually referred to as 'language-specific phonetics') may find their most coherent explanation by a slight enriching of the phonological feature set.

The last point deserves some further discussion. Why, if there exist additional UG features, have these features not been noted to function *contrastively* within a single linguistic system?

We believe the explanation for this may be found in the following considerations. First, the set of attested languages is a subset of the set of attestable languages (where 'attestable' includes all linguistic systems which could develop diachronically from existing conditions—e.g., all dialects of English or Chinese or any other language in 400 years, or 4000 years, etc.). In addition, the set of attestable languages is a subset (those which can evolve from current conditions) of the set of humanly computable languages. (In our opinion, the human phonological computation system can *compute* a featural change operation such as /p/ → [a]/ _ d but it is of vanishingly small probability that such a rule could arise from any plausible chain of diachronic changes.) Finally, the set of humanly computable languages is itself a subset of formally statable systems (which could include what we take to be humanly impossible linguistic processes such as /V/ → [V:] in prime numbered syllables). The key point here is that the set of *diachronically* impossible human languages is not equivalent to the set of *computationally* impossible human languages.[37]

It is not difficult to imagine that featural distinctions which are transduced to very similar outputs (as in the case of Danish [e] and Japanese [e] discussed above) may present a major challenge to an acquirer. It may in fact be the case that systems which contain *both* Danish [e] and Japanese [e] would – like lengthy chains of center-embedded structures in syntax – be unattested for processing reasons (strictly speaking, because of the dependence of acquisition on accurate processing). However, *UG is a theory of the human linguistic computational capacity, not a*

[37] For a complete discussion of these distinctions, see Hale and Reiss (2000a).

theory of the (accidental) set of attested human languages. The set of features posited for UG should be precisely the set required to account for the properties of the human genetic endowment for phonological computation. We believe that this set of features can be discovered only through the assumption of a coherent learning theory and the development of a corresponding detailed account of cross-linguistic variation in acoustic spaces—excluding true microvariation, as outlined here.

Acknowledgements

The authors would like to thank the audiences of the 6e Journées Internationales du Réseau Français de Phonologie (June 2004) and the 25th GLOW Colloquium in Amsterdam (May 2002) as well as an anonymous reviewer for their comments and suggestions.

References

Anderson, S., 1985. Phonology in the Twentieth Century: Theories of Rules and Theories of Representations. University of Chicago Press, Chicago.

Appelbaum, I., 1996. The lack of invariance problem and the goal of speech perception. ICSLP-1996 1541–1544.

Aslin, R.N., Pisoni, D.B., Hennesy, B.L., Perey, A.J., 1981. Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. Child Development 52, 1135–1145.

Bender, B., 1968. Marshallese phonology. Oceanic Linguistics 7, 16–35.

Best, C., McRoberts, G., 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. Language and Speech 46, 183–216.

Browman, C., Goldstein, L., 1990. Gestural specification using dynamically-defined articulatory structures. Journal of Phonetics 18, 299–320.

Choi, J., 1992. Phonetic Underspecification and Target Interpolation: An Acoustic Study of Marshallese Vowel Allophony. UCLA Working Papers in Phonetics 82, UCLA.

Chomsky, N., Halle, M., 1968. The Sound Patterns of English. The MIT Press, Cambridge, MA.

Eimas, P.D., Siqueland, E.R., Jusczyk, P.W., Vigorito, J., 1971. Speech perception in infants. Science 171, 303–306.

Flege, J.E., 1991. Perception and production: the relevance of phonetic input to L2 phonological learning. In: Huebner, T., Ferguson, C.A. (Eds.), Crosscurrents in Second Language Acquisition and Linguistic Theories. John Benjamins, Amsterdam, pp. 249–289.

Flege, J.E., 1995. Second language speech learning: theory, findings, and problems. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross Language Speech Research. York Press, Timonium, MD, pp. 233–277.

Flege, J.E., Schirru, D., MacKay, I., 2003. Interaction between the native and second language phonetic subsystems. Speech Communication 40, 467–491.

Fodor, J., 1976. The Language of Thought. Harvester Press, Sussex.

Gussenhoven, C., Kager, R., 2001. Phonetics in phonology. Phonology 18 (1).

Hale, M., 2000. Marshallese phonology, the phonetics-phonology interface and historical linguistics. The Linguistic Review 17, 241–257.

Hale, M., Kissock, M., 1997. Nonphonological triggers for renewed access to 'phonetic' perception. In: Sorace, A., Heycock, C., Shillcock, R. (Eds.), Proceedings of the GALA'97 Conference on Language Acquisition. University of Edinburgh, Edinburgh, pp. 229–234.

Hale, M., Kissock, M., in press. The phonetics-phonology interface and the acquisition of perseverant underspecification. In: Ramchand, G., Reiss, C. (Eds.), Handbook on Interface Research in Linguistics. Oxford University Press, Oxford.

Hale, M., Reiss, C., 1998. Formal and empirical arguments concerning phonological acquisition. Linguistic Inquiry 29, 656–683.

Hale, M., Reiss, C., 2000a. Phonology as cognition. In: Burton-Roberts, N., Carr, P., Docherty, G. (Eds.), Phonological Knowledge: Conceptual and Empirical issues. Oxford University Press, Oxford, pp. 161–184.

Hale, M., Reiss, C., 2000b. Substance abuse and dysfunctionalism: current trends in phonology. Linguistic Inquiry 31, 157–169.

Hale, M., Reiss, C., 2003. The subset principle in phonology: why the *tabula* can't be *rasa*. Journal of Linguistics 39, 219–244.

Hammarberg, R., 1976. The metaphysics of coarticulation. Journal of Phonetics 4, 353–363.

Jackendoff, R., 1990. Semantic Structures. MIT Press, Cambridge, MA.

Jakobson, R., Fant, G., Halle, M., 1963. Preliminaries to Speech Analysis. The MIT Press, Cambridge, MA.

Jusczyk, P., Luce, P., 2002. Speech perception and spoken word recognition: past and present. Ear and Hearing 23, 2–40.

Keating, P.A., 1988. Underspecification in phonetics. Phonology 5, 275–292.

Kingston, J., Diehl, R., 1994. Phonetic knowledge. Language 70, 419–454.

Kissock, M., 2002. The initial state: recent empirical evidence. Paper presented at the 10th Manchester Phonology Conference, Manchester, England.

Kuhl, P.K., 1991. Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. Perception and Psychophysics 50, 93–107.

Ladefoged, P., 2001. A Course in Phonetics, fourth ed. Harcourt, Inc., Orlando.

Liljencrants, J., Lindblom, B., 1972. Numerical simulations of vowel quality systems: the role of perceptual contrast. Language 48, 839–862.

Lisker, L., Abramson, A., 1970. The voicing dimension: some experiments in comparatic phonetics. In: Proceedings of the 6th International Congress of Phonetic Sciences, Prague, 1967, Academia, Prague, pp. 563–567.

Mack, J., 1989. Consonant and vowel perception and production: early English–French bilinguals and English monolinguals. Perception and Psychophysics 46, 187–200.

Maddieson, I., Wright, R., 1995. The vowels and consonants of Amis. UCLA Working Papers in Phonetics 91, 45–65.

Mayberry, R.I., Lock, E., 2003. Age constraints on first versus second language acquisition: evidence for linguistic plasticity and epigenesis. Brain and Language 87, 369–384.

Maye, J., Gerkin, L., 2000. Learning phoneme categories without minimal pairs. Proceedings of the 24th Annual Boston University Conference on Language Development, pp. 522–533.

Peperkamp, S., Dupoux, E., 2002. Deconstructing phonology: the case of loanword adaptations. Paper presented at the 2nd North American Phonology Conference, Concordia University, Montreal.

Pierrehumbert, J., 2001. Words and wordoids. Paper presented at the Workshop on Early Phonological Acquisition Carry-le-Rouet, France.

Polka, L., Werker, J.F., 1994. Developmental changes in perception of non-native vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance 20, 421–435.

Pulleyblank, D., 2001. Defining features in terms of complex systems: is UG too complex? Paper presented at the Workshop on Early Phonological Acquisition, Carry-le-Rouet, France.

Pylyshyn, Z., 1984. Computation and Cognition: Toward a Foundation for Cognitive Science. MIT Press, Cambridge, MA.

Rivera-Gaxiola, M., Silva-Pereyra, J., Kuhl, P.K., 2005. Brain potentials to native- and non-native speech contrasts in seven and eleven-month-old American infants. Developmental Science 8, 162–172.

Sapir, E., 1933. The psychological reality of phonemes. Journal de Psychologie Normale et Pathologique 30, 247–265.

Scobbie, J., 2001. Sounds and structure: covert contrast and non-phonemic aspects of the phonological inventory. Paper presented at the Workshop on Early Phonological Acquisition, Carry-le-Rouet, France.

Steriade, D., 2000. Paradigm uniformity and the phonetics/phonology boundary. In: Pierrehumbert, J., Broe, M. (Eds.), Papers in Laboratory Phonology VI. Cambridge University Press, Cambridge.

Stevens, K.N., 1972. The quantal nature of speech: evidence from articulatory-acoustic data. In: Denes, P., David, Jr., E. (Eds.), Human Communication: A Unified View. McGraw Hill, New York.

Stevens, K.N., 1998. Acoustic Phonetics. MIT Press, Cambridge, MA.

Vaux, B., 2002. Explaining vowel systems: dispersion theory vs. evolution. Paper presented at the 2nd North American Phonology Conference, Concordia University, Montreal.

Werker, J., Tees, R., 1984. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 7, 49–63.

Werker, J., Fennell, C., Corcoran, K., Stager, C., 2002. Infants' ability to learn phonetically similar words: effects of age and vocabulary size. Infancy 3, 1–30.

Yamada, R., 1995. Age and acquisition of second language speech sounds: perception of /r/ and /l/ by native speakers of Japanese. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross Language Speech Research. York Press, Timonium, MD, pp. 305–320.