Against Formal Phonology

Robert F. Port and Adam P. Leary
Department of Linguistics, Indiana University
Bloomington, Indiana, 47405
First Author: 812-855-9217
port@indiana.edu
adamlear@indiana.edu

June 2, 2005

Chomsky and Halle (1968) and many formal linguists rely on the notion of a universally available phonetic space defined in discrete time. This assumption plays a central role in phonological theory. Discreteness at the phonetic level guarantees the discreteness of all other levels of language. But decades of phonetics research demonstrate that there exists no universal inventory of phonetic objects. We discuss three kinds of evidence: First, phonologies differ incommensurably. Second, some phonetic characteristics of languages depend on intrinsically temporal patterns, and, third, some linguistic sound categories within a language are different from each other despite a high degree of overlap that precludes distinctness. Linguistics has mistakenly presumed that speech can always be spelled with letter-like tokens. A variety of implications of these conclusions for research in phonology are discussed.*

[A slightly revised version of the manuscript will appear in *Language*, probably in the December, 2005 issue.]

* Acknowledgments. The authors are grateful to Keith Johnson, Dan Dinnsen, Ken de Jong, David Pisoni, Terry Nearey for their discussion of these issues.

The generative paradigm of language description (Chomsky 1965, Chomsky 1964, Chomsky et al. 1968) has dominated linguistic thinking in the United States for many years. Its specific claims about the phonetic basis of linguistic analysis still provide the cornerstone of most linguistic research. Many criticisms were raised against the phonetic claims of the *Sound Pattern of English* (Chomsky et al. 1968), some from early on (e.g. Lisker et al 1971, Hockett 1968, Sampson 1977, Keating 1984) and others in recent years (Pierrehumbert 2000a, b, Steriade 2000), addressing various aspects of the phonetics model. The goal of this paper is to develop one very general criticism: that a fundamental mistake of the generative paradigm is its assumption that phonetic segments are formal symbol tokens. This assumption permitted the general assumption that language is a discrete formal system. This bias forced generative phonology to postulate a phonetic space that is closed and contains only static symbolic objects. We show that theories of phonetics satisfying these constraints have little to no support from phonetic evidence and thus that the formal-language assumption is surely incorrect.

1. Classical Phonetics and the IPA

First, we should ask whether the majority of phoneticians find agreement with generative linguists on basic issues like the closure of the phonetic space. For example, do phoneticians generally agree with phonologists that we will eventually arrive at a fixed inventory of possible human speech sounds? The answer is no. Although there may be many differing views, we treat the latest edition of the *Handbook of the International Phonetic Association* (1999) and the views expressed in Ladefoged & Maddieson's *Sounds of the World's Languages* (1996) as broadly representative. The IPA has, for about a hundred years, promoted and periodically updated an alphabet of graphic symbols for technical phonetic purposes combined with suggestions for how to use these symbols to transcribe linguistic utterances for both pedagogical purposes and for academic communication. Although earlier editions offered very little explanation about what was being assumed in general about the phonetic space, the recently revised *Handbook of the IPA* (1999) presents its own theoretical rationale. The IPA makes no claim about the limits of the phonetic space nor does it posit any fixed number of possible phonetic distinctions. Basically, the IPA simply offers an alphabet of graphic symbols each of which has an approximate articulatory or auditory interpretation but whose precise meaning requires elaboration by the user of the alphabet.

The IPA is clear that it assumes that languages each have a small number of segmental phonemic contrasts and that only some properties of speech sound are phonetically relevant, in particular, properties that are invariant across individual speakers, speaking rates, styles of voice quality, etc. (IPA: 3, 27). An alphabet of symbols of modest size is offered (about 100 Cs and Vs plus about 25 suprasegmentals in the latest edition) along with an expandable list of diacritic marks. Each of the symbols can be used for a range of different but similar sounds. It is suggested, in particular, that they can be employed in 3 basic ways for linguistic description (leaving aside various special-purpose uses like speech pathology records or comparison of the phonetic detail of one language with another). First, the graphic symbols may be used for **phonemic transcription** (or broad transcription) where the symbols exhibit the least precision since only the contrastive sounds in a single language are distinguished. It is assumed that the reader speaks the language and knows how to read the text with appropriate allophonic variation. The second way is **allophonic transcription** or systematic phonetic transcription where many additional context-sensitive details are made explicit in the transcription – especially variants due to specific contextual features or

ones that are likely to be noticed by speakers of other languages. Both of these methods assume that a phonological analysis is available for the language that provides guidance on how the alphabet is applied in this language.

There is a third way to use the phonetic alphabet – the method that many nonphoneticians may think the IPA was designed for. **Impressionistic transcription** applies the phonetic alphabet to speech in an unknown language to notate all properties that are potentially relevant to the transcriber's purposes. Ideally, the impressionistic transcription would be objective and exhaustive in some sense. But, of course, many decisions regarding a transcription depend on knowing the language (34). For example, should one transcribe [lɛ] vs. [lʲɛ], [oa] vs. [owa], [tʃ] vs. [č], or [bər] vs. [bɚ], etc.? The answers to questions like these depend not simply on the actual gestures or sounds themselves, but also depend on many other facts about the phonology of the language. The graphic symbols can be interpreted in some absolute or objective sense only in a vague way. Specific choices are made depending on many aspects of the phonology of a language aside from the actual speech gestures of the speaker. Furthermore, some sound distinctions depend critically on phonetic perception skills acquired very early in life (Logan et al. 1991, Strange 1995a). There is little reason to believe that any linguist or phonetician can do objectively accurate transcription in an arbitrary unknown language. Thus, as far as the IPA is concerned, impressionistic transcription is only possible to a limited extent. So, despite the assumptions of some phonologists, nothing in the principles of the IPA implies the possibility of any `authoritative' or `canonical' impressionistic transcription of an arbitrary utterance.

For our purposes here, the important point is that the IPA (1999) is quite explicit that (1) a phonetic transcription can never be done without considerable phonological information, and also (2) nothing in the *Handbook* suggests there is a `complete' set of speech sounds. The alphabet and the set of diacritics are always subject to expansion depending on new research or changes in the languages of the world. These assumptions contrast strongly with the approach of Chomsky & Halle and with the approach of phonologists working in the generative tradition. The IPA assumes an open and potentially unlimited set of possible speech sounds (leaving aside, of course, the many nonlinguistic variations like those due to speaking rate, speaker idiosyncrasies and vocal tract size), whereas Chomsky & Halle assume a closed set of phonetic options employed by phonetician-linguists whose transcriptions can be taken as the raw data for phonological analysis. One goal of this paper is to explore the issue of why such a bold assumption is made and the degree to which it might be valid.

In *The Sounds of the World's Languages*, Ladefoged et al. (1996) also assume that the space of possible speech sounds is indefinitely large. They restrict their attention to the sound differences that are used for minimal distinctions between lexical entries in some language or which may not be found in any extant language but which are noticeable differences between languages. After all, `the next generation of speakers…may even create sounds that have never been used in a human language before' (369). They acknowledge that it is very risky to say that there is some specific number of places of articulation or states of the glottis, etc. They nevertheless attempt to develop a set of mainly articulatory (but also acoustic) parameters which they hope will be sufficient to differentiate all the possible contrasts in consonants and vowels (cf. Ladefoged 1980). Thus they do not claim it is possible to describe a closed set of `phonetic capabilities' of the human species, but hope that their continuous acoustic and articulatory parameters will be sufficient to differentiate all of them that appear (2-6). Ladefoged & Maddieson (1996, Ladefoged, 1980) explicitly claim there is no closure to the set of possible speech sounds thus finding themselves in sharp contrast with the phonetics of *Sound Pattern of English* and most modern American phonologists.

Phoneticians in general, and at least some phonologists (Bybee 2001, Pierrehumbert 2000b), seem to deny the notion of an apriori, closed inventory of sound types in human language. Does this mean that categories of speech sound are viewed as a mistake altogether? Are phoneticians simply denying the existence of any sound categories? No. Ladefoged and Maddieson and the IPA base their approach on the notion of distinct sound contrasts within languages. Certainly, the present authors take for granted that individual languages exhibit sound categories that yield contrasts between lexical entries. The many examples of minimal pairs show that languages do employ distinctive categories of speech sounds. The difference between the generative school and the phoneticians has to do with the nature of the categories and how they are acquired and represented. There is a critical difference between the notion of categories of sound and actual symbol tokens. The generative school insists that the sound categories are cognitive symbol tokens. But to call linguistic sound categories symbol tokens is a very strong assumption and one that need not be accepted in order to account for distinctive sound contrasts in languages. Later in this paper, we shall try to suggest how an appropriate set of sound categories can be observed and acquired by children without assuming all the formal power of actual symbolic units. First, however, we need to clarify what is meant by the assumption that language is a formal system.

2. Language as Formal Knowledge.

Classical generative phonology is built upon two basic assumptions: that language is a kind of knowledge and that linguistic knowledge is formal. The achievement of Chomsky and Halle was to follow out the consequences of the idea that human linguistic knowledge is fully expressed using some formal algebra of symbol tokens. The goal of linguistic research given these assumptions should be to discover the formal algebra, i.e., the grammar, available to an adult speaker for employment by the linguistic performance system (Chomsky, 1965). The symbolic knowledge assumption is taken to permit exploitation of all the capabilities of discrete mathematics to model linguistic knowledge. The powers of discrete mathematics include (a) allowing an unlimited number of computational steps, (b) use of symbol tokens that are perfectly stable over time, and (c) symbol tokens that can be `read' and `written' without error. Finally, (d) discrete mathematics offers a metaphorical time scale that is also discrete, so all the messiness of continuous time is ruled out of linguistics. Discrete time plays two roles. First, it governs the operations of the system implementing the grammar, such as the jump from the left side of a rule to the right-hand side. Rules take place in a single step. And secondly, the serial order of discrete symbols models time as integer-valued, as $t = 0, 1, 2, 3, \ldots$, rather than with rational and real numbers. Thus $t = 2.4$ is meaningless and undefined. Generative theorists and practitioners do not hesitate to exploit all these properties in their reasoning about phonology.

As with other mathematical systems, description of the system requires spelling out an apriori set of symbol types from which the complex representations in speakers' minds can be constructed. Thus, for basic arithmetic, one must postulate at least the integer 1 and for theorems in propositional logic simply postulates propositions like p and q. For language, Halle and Chomsky proposed an inventory of apriori symbol types, including NP, [Vocalic], [Voiced], [High], and so on. Any specific list is, of course, assumed to be tentative, subject to additional research. The goal of this essay is to argue that, despite the fact that the mathematics of formal string grammars was inspired by alphabetic written language, human language capability cannot actually be a formal grammar. The argument developed here focuses especially on the problem of the symbol tokens, that is, on the universal phonetic space and the issue of discrete vs. continuous time as employed in generative phonology.

There are several awkward consequences of the assumption that all of language is symbolic knowledge. The first awkwardness is that formal symbols can only be static. Just as my knowledge of, say, who the President of the U.S. is seems to be a description of the state of my memory system at some point in time, similarly, the Chomsky-Halle view is that linguistic knowledge can be fully described as a static structure. This provides a reasonable model for phonetic segments where the articulators are relatively static, as in fricatives and steady-state vowels and sonorants. However, for the speech segments that involve movements, such as stops, glides, affricates, diphthongs, etc., and for all properties that depend on time, the segmental description is clumsy and unrevealing.

This awkwardness results because speakers only produce and perceive speech as an event in time (leaving aside the technologies of reading and writing). So, if the time scale of the language itself is discrete only, then apparently during speech perception, the temporal information must be stripped away (or recoded into serial symbols) to produce the abstract, static linguistic description for linguistic cognition. Real linguistic processing takes place in discrete time. Then at the time of speech production, much of that same information must be reapplied in the conversion from digital to analog mode. So, any thinking that involves language is said to happen in discrete time but whenever speech is used, either in talking or in listening to speech, users engage in an activity in continuous time. As is demonstrated below, many details of the temporal patterning of speech turn out to be critical to the proper specification of linguistic structure for a speech perceiver. So the awkwardness – or we might even say the paradox – is that human cognition is thus asserted to have two parallel kinds of time – one continuous and one discrete with temporal aspects alternately stripped away and then reinserted.

The second reason it is awkward to assume that the entire grammar is static knowledge is that the hypothesis makes the prediction that all nonstatic features of language must be universal. If every way that languages may differ is expressible in discrete symbols, then there is no possibility of language-specific temporal structures – no characteristic rhythmic patterns or distinctive timing properties. But this is obviously not the case: languages differ greatly in characteristic timing patterns and one of the most prominent features of foreign accent is inappropriate timing (Port et al. 1983, Flege et al.1992, Eisen et al. 1992, Tajima et al. 1997, Tajima et al. 2003). Given this situation, the theory can be retained as long as some static (universal) features can be claimed to cause the temporal effects as epiphenomena due to `temporal implementation rules' (Chomsky et al. 1968, Klatt 1976, Port1981b, Keating 1984). So the distinction between timeless-and-static (competence) vs. temporal-and-active (performance) does not apparently line up with properties-of-the-language vs. properties-of-the-speaker. For many years, linguistics has followed Chomsky's insights but doing so has forced linguistic thinking to rule out of the field many phenomena that appear, prima facie to be relevant for the description of language – such as all the temporal phenomena that can't be expressed in terms of serial order. The goal of the next section is to outline in more detail why the assumptions of modern phonological theory create a serious problem for dealing with time, and then, in the following section, to review some of the language-specific phenomena of phonetics that create difficulties for the static theory of language.

2.1 Time and Formal Knowledge. No one denies that speech is produced in time, that is, that the sentences, words, consonants and vowels of human language are always extended in time when they are uttered. Still, on the generative view, since language is viewed as a body of symbolic knowledge, then, because knowledge is static and symbols are serially ordered, it must be concluded that temporal extension is not an intrinsic property of language and that the temporal

patterns of language (other than those representable in serially ordered symbols)[1] are not relevant and will not be revealing about language itself. Generative linguistics thus presumes that the detailed temporal layout of speech arises from outside language at the stage where the static symbol structures of the language are performed by the human body. It is only the segmental, discrete symbolic transcription, a matrix of feature values, that represent `the phonetic capabilities of man' (Chomsky et al., 1968: 295) not any events distributed in time. As Halle has noted (1985:106), lines of printed text are good models of the hypothesized cognitive representations. The cognitive form of language has serially ordered, discrete words composed from a small inventory of meaningless sound- or gesture-related segments, strikingly similar to the letters on a printed page.

Cognitive symbol strings may be `implemented' in time by the linguistic `performance' system if a linguistic construction happens to be spoken. Speech is language as filtered or distorted by the performance system that maps language into speech. From the generative point of view, then, speech performance is derivative and is merely one possible output mode -- one of several ways (along with writing) to get language from the mind out into the body and the world. Speech could be said to impose time on an intrinsically nontemporal structure.

Despite many criticisms of this theory of phonetics from within phonology (e.g., Pierrehumbert 2000a, b, 2001, Steriade 2000, Boersma 1998, Bybee 2001), this view is still widespread in early-21^{sr} century generative research on language. It seems that the issue on which to criticize the Chomsky-Halle symbolic model of language most effectively is to demonstrate how it is forced to impose discrete time despite evidence of many kinds of temporal patterns in languages. The problem stems directly from that most fundamental Chomskyan distinction: Competence vs. Performance. On one hand, there is a formal world, the Competence world, where the serial order of timeless symbols provides a time-like framework for the data structures of natural language (Chomsky 1965). Formal operations apply in discrete time to these data structures just as they apply in a derivation in formal logic or a mathematical proof. And just as in the formal structure of some mathematical systems and computer programs, linguistic structures (like words, phrases, sentences) are composed from simpler formal parts in a hierarchy of levels. These units are assumed to be distinct in just the way that bits in a computer are distinct. Formal operations like rewrite rules or constraint evaluations take place between successive ticks of a serial clock. Given the formal nature of language, any time that might be required for the operations to take place is understood to be merely epiphenomenal and not directly relevant to the operations or formal structure and thus not linguistic in any way. In discrete time the discontinuous and instantaneous clock ticks are the only temporal locations that exist. In a computer program which is similarly discrete, for example, changing the clock rate on the computer has no influence on the execution of the program itself except that it runs a little faster or slower when compared to an external clock.

On the other hand, there is also a physical world of brains and bodies living in continuous time – the world of Performance. In the formalist framework, the structures of language are assumed to be `implemented' in time (see Scheutz 1999 for careful discussion of the notion of implementation). Somehow the brain is supposed to support these two parallel, time-locked scales for time, one discrete time and one continuous. This contrast between the formal and the

[1] For example, Hayes' (1995) discussion of meter in various languages defines meter in serial-order terms, not in terms of intervals even though hypotheses like stress-timing are about the equality of time intervals. M. Liberman (1975), however, clearly appreciated the distinction between a `metrical grid' that was continuous-time vs. one that is defined in discrete-time (Chomsky 1964).

physiological is related, at least historically, to the distinction between the mind and the brain or between the soul and the body. As far as linguistics is concerned, the implementation processes may be of scientific interest but they are not linguistically relevant. Linguists can easily afford to ignore, indeed ought to ignore, all such matters of Performance. The reason is the same as the reason why programmers pay no attention to the clock rate of the computer that will run their programs.

This entire point of view is deeply misguided. There are many reasons to reject the computational model of cognition as a whole (see van Gelder and Port 1995, Thelen et al. 1994, Clark 1997), but we discuss just a few of the reasons here. The main problem is that divorcing Competence from Performance creates a gulf, an incommensurability, between serial, discrete time (chopped into static moments) and the continuous time of biological systems. For speech production, converting from discrete to continuous requires some form of low-pass filter to slow down the abrupt changes between segments plus a means to implement timing rules. And for perception, conversion from the roughly continuous analog acoustic signal to a discrete-time representation requires some way to measure durational patterns not just sequences of segmental events. In a computer, this perceptual function could be achieved by using a sampling clock at some constant rate to take snapshots of an external continuous signal at a rate fast enough to capture all the temporal information of interest (which lies below the `Nyquist frequency'). At each click of the analog-to-digital clock, the value of the analog signal is measured and stored. Then the digital computer could process the now serially-ordered representation – using numbered memory locations to stand in for time. Unfortunately, there is no evidence for such a sampling clock in human nervous systems or for a memory that is exclusively serial (Port 1990, Port et al. 1995). The method of discrete sampling just described is a useful technological trick but is not a plausible biological process. Of course, all that is achieved by this sampling process is transduction into the nervous system; pattern recognition processes have not yet begun.

It can be seen that trying to reconcile two parallel types of cognitive time, one discrete and the other continuous, leads to many conceptual problems and paradoxes. Phonologists have shown little interest in working out any continuous interpretation of their discrete models. The incommensurability between discrete and continuous time is surely one reason why linguists often consider most cognitive disciplines outside linguistics to be irrelevant (e.g., experimental psychology, neuroscience and experimental phonetics). They may assume these time-dependent fields can have no direct impact on language, a pure symbol system. (Of course, the discrete subdisciplines of mathematics and logic are taken seriously indeed). Correspondingly, this is why scientists from many other disciplines have great difficulty understanding the basic mission that linguistics has taken on. If one believes that cognitive and linguistic events could not, in fact, exhibit symptoms of existence either in space or time, then, since real physical and physiological events do, there is no way to make them fit together.[2] In recent years, some phonologists have

---

[2] In a recent paper Chomsky rejects arguments like ours that formal systems are incompatible with biological principles. He suggests the incompatibility is really only between formal systems and biological principles as understood today (Chomsky 2000). This reveals more about our primitive level of understanding of biological systems, he argues, than about any difficulties with linguistic theory in principle. Linguists should continue doing what they do. Sooner or later the scientific understanding of biological systems will catch up and the reconciliation will seem perfectly obvious and natural. At least his making this argument shows that Chomsky understands the degree to which his approach is committed to mechanisms that are nonbiological. Jackendoff (2001) takes a similar stand arguing that nonlinguist scientists fail to appreciate the challenges posed by linguistic data or the achievements of linguistic analysis.

claimed that phonological rules are so formal and abstract that we are invited to imagine many thousands of rules or constraint evaluations but not to be concerned about the the time required to do them (McCarthy 2002).

2.2 Formal Symbol Systems. To appreciate the seriousness of the problem of the incommensurability between real time and discrete time, it may be useful to review some of the essential properties of formal symbolic systems. Linguists often assume language is symbolic at all levels, but less attention has been paid to exactly what properties a symbol token in such a system must exhibit in order for the computational system to work as intended. In western science, the notion of the symbol as a physical token `standing for' something else seems to be a fairly recent idea. It may have been first formulated by C. S. Pierce late in the 19th century (see Fetzer 2001). This idea contributed to twentieth century mathematics dealing with strings of tokens (e.g., the work on discrete models of computation by Turing, Shannon and Chomsky), and eventually to both programming languages and modern linguistics. Today, it seems symbols are employed in three basic domains: (1) for doing mathematical reasoning (e.g., mathematics, logic, etc), (2) in computer software, and (3) in theories of cognition (e.g., Chomsky 1965, Newell et al. 1976, Fodor 1975). In formal reasoning (e.g., doing logical proofs, long division, writing or debugging a computer program, etc.), operations are performed on symbolic structures by trained human thinkers. Throughout training and practice, steps in the formal reasoning process are typically supported by `props' external to the body. That is, conscious formal reasoning that requires more than a couple serial steps always relies on a blackboard or a piece of paper (see Clark 1997, 2004). An important form of mental scaffolding is writing graphic tokens on paper. In the past half century, much more powerful scaffolding has become available: running programs on a computer that reads and writes such tokens. In computer hardware, formal methods are physically automated using symbol tokens coded into physical bits and manipulated by a machine in discrete time. The third domain for symbolic theories lies, of course, in a particular view of various cognitive operations involved in human language and human reasoning (Chomsky 1965, Newell and Simon 1976, Fodor 1975, Fodor et al. 1988). The symbol tokens proposed for language include, of course, sentences and words but all morphology is eventually mapped onto strings of phonological feature vectors.

Symbol tokens must exhibit the properties shown in (1) to function as advertised in symbolic systems (see Haugeland 1985 for further discussion).

(1) Properties of Symbol Tokens
   a) a symbol token is either apriori, or composed of apriori tokens
   b) it can be perfectly recognized and perfectly produced by the symbol processor; that is, it is digital
   c) it is static, i.e., definable at a single timepoint in a discrete-time system

All symbols are either apriori or composed from apriori atoms. Some set of units is available at the time of origin of any symbolic system from which all further data structures are composed. In the case of logic or mathematics, an initial set of specific units is simply postulated e.g., `Let there be the integers (or proposition p, or points and lines, etc.).' In computing, the

---

But Jackendoff fails to consider the degree to which those linguistic analyses rest on a foundation of highly speculative assumptions and does not address the problem the realtime operation of formal cognitive models.

analogous aprioris are the physical bit string patterns (that is, voltage patterns in separate `wells' in a silicon chip) and the hardware instruction set that causes particular operations to occur. Of course, the units and primitive operations were all engineered into the hardware itself, and are thus obviously apriori from the perspective of the programmer. Similarly, in phonology, according to Chomsky and Halle (1968), it is fairly obvious `that there must be a rich system of apriori properties – of essential linguistic universals.' This follows from the fact that children acquire language very quickly with no tutoring despite wide differences in intelligence (4). The theory supposes that children are able to use their innate phonetic alphabet to represent words and morphemes roughly correctly when they are spoken. If someone says *cookie* several times, children can recognize the identity using their apriori transcription scheme. The phonetic alphabet provides an important bootstrap for language learning.

So the first problem for a symbolic model of language is what the apriori symbol tokens are. Since they are premises of the theory, there must be a finite number at most. What is the initial vocabulary of symbol tokens from which the morphology and syntax are composed? In Chapter 7 of *SPE*, Chomsky and Halle offered a set of 40-50 phonetic features for use by linguists as a partial answer. Very few features have been added to this list since 1968. The discovery of the correct list of all innate symbols is one of the primary missions of research in linguistics (Chomsky 1965). However, since 1968, the majority of phonologists have simply relied on the feature set proposed in Chapter 7 with virtually no effort expended at carefully evaluating the appropriateness of this list. This set includes at least segmental features like [Consonantal], [Voiced], [High], [Coronal], and [Continuant] which are combined into a vector of values of all the features for each phonetic segment.[3] These atoms serve as building blocks for the construction of descriptive statements about various languages by the language learner (as well as by the linguist).

In order to function as intended, the symbol tokens must be, using Haugeland's term, digital, that is, perfectly distinct from each other and reliably recognizable by the computational equipment that implements the system (Haugeland 1985). This is an absolute requirement in order for the computational mechanisms to successfully manipulate the symbols during the processing of rules. The atomic units from which all linguistic structures are constructed must be physically discrete because it is only their physical form that determines what operations apply to them. Thus the numerals, p and q, x and y, etc. must be reliably distinct for the system (i.e., the logician, mathematician, computer, etc.) using them. Fodor & Pylyshyn (1988) point to this property of physical distinctiveness as a key argument for the superiority of computational models over neural network models, since in computational models, patterns with similar content are built from physically similar parts. Thus, for Chomsky and Halle, the similarities and dissimilarities of [t] vs. [d] are directly expressed in the physical form of some units corresponding to plus and minus values of the phonetic features in the linguistic description. On the other hand, in connectionist

---

[3] One occasional source of confusion has been Chomsky and Halle's insistence that these features were to be understood as `scales' even though they never proposed any phonetic features that were not binary. They suggested the scales might have multiple values, but did not claim they were continuous-valued scales. (If they had it would have created the problem that their phonetic space would have an infinite set of elements.) Of course, anywhere the phonetics might suggest a continuum (e.g., vowel height, place of articulation between the teeth and the uvula, voice-onset time, etc.), they simply postulated enough features to cover the phonological categories they needed with binary features. From our perspective, binarity is not the issue. The problems are discreteness and the claim that their phonetic basis is nevertheless independent of phonological considerations.

models, they claimed, activation patterns cannot be physically examined to determine anything about content (so that application of a rule could depend on the pattern found). In backpropagation-based neural network models, the content of patterns is not, they claimed, directly specified by the physical pattern of node activations itself.

Modern computers, of course, are digital since the read and write operations on bit strings produce errors only once in trillions of cycles (and use software error correction codes to fix almost all that occur). The formal-language assumption implies the brain would have to identify and produce phonetic feature tokens nearly as well as a computer if we are to imagine a complex phonological grammar with hundreds or thousands of formal rules or constraints to be executed. For the program-executing device, any two atomic units must either be identical or else perfectly distinct. And any failures of discreteness should be quite catastrophic since such errors are randomly related to content. We should expect the equivalent of a `crash' whenever discreteness fails. Yet, we know that human language exhibits great robustness to errors of many kinds (including high background noise, foreign accent, food chewing, cochlear implant, etc.). How can this discrepancy be accounted for?

The third property of symbol tokens that is relevant for our purposes is, (3c), that symbol tokens must be static. Since symbolic or computational models always function in discrete time, it must be the case that at each relevant time point (that is, e.g., at each tick of the clock that governs a discrete system in real time), all relevant symbolic information is available. For example, if a rule converts apical stops into flaps, then there must be some time point at which the features that figure in the rule, e.g., [–continuant, +voice, +apical] etc., are all fully specified and mutually synchronized while the rule applies in a single time step. Thus, tokens in a symbolic system cannot either unfold asynchronously (i.e., each on its own timescale) or in continuous time, but must have some discrete symbolic value only at each relevant clock tick[4]. In between the ticks, of course, the state of the computational system is undefined. In simple terms, a discrete system does not permit any form of nondiscrete-time description or continuous-time predictions.

Finally, in addition to the properties listed in (3) it seems clear that the apriori symbol token set must be limited in size. An alphabet making extremely fine distinctions would be difficult for the child to use (since different productions of the same words would, in general, be different) – nor could linguists make predictions since every utterance is likely to be different from every other utterance. So, practically speaking, the segments must come from a small list, if they are to serve a bootstrapping role for infants and to serve linguists as a basis for comparison across languages. In a computer there are only two apriori symbol tokens, usually called 0 and 1 (organized in sets of, e.g., 16 or 32 bits). If new phonetic aprioris may be added to the theory without limit, a theory threatens to become adhoc. Jakobson's distinctive feature set, at 12-15 features, was very small indeed (Jakobson, Fant and Halle 1952), but even Chomsky and Halle enlarged their set only to 40 or 50 phonetic features.[5] The Chomsky and Halle features, for example, can be combined to implicitly specify only 3 or 4 values of voice-onset time (VOT). But what if there are not just a

[4] Of course, there is nothing to prevent simulation of continuous time with discrete sampling, but this is not what the symbolic, or computational, hypothesis about language claims. See Port et al. 1995, for discussion of time sampling in cognitive models.

[5] Ladefoged (1965) estimated that these features would produce over 12,000 consonant segments and several thousand vowels (quoted in Sampson 1977). Is this too large a number to be plausible? It is hard to say without more information, but not necessarily. Probably a million would be too many.

few values of VOT under linguistic control, but several hundred (e.g., to include language-specific and context-determined VOT targets)? Then, in general, due to random noise, repetitions of a word by the same or different speakers will tend not to be heard or transcribed the same by the infant language learner. Every word would have a huge set of variant phonetic spellings which could confound the learning of vocabulary. The innate phonetic alphabet must be very small (and the equivalence classes fairly large) to keep this problem under control. Of course, a final reason for a small list is simply that each feature or segment type is claimed to have some innate special-purpose identification mechanism.

It can be seen that these properties of the space of phonetic symbols suitable for a formal theory of language are not trivial. Very strong demands are placed on these tokens which imply many testable empirical claims. It seems to us that the assumption that the phonology of a language is a formal system was made without full consideration by linguists of what is required in cognitive hardware for this to be true. The issue we address below is whether there is phonetic evidence to support these claims.

2.3 Biological Instantiation. Could discrete operations on symbolic tokens take place in a human brain? True formal symbols assume some rather nonbiological properties, such as, timelessness and practically perfect performance (digitality). It is one thing for humans to manipulate arithmetic symbols consciously leaning on the support of paper and pencil so each step can be written down in visual symbols and checked for accuracy. But that is not what is called for by a symbolic or computational model of linguistic cognition. A computer can meet the requirements of formal operations due to specialized hardware with an apriori discrete-time clock to process the symbolic structures encoded as bit strings. But it is another matter altogether to assume that formal structures are actually processed in discrete time by the human brain. The difficulty is that if we study language as a facet of actual physical humans, then all its processes and its products must have some location and extent both in real time and in space. If language is symbolic, that is, if it uses technical symbol tokens, then our brains must do `cognitive discrete time' for all grammatical processes. After all, the computer is the purest example of an implemented symbol system running in time. Bit strings are discrete, with an apriori vocabulary of size 2 and they exist in real time and physical space as electrical charges in discrete cells in silicon at each tick of the computer's clock. The hardware was engineered so that at each clock tick, every bit has the value of either zero or one. But what of the human brain? How might it approximate discrete time with discrete symbols?

If natural language is implemented as a grammar in a discrete-time symbol processing device, we should find plenty of evidence of this discreteness. It should be very easy to see. How would one verify that a conventional movie-theater cinema is actually a discrete-time system? Look at how it displays a stroboscopic pattern. When the periodic pattern is at or very close to the sampling frequency (or its multiples), the motion is slowed or stopped (which is why rolling wagon wheels appear to move slowly forward or backward). At the very least, the mechanism must be accessible to scientific research methods that investigate events in space and time. Of course, there are many kinds of periodic and temporally discrete behaviors in the human brain including periodic oscillations in, e.g., EEG.[6] None of these periodicities have either the constant rate or temporal synchronies that would seem to be necessary.

---

[6] Of course, single neural units exhibit action potentials that are discrete: fire or not fire. Mathematicians and neuroscientists were inspired, early in the 1950s , to propose digital models of neural activity. But soon it became clear that for any unit to fire depends on temporally integrated inputs from a large number of other neurons so digital models of neurocognition have given way to essentially continuous ones. Synchronous

But if one still has faith there might be a discrete-time level of activity involved in language, then clearly the correct way to study such phenomena is by gathering data in continuous time in order to discover the conditions for temporal discreteness and to understand how discrete performance is achieved. The view of SPE is that there is no need to investigate these issues since there is a sharp apriori divide between language as a serial-time structure (implementing a theory of phonology) and speech as a continuous-time event. This claim of the independence of cognitive time from physiological time makes strong predictions about neurocognitive behavior that linguists have not bothered to investigate and which unfortunately provide no empirical support for the formalist hypothesis.

Another reason for rejecting the view that language is essentially formal is that it is clear that language is biologically a spoken medium not a written one. When it comes to doing linguistic research, however, most linguists, including phonologists, rely exclusively on forms of written language – on segmental phonetic or orthographic transcriptions – as their input data and ignore all continuous-time aspects of real speech. All written versions of language are derived from speech, however, by perceptual processes that depend on some particular human transcriber. But it is well-known that transcription is influenced by one's native language to large degree (see Strange 1995b for a review of many issues). Thus many of the properties of orthography, such as its discreteness, the restriction to a small inventory of elements and regularization into words and sentences, may be inadvertently and artificially imposed by the transcriber. Phonetic transcriptions must always be viewed as inherently untrustworthy.

Chomsky and Halle apparently hoped, nevertheless, that this kind of discrete phonetic inventory would turn out to account for all linguistic expressions. They may well have assumed that phonetic research would soon converge on some specific list of universal atoms so that all utterances would be discrete structures assembled from this set of sound atoms in the same way that the English words on this page are composed from a small inventory of letters. But why isn't the identity of the segment or feature list as transparent to speakers (as well as to linguists) as the identity of the graphic letters on this page is? To find out how many letter shapes are used in this issue of the journal, one just needs to make a list. But there is no consensus in the field on how many phonemes (or other phonological primitives) are in use in a language like English or any other language. Of course, the sounds needn't all be identical across languages, just as individual letters exhibit variable forms. But will transcribers with different native languages be able to produce identical transcriptions in a universal alphabet? There is surprisingly little direct investigation of a native-language effect on the transcriptions of language professionals (e.g., linguists, language teachers, speech therapists, etc.), but many findings in the phonetics literature plus personal experience teaching phonetics to international students support deep skepticism that anything approaching identity of transcription is possible (Lieberman 1965, Shriberg et al. 1991, Eisen et al. 1992, Flege 1993). And the sources of these differences would be very difficult to eradicate (Logan et al. 1991, Strange 1995b). Furthermore, there are regularly new results about the phonologies of languages displaying properties that are not obviously describable with the tools of discrete symbol identification (e.g., properties such as C/V ratio, mora timing, foot structure, rhythmic performance, etc).

---

firing in certain physically remote cortical regions has also been discovered but this involves only a tiny fraction of the cortex and is not a plausible implementation of what linguistic models require.

12

3. Pro and Con Evidence.

It must be acknowledged that the hypothesis that there is a fixed set of discrete symbols used in all languages does have various kinds of supportive evidence despite our disagreements with it. Even though empirical arguments in support of this hypothesis have always been avoided by Halle, Chomsky and other generativists in favor of simple assertion, there are a variety of kinds of evidence that could be raised to support it and may contribute to the intuitive rightness of formal descriptions. We first present the arguments in favor of such a phonetics in (2), each with a brief counterargument, and then summarize the main arguments against a fixed universal phonetic alphabet in (3).

(2) Possible Evidence Supporting a Fixed Alphabet-like Phonetic Inventory

a) there is a matrix-like structure to phonological inventories in most languages suggesting discrete feature combination

b) introspection reveals discreteness of categories

c) there are some apparent universals of speech sounds

d) the same sounds sometimes occur in different languages

The proposal that the linguistically relevant sounds of language are discrete and drawn from a fixed set seems initially plausible for a variety of reasons, including, as in (2a), the obvious table-like structure of lexical items in a dictionary. Looking at English, we find word sets suggesting a front-vowel series, like *beat- bit- bet- bat* with the same vowel series repeated in sets like *seal- sill- sell- Sal* and *reefer- rift- left- laughter*, etc. Similarly for consonant contrasts, we find *bad- pad, Bill- pill, black- plaque, Libby- lippy,* etc. and analogously, *dad-tad, dill-till, drip-trip*, etc. Continuing this process can lead to a segment matrix like:

b	d	g
p	t	k
m	n	ŋ

All languages have many such tables of minimally distinct sets of words in their lexicon. The key observation to note about these cases is that generally there appear to be few if any tokens exploiting regions between members of the maximum set of vowel and consonant categories. The discovery of such word sets in most languages has suggested to some that languages may employ some discrete set of V and C contrasts for `spelling' lexical items (Fischer-Jorgensen, 1952). In short, it is obvious that within specific languages or dialects, there is a fairly small set of perceptually salient sound categories that seem to be permuted to make much of the vocabulary.

An important observation, however, is that what these phenomena support is only phonological categorization on a language-specific basis. This evidence offers no support at all for universal phonetic sound classes. It seems that meaningful words and morphemes in every language imply some kind of limited set of contrastive sound types within that language. But how speakers arrive at these categories is the critical research question. The claim of universal sound categories is only one possible explanation for these observations. Another way to interpret this matrix-like structure is as evidence for **symboloids** – categorical patterns of sound that resemble symbols to some degree but are not technical symbol tokens (van Gelder and Port 1995).

The second kind of evidence in support of a fixed phonetic inventory, (2b), is our introspective interpretation of sound categories. Speech just sounds discrete to most of us. So,

when we listen to speakers of other languages, we still tend to hear discrete categories (those of our native language and possibly of other languages we are experienced with). And if someone makes a slow vowel glissando from, say, [i] to [æ], the vowel may, for an English speaker, seem to jump perceptually from [i] to [e] to [ɛ] to [æ]. Experiments have verified this categorical tendency repeatedly (Liberman et al. 1968, Kuhl et al. 1995). The `categorical perception effect' shows that during the identification of one's native vowels and consonants differing along some acoustic-phonetic continuum, listeners exhibit sharp category boundaries. The evidence for categories is that subjects have reduced ability to discriminate stimulus differences within a category but enhanced ability to distinguish between categories. These results show there is a strong bias toward sound categories. But the experiments were done on adult speakers. Such introspectively discrete categories are persuasive as perceptual effects, but we must keep in mind that speakers of different languages will hear a different number of categories and hear the category boundaries in different places than other listeners. We are aware of no studies showing any widespread tendency toward cross-language categories – as required by the hypothesis of universal phonetics. The fact that we can use the IPA graphic symbols in this way only shows how easily we can ignore the differences in usage of the symbols between languages.

The third kind of evidence supporting a fixed universal phonetic space is, (2c), that some speech production data suggest phonetic universals. For example, in English, in word-initial position, there is an obvious phonological contrast between the distributions of measured VOT values for words like *tip* and *dip*. Looking across languages, the variable of VOT seems to exhibit three clear modes: commonly called prevoiced, short-lag (or unaspirated) and long-lag (or aspirated) (see Lisker et al. 1964, Figure 8 showing a VOT histogram of utterance-initial stops from 10 languages). Thus, some aspects of VOT data could be said to support a universal inventory with only a few categories of VOT. Similarly, no language seems to use more than about four levels of contrastive VOT or vowel height. Perhaps that is as many VOT types as there are.

While it is true that Lisker and Abramson's cross-language data on VOT did have 3 prominent modes, individual languages exhibit target VOTs at many different locations along the VOT scale, not just at or near the 3 modes. Thus, in their data the mean value for word-initial /k/ in English was about 50 ms while in Korean it was about 125 ms. So their data do not support the Chomsky-Halle claim that only 4 values of VOT targets are possible in natural languages. Measurements of VOT in English across linguistic contexts (varying place of articulation, stress, vowel identity, presence of following glides, word length, etc.) reveal distributions suggesting a very large number of VOT targets (Lisker et al. 1967, Port et al. 1979, Cho et al. 1999). Of course, given articulatory noise, some of these targets may be, for practical purposes, the same across languages. But as Pierrehumbert (2001) puts it `it is not possible to point to a single case in which analogous phonemes in two different languages display exactly the same phonetic targets and the same pattern of [contextual] variation.'

As shown in (2d), another suggestive fact is that some sounds are found in many different languages. For example, many languages have sounds described roughly as [i, ɑ, u] or [d, n, l, t, s], etc. The reason for this, one might hypothesize, is that speech sounds are all drawn from a fixed, universally available list. Still, although some sounds seem to appear in many languages, (a) they are in fact not exactly the same, merely similar (e.g., Port et al. 1980, Flege et al.1986, Bybee 2001: 66, Maddieson 2003), and (b) there are also many sounds that are isolates, that is, sounds which have been found in only one or a very small set of languages. Given nearly identical vocal tracts, the universals of physical acoustics and the constraints imposed by human hearing, it is no surprise that different linguistic communities discover many similar sounds to use for communication. But

14

this provides no justification to postulate innate universals (e.g., Stevens 1989) to account for such similarities. Some sounds have apparently useful properties, e.g., they are relatively easy to produce and/or yield distinctive acoustic effects, so many languages have gravitated toward them over the generations for word specification (Lindblom 1990). There is no reason to additionally invoke an innate origin for these sounds.

Despite the above arguments in its favor, there are many arguments that can be raised against fixed universal phonetics stemming from the phonetics literature that seem to be largely overlooked by phonologists and whose implications are ignored even by some phoneticians. These arguments are summarized in (3),

(3) Evidence Against a Fixed Phonetic Inventory
a) Transcription is very difficult, inconsistent and errorful
b) Phonetic spaces are highly asymmetrical
c) Language-specific categorization appears very early in language acquisition
d) Many language-specific phenomena are incompatible with serially ordered phonetic symbols.

Phonetic transcription is notoriously difficult and inconsistent. As discussed above, even professional students of language will transcribe utterances in their own language as well as other languages quite differently (Lieberman 1965, Shriberg et al. 1991). Further, every linguist knows that very often it is not obvious what phonological units make up a stretch of speech -- or even how many segments there are. To take a typical simple example, is an English syllable like chide made of 3 sound units (CVC), or 4 (e.g., CCVC) or 5 (as CCVGC)? The initial consonant has two parts and the medial vowel has either two parts or a gliding motion both acoustically and articulatorily. Linguists have endorsed varying analyses of these cases in various languages through the years. So, despite the intuitions mentioned in 2b, neither speakers nor linguists have clarity or unanimity of intuitions on the number of vowels or consonants in these cases. Languages also typically exhibit `positions of neutralization' where identity is also quite unclear. For example, is the vowel in *beer* the same as the vowel in *bead* or the one in *bid*? What is the stop after [s] in *store*? One could make an argument for either /t/ or /d/ or even for a third alternative. The phonetic transcription of such words is completely uncertain for both native speakers and linguists. Is at least the number of syllables in a text clear? It may be clear for words like *sap, paper, banana*, etc. But what about words like *mare* (cf. *mayor)* or *air* (cf. *error), memory* (cf. *mammary*) or *police* (often pronounced *p'lice*)? Cases like these are found in every language -- as anyone knows who has ever tried to do phonological description. So it is simply not true that the phonetic introspections of native speakers (or anyone else) are always clear and unambiguous about the number or identity of segments (or syllables or any other units). They are only consistent in the simplest cases.

There are also many asymmetries in phonetics, as in 3b. Chomsky and Halle claimed the universals of phonetics are always in the form of segmental features, like [Voicing, Height, Nasal], etc. because some phonological properties can be combined independently. But only a few phonetic objects are combinatory in this way, like, e.g, vowel height and backness or stop place and some manners. The features [Lateral] and [Retroflexed], for example, are not combined with any other place of articulation than apical. So what is the rationale for creating a place-independent feature for these properties? And the apical flap seems perceptually like a segment, not a feature. If flapping is really a component `feature,' it is oddly combinable only with apical place. The implication of such observations is that the description of speech sounds as a uniform matrix of

values for features that are mutually independent, as suggested above in (2a), represents a Procrustean schematization of phonetics. This regularization may offer methodological conveniences for writing formal rules but, looked at critically, has only very sketchy empirical support.

As suggested in 3c, children learn to categorize the speech sounds in a way peculiar to their native language before they do much talking (Werker & Tees, 1984). It is clear now that the strongly categorized perception of speech sounds that adults experience is not innate but is actually learned very early in life – much of it before the acquisition of a child's first word. Throughout the first year of life, children can differentiate many sound contrasts in many languages.[7] But by the beginning of the second year, infants lose the ability to differentiate some sounds that are not employed in the ambient language (Werker et al. 1984, Kuhl et al. 1995). This implies that all adult humans are burdened with strongly biased perceptual systems. This bias is guaranteed to distort, not only the sounds of our native language (so we fail to notice some differences that speakers of other languages might easily detect), but the sounds of any other language as well (where we will miss distinctions that are obvious to its speakers). There is no evidence that taking a course or two in phonetic transcription, as linguists do, will eliminate these biases. This evidence alone shows authoritative phonetic transcriptions are quite impossible.

The fourth argument against universal phonetics, 3d, is that much additional evidence against the claim of a universal segmental phonetic inventory has been provided by the phonetics literature of the past 40 years. This research tradition, found in the *Journal of the Acoustical Society of America, Journal of Phonetics, Phonetica, Language and Speech*, etc., has demonstrated the overwhelming variety of ways in which languages can differentiate classes of words. Only a few specific cases are presented here that happen to focus on temporal issues. Section 3 presents a sampling of the specific phonetic evidence showing the Chomsky-Halle claims to be insupportable. We also present evidence that words and other apparent linguistic units are sometimes nondiscretely different from each other (unlike printed letters and zeros and ones). Data show that linguistic units like words and phonemes are not always timeless static objects, but turn out sometimes to be necessarily and essentially temporal. By this we mean they are defined in terms of non-segmental properties (such as durational ratios) and distributed widely across the time of production of an utterance. If we can successfully show the existence of even one linguistic structure in any language that is essentially temporal (as opposed to merely implemented temporally), or if a case of genuine category nondiscreteness exists, then the bold symbolic phonology assumption – the claim that phonetics and phonology are invariably discrete and static – would be very seriously compromised.

4. Temporal Phonetics.

Research on speech production and perception demonstrated from the earliest era (Joos 1948, Liberman et al. 1956) that manipulation of aspects of speech timing could influence listeners' perceptual judgments. Thus, vowel duration may influence judgments of vowel Length and

[7] Of course, the ability to discriminate a difference is not the same as and does not imply the ability to represent what those differences are. So, the famous study by Eimas et al. (1971) showing the ability of newborns to discriminate a place of articulation difference is a surprise, but does not offer support for the Chomsky-Halle claim that infants can use some universal phonetic perception system to `transcribe,' that is, cognitively represent to themselves, the ambient language.

16

consonant Voicing in many languages and VOT (Lisker et al. 1964, 1967) influences judgments of Voicing – to mention just a few examples (see classic reviews by Lehiste 1970 and Klatt 1976). Linguistic theorists were forced to address the problem of the discrepancy between symbolic phonetic transcriptions and a real-time description of speech. Chomsky and Halle addressed part of the discrepancy by postulating universal implementation rules to convert serially ordered segmental feature vectors into continuous-time speech gestures. Halle and Stevens (1980) proposed hypothetical implementation rules that would interpret, for example, a static feature of [Glottal Tension] in a way that results in a delay in the voice-onset by some number of milliseconds after the release of the stop. So the temporal effect of long VOT was interpreted as merely an articulatory epiphenomenon of a change in a static feature value (see Lisker and Abramson 1971 for an early critique of such interpretations).

Notice that the implementation-rule solution rests on an important claim about the phonetic implementation, one that is vulnerable. The Halle-Stevens-Chomsky account of speech timing is tenable relative to their theory only if the phonetic implementation processes are universal. This is critical because only if the implementation of discrete phonetic symbols works the same for all languages could it be true that utterances are composed entirely of symbols and differ from each other linguistically only in symbol-sized steps. The phonology is supposed to specify the language-specific properties of speech, while the phonetic inventory and its implementation is supposed to be universal. This must be true if the segmented phonetic space is to include all `the phonetic capabilities of man' (Chomsky et al. 1968). Of course, if one may multiply static features without limit, then one can claim that any new awkward temporal data simply requires the addition of a new apriori phonetic feature that may not have been seen before but has just the temporal consequences one observes. The following sections present specific evidence that appears incompatible with the now-traditional story of a universal discrete phonetic inventory.

4.1 English and German voicing. Data gathered over the past 30 years make it fairly clear that one distinction between some pairs of words in English is an intrinsically temporal property. English and German offer cases where two sound classes differ from each other in a particular durational ratio between some adjacent acoustic (or articulatory) segments in syllable rhymes. English has a contrast among stops and fricatives, e.g., /b, d, g, z / vs. /p, t, k, s /, differing in [voicing] or [tenseness] in non-syllable-initial positions. Pairs of words like *lab-lap, build-built* and *rabid-rapid* contrast in this temporal feature, as do German pairs like *Bunde-bunte* (`club'-Plur, `colorful'-Nom, Sngl) and *Egge-Ecke (*`harrow'-Nom, `corner'-Nom*).* One characteristic of this contrast in both languages is that it depends significantly on a pattern of relative timing to maintain the distinction. If two segment types differ from each other in duration, one might argue that this results from a static feature that has some concomitant temporal effects (Halle et al. 1980). But if specification of the feature requires comparing the durations of two or more segmental intervals, then the claim that this is achieved by implementing symbolic units in independent ways begins to strain credulity. For words with syllable-final or post-stress voiceless consonants, like English *lap, rapid, lumper, scalper*, the preceding stressed vowel (and any nasal or glide) is shorter while the stop closure is longer in the /p/ words relative to corresponding words with /b/, e.g., *lab, rabid, lumber, album* (Peterson et al. 1960, Lisker, 1984, Port, 1981a, 1981b).[8] A simple measure that

[8] The other main cue for this feature is glottal oscillations during the closure – but in English, stops without glottal oscillations still sound voiced if the closure duration is short relative to the preceding vowel.

makes use of these opposing durational trends is the ratio of the vowel duration to the stop duration. This value changes from around unity for voiceless obstruents like /p/ and /s/ (where the V and following C are about the same), to values of two or three for voiced obstruents like /b/ and /z/, where the V and any additional sonorants are noticeably longer than the C.

Of course, since speakers typically talk at different speaking rates, the absolute durations of the segments are highly variable when measured in *ms*. For example, Port (1981b) had subjects produce minimal word sets like *dig, digger, diggerly* and *Dick, dicker, dickerly* (some real and some plausible nonsense words). The stressed vowel /ɪ/, as well as the /k/ and /g/, got shorter as additional syllables were added, but the ratio of vowel duration to stop closure duration remained nearly constant in words with the same voicing feature. The ratios did change, for example, between the words above with /ɪ/ and the nonsense words with the long vowel /i/ (in *deeg, deeger, deegerly*, and *deek, deeker, deekerly)*. So the vowel durations are affected by the vowel change (from /ɪ/ to /i/) while the stop closures are not. Clearly absolute durational values in milliseconds cannot be employed to specify this voicing information, since in that case listeners would both perceive more /k/s and /s/s after short vowels and at slow rates and more /g/s and /z/s after long vowels and at faster rates. The ratio of V to C tended to be relatively invariant over many, though not all, changes in context.

Additional evidence is that perceptual experiments with manipulated natural speech or synthetic speech confirm that it is the relative durations that determine judgments between minimal pairs like *lab-lap* and *rabid-rapid* whenever other cues to the voicing feature are ambiguous (that is, in particular, when the consonant closure does not have glottal pulsing) (e.g. Lisker 1984, Port et al.1982). Port (1981b) called this relationship `V/C ratio,' the relative duration of a vowel to the following obstruent constriction duration (or equivalently its inverse C/V). This ratio is relatively invariant across changes in speaking rate, syllable stress and segmental context as shown in *Figure 1* (Port et al. 1982).

In several other Germanic languages, similar measurements of speech production timing (Elert 1964, Port et al. 1983, Pind 1995) and perceptual experiments using manipulations of V and C durations have shown similar results -- that listeners pay attention especially to the relative duration of a vowel and the constriction duration of a following obstruent (Port et al. 1983, Pind 1995, Bannert 1974). In Swedish the longV-shortC vs. shortV-longC contrast is partly independent of voicing, with minimal pairs like *vit-vitt* [vi:t, vit:] white-Basic, white-Neuter, and *bred-brett* [bre:t, bret:] broad-Basic, broad-Neuter (Sigurd, 1965) and Icelandic *baka-bakka* to bake, burden-Acc (Pind, 1995).

According to the Chomsky-Halle theory, the V/C durational ratio must be a temporal universal that is presumably triggered by some segmental feature. It is difficult to preclude such a claim, but this is surely adhoc. It is one thing to say that some static feature causes a delay or lengthening of some segment, but quite another to claim that a static feature causes a conspiratorial adjustment of the relative duration of a vowel to a following consonant closure. An objective look at the phenomena suggest that these languages exhibit a contrastive feature that is intrinsically temporal.

18

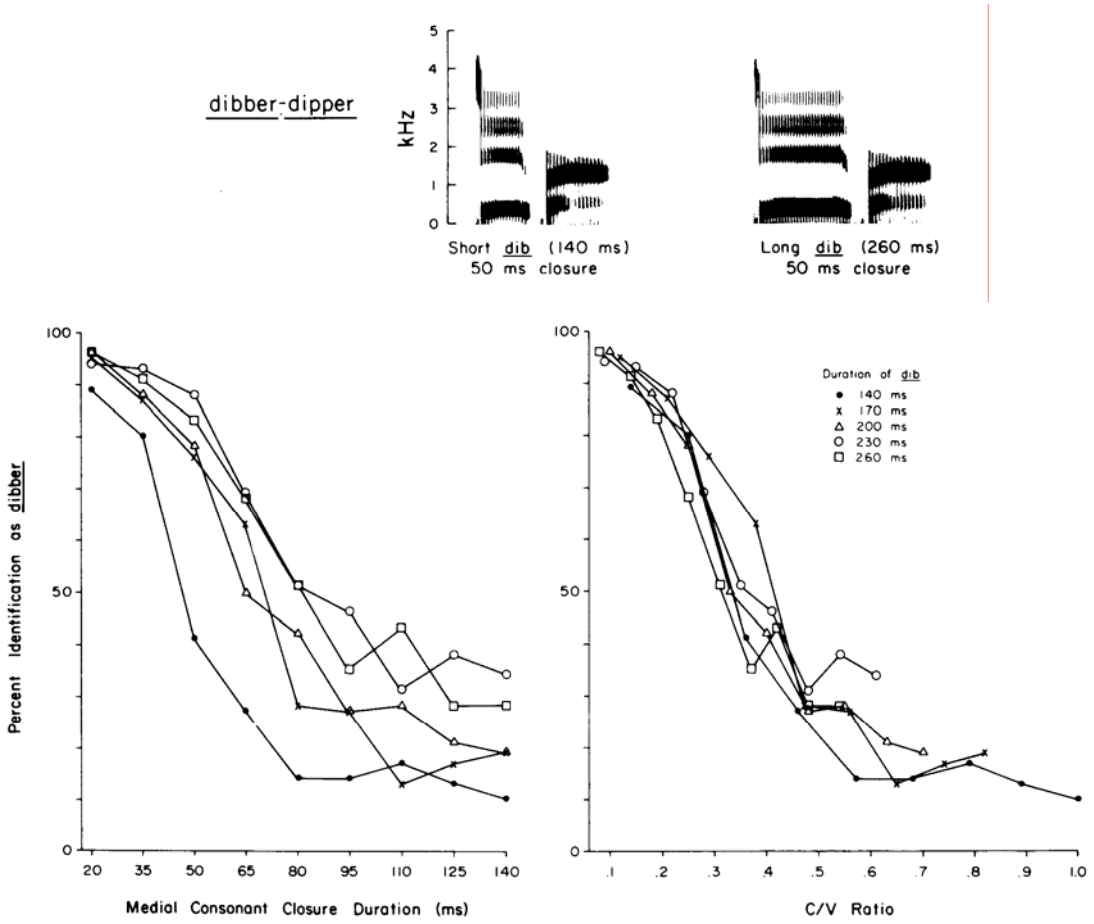

Figure 1. Illustration of some stimuli and results from Experiment 1 of Port & Dalby's (1982) study on consonant/vowel ratio as a cue for voicing in English. The top panel shows sound spectrograms of some of the synthetic stimuli. These examples show the shortest (140 ms) and longest (260 ms) vowel durations for /dɪb/. For each vowel duration step, nine different silent medial-stop closure durations were constructed. Subjects listened to the stimuli and were asked if they heard dibber or dipper. The bottom panel shows the results of the forced-choice identification. The left bottom panel shows the identification scores as a function of medial-stop closure-duration in ms. For the shortest vowel the C duration at the crossover is about 50 ms and for the longest, over 80 ms. The bottom right panel shows the same data, plotted as function of the single variable of consonant/vowel ratio. Note the large reduction of variation as all perceptual boundaries (50% identification) for dibber vs. dipper cluster near a C/V ratio of 0.35 or a V:C ratio of about 3:1.

4.2 Temporal Implementation Rules. Even leaving aside these descriptive concerns, there are still major difficulties in principle with any rules of temporal implementation that depend on phonetic context. Since the rules are static, they can only specify duration as some kind of number, that is, as some static quantity that will be interpreted in temporal terms by the performance system.

Let's assume for the moment that implementation rules could supply an inherent duration in *ms* for each segment type, e.g., 45 *ms* for a [b] closure and 60 *ms* for a [p] closure. Then a context implementation rule could adjust the duration of the preceding vowel (via multiplication or addition/subtraction) to be longer before a [b] (or shorter before [p]). The result of these two rules would then be a target duration in *ms* for both the vowel and consonant closure (see Klatt 1976, Port 1981b, Van Santen 1996 for temporal implementation schemes of this general form).

The first problem is the issue of what use there might be of the target durations specified in milliseconds (whether they are stored or computed). Who or what will be able to use these numbers to actually achieve the target duration of some number of *ms* for a segment? There is no serious model in the motor control literature that could employ such absolute specifications. A new theory of motor control is needed to make use of such `specs' to generate speech gestures with a particular target duration (see Fowler at al.1981, Port et al. 1995). The general difficulty is that a motor execution system that is to interpret specifications in milliseconds would need to have its own timer (independent of any clock for cognitive discrete-time) in order to determine when *N ms* have transpired. The second problem is that durations in milliseconds seem fundamentally misguided since speakers talk at a range of rates. So for this reason alone, it seems that it should be relative durations that any rules compute, not absolute durations (see Port 1990, Port et al. 1995). Third, these models, such as Klatt (1976) and van Santen (1996) typically specify durations one segment at a time. Longer intervals (such as interstress intervals, syllable durations, etc.) can get their duration only when the individual segments have been computed and added up. But such a system has no way to arrange for global timing patterns (e.g., periodic stress timing or mora timing). Yet humans find it very easy to produce speech with regular periodicity at a global (e.g., syllable, foot or phrasal) level, such as when chanting, singing or reciting poetry (Cummins et al. 1998, Tajima et al. 2003, Port 2002, 2003).

Despite these implausible features, one cannot prove the impossibility of an account using temporal implementation rules. After all, if formal models can implement a Turing machine, they can handle relational temporal phenomena by some brute-force method. But an implementational solution along this line is only interesting if some specific constraints are applied to the class of acceptable formal models, as Chomsky has frequently pointed out (1965). And, if one can always postulate additional phonetic symbols with temporal consequences to the inventory and apply as many rules as you please, then the resulting proliferation of new universal symbols would surely undermine credibility.

Yet, short of proliferating new features with exotic effects, an implementation rule for the voice timing effect in English and German cannot be universal. Most languages in the world (including, e.g., French, Spanish, Arabic, Swahili, for example) do not exploit the relative duration of a vowel to the following stop or fricative constriction as correlates of voicing or anything else (Chen 1970, Port et al.1980). We know from classroom experience that in cases where English stimuli varying in vowel and/or stop closure duration (with unvoiced stop closures) lead native English speakers along a continuum from, e.g., <u>rabid</u> to <u>rapid</u> – the same stimuli with varying V/C ratio will tend not to change voicing category at all for at least Spanish, Chinese and Korean listeners. Their voicing judgments are quite insensitive to V/C ratio. They primarily pay attention to glottal pulsing during the constriction. Such durational manipulations may affect the naturalness of the stimuli, but do not make them sound more Voiced or less Voiced for speakers of most languages outside of English and its Germanic relatives.

The conclusion we draw from this situation is that English and German manipulate V/C ratio for distinguishing classes of words from each other. English listeners, for example, make a

categorical choice between two values of a feature that might be described as `Voicing' (or as `Tenseness' or `Fortis/Lenis'). But there is nothing universal about this property. It just happens to be a way that several closely related languages control speech production and speech perception to distinguish classes of vocabulary items. Thus, we have a temporal pattern which apparently must be a learned property of the phonological grammar of specific languages serving as a `feature' for contrasting sets of words. Could we call this distributed temporal relationship a symbol or a segmental phonetic feature? Sure, but to do so would be to distort this technical concept beyond recognition. The term symbol or feature would then become a very a loose kind of metaphor that only masks what it really is — an intrinsically temporal pattern that is categorically perceived by speakers of certain languages that use this relation like a discrete feature (viz., *Ruby-rupee, bend-bent*, etc.) (Bybee 2001). But it cannot really be a symbol token, that is, a formal object analogous to a bit-string, available for manipulation by a symbol processing machine.

To return to the main argument of this section, such a language-specific, inherently temporal specification for categorical features or phonemes should not be possible according to the formal theory of language. All cross-language differences should be static and segment-sized (that is, discrete in time). And any effects that demand temporal description should be universal. But clearly this is not the case.

4.3. Phonological Nondiscreteness. Thus far, we have presented evidence of the importance of language-specific temporal patterns as intrinsic aspects of linguistic sound categories. Since they are not static, they cannot be symbolic. Another kind of evidence against the hypothesis that phonetics and phonology are symbolic would be a convincing demonstration of patterns that are not discretely different. The question is whether there are patterns that are consistently different but not different enough that they can be reliably differentiated. That is, are there any sound patterns that overlap, not just occasionally, due to noise in the production or perception system, but patterns that consistently or intrinsically overlap with respect to important articulatory features? This may seem a difficult set of criteria to fulfill, but in fact such situations have been demonstrated several times in several languages (see Warner, et al. 2004).

The best studied case is the **incomplete neutralization** of voicing in syllable-final position in Standard German. Syllable-final voiced stops and fricatives, as in *Bund* and *bunt*, `club' and `colorful,' are described by phonologists (Moulton 1962) and phoneticians (Sievers 1901) as neutralizing the voicing contrast in favor of the voiceless case. So although *Bunde* and *bunte* (with suffixes) contrast in the voicing of the apical stop, the pronunciation of *Bund* and *bunt* seems to be the same. Both words sound like [bʊnt]. The difficulty is that they are not pronounced exactly the same (Dinnsen et al. 1971, Port et al. 1986, Port et al.1989, Fourakis et al.1984). These pairs of words actually are slightly different as shown in the schematic recorded waveforms in *Figure 2.* If they were the same, then in a listening task you would expect 50% correct (pure guessing – like English *too* and *two* would show). If contrastive, one would expect at least 99% correct identification under good listening conditions with motivated subjects (just like *Bunde* and *bunte* would show). Instead, the two words are different enough that listeners can guess correctly which word was spoken with only about 60-70% correct performance (Port et al. 1989). This unexpected level of performance shows that the word pairs are neither the same nor distinct. Their discriminability, or *d′* (d-prime) (Green et al.1966), is somewhat less than 1 (Port et al. 1989).

Figure 2. Schematic waveforms of recorded minimal pairs averaged across 5 tokens each for 8 German speakers in Port et al. (1985). The onset of the first vowel (grey rectangle) begins at 0 ms, the wavy rectangle is the period of voicing visible during stop closure, the white box is the voiceless portion of stop closure, and the last rectangle represents the visible stop burst duration at release. These measurements show reduced characteristics relative to normal stop voicing cues: slightly longer vowel durations for the voiced word as well as slightly longer voicing into the closure, somewhat shorter closure durations and weaker bursts. All stops sounded voiceless to the experimenters. These results do not support the notion of a static, binary voicing feature ([±voice]). The timing for the voiced and voiceless word pairs is very similar and the distributions overlap greatly.

The voicing contrast is almost neutralized in this context (close enough that both `sound the same' to us), but when you look closely they turn out to be different. The differences can be measured on sound spectrograms, but for any measurement or combination of measurements one chooses (e.g., vowel duration, stop closure duration, burst intensity, amount of glottal pulsing during the closure, etc.), the two distributions overlap greatly. If an optimal linear combination of these measurements is computed (using discriminant analysis) then the two classes still overlap but can be correctly classified with 60-70% accuracy – about the same as native-speaking listeners do (Port et al. 1989). The Port-Crawford study ruled out the possibility that the different production patterns reflect the use of orthographic presentation (where the distinction is maintained) or that subjects were trying to be cooperative by producing the difference they thought the experimenters were looking for. This unsettling array of facts led Manaster Ramer (1997) to express concern that if the incomplete neutralization phenomena were correct, then it would imply that linguists could not rely on their own or anyone's auditory transcription. We agree entirely with his conclusions. Introspective analyses, such as phonetic transcription, cannot be trusted and are an inadequate basis for linguistic research.

If this difference is not some experimental artifact, one might reasonably ask along with Manaster Ramer, why phoneticians and linguists have failed to note this in their transcriptions of German. The reason is that the goal of phonetic transcription is not typically to record everything,

22

but to record only what is likely to be relevant for a native speaker-hearer (*IPA Handbook*, 1999). The differences shown in Figure 2 are too small to be very useful for communication. Both sets really sound like they end in [t]. Of course, when subjects face a forced-choice identification test of isolated minimal pairs, as Port and Crawford asked of their subjects, we find that such differences can still be exploited for perception yielding better than chance performance. The key fact for the purposes of this essay is that these word pairs lack an essential property of any symbol token (Haugeland, 1985, Port, 1997): they are neither discretely different nor are they the same. So what should their phonetic spelling be, voiced, unvoiced or something else?

A similar problem occurs in American English in the neutralization of pairs like *butting* and *budding*. The voiceless and voiced stops in *butt* and *bud* are typically, at first listen, neutralized to a flap before a vowel-initial suffix (or vowel-initial word: *Say that again*). But for most American speakers, spectrograms of the two words are somewhat different in that *butting* looks and sounds more /t/-like. The spectrogram shows slightly longer closure, slightly shorter preceding vowel, slightly stronger burst and less glottal pulsing during the closure relative to *budding* which is more /d/-like (Fox et al. 1976). And the percent correct identification in a forced choice task gives a score in the 60-75% range (Port unpublished data). In these cases, both the English and German speakers are consistently producing a very small difference in articulatory detail. What they produce lies nondiscretely between two categories. Other replicated examples of incomplete neutralization are word-final voicing neutralization in Russian (Pye 1986), Polish (Slowiaczek et al. 1985) and Dutch (Warner et al. 2004). It is interesting and probably important that all the cases of incomplete neutralization mentioned are context-sensitive. That is, they do not represent a general collapse of a distinction in the lexicon. In certain contexts, a distinction has been largely lost but whatever `process' achieves the neutralization does not completely wipe out the effects of an `underlying spelling' of the lexical items. These cases present troublesome violations of the claim that the phonetics of languages is based on a discrete or digital inventory, and that these discrete phonetic units function like the tokens of a formal system. Instead, speakers can occasionally leave a distinction only partially neutralized by using fine articulatory control. A pattern of overlap might be evidence of a sound change in progress (Pierrehumbert 2001), but in a few cases like final devoicing and flapping of apical stops, the pattern of systematic but minute differences appears to be quite stable over time.

One way to defend the Chomsky-Halle theory despite the phenomena of incomplete neutralization and intrinsically temporal cues is to postulate a far more finely divided phonetic space, one that is as detailed as we need. But unfortunately if the space includes enough detail to differentiate the two flaps of American English or the two final /t/s in German, then the size of the phonetic space must be multiplied by many orders of magnitude. And the majority of these categories will not be reliably distinct from their neighbors (that is, they might get 60-70% correct identification under ideal conditions). If the innate phonetic space has that level of detail, then how could a child's representation of speech in this alphabet solve the problem of rapid language acquisition? It would mean that scarcely detectible differences in the production of words would complicate the child's `transcription'. This degree of detail would mean essentially no phonetic categorization at all.

4.4. Summary of the Argument. The argument we have been presenting can be summarized as follows:

1) The claim that the phonology of a language is a formal system employing phonetic units as its ground level implies an apriori inventory of phonetic atoms that are discretely different

from each other and static as described by Chomsky and Halle, 1968. These tokens play a role analogous to bit strings in a digital computer and serve as the vocabulary for higher-level structures like morphemes. For the same reasons as in a computer, they must be discrete and be reliably produced and recognized if a formal cognitive system is to execute rules successfully.

2) Evidence: Some languages, including English and German, employ patterns of relative duration to distinguish classes of lexical items. These contrasts violate the requirement that all distinctive phonetic elements be definable in static terms. The attempt to treat the production command as a static phonetic symbol which is then implemented with temporal stretching or compressing of the segment cannot account for the data without elaborate and speculative machinations, nor does this approach provide plausible instructions to the motor system.
3) Evidence: Some languages, like English, German, Dutch, Russian and Polish, have consistent phonetic categories (i.e., target articulations) that are different from each other but not discretely so. Even in exactly the same contexts, the distributions of the two units overlap almost completely along any measurable phonetic dimensions and are very imperfectly identifiable when listened to even under ideal conditions by native speakers. These target articulations violate the requirement that all phonological elements be either the same or discretely different.
4) Altogether these phenomena demonstrate that the phonological structures employed by human speakers cannot provide the digital foundation for a formal system of language that is required by generative phonology as well as by all formal theories of language. Indeed, many formal theories of cognition would seem to require formally distinct lexical items (e.g., Fodor 1975, Newell et al. 1976). There are many properties of the sound systems of languages that cannot be described if we are limited to descriptions that comply with the technical constraints of formal symbol systems. In an *SPE*-like model for symbolic phonetics, it is the discreteness of phonetics that guarantees the discreteness of phonology. In our view, language may resemble a formal system in some respects but it cannot, in fact, be a formal system.

If a loose approximation to a formal system is all that is required, for example, if one were designing a practical orthography for a language or trying to facilitate adult language learning, then a simplified formal approximation to phonology (as used by most phonologists) is likely to be quite useful. We do not deny that the phonologies of languages exhibit symbol-like properties, such as reusable and recombinable sound patterns. A small inventory of segment-sized, graphically represented phonological categories can provide a practical scheme for representing most languages on paper. But what is in speakers' heads is apparently not symbols analogous to graphical letters. The term *symboloid* seems like an appropriate term for these cognitive patterns. But for scientific research into human phonological behavior, a schematized formalism that cannot account for the many problems of description discussed here will not be sufficient. Linguists, just as much as psychologists and neuroscientists, urgently need a continuous-time model for language. There is no way to make an alphabet do the job of providing a phonological description of the lexicon of a language. To provide suitable descriptions, we need to take account of the dynamical neurocognitive mechanisms that support such patterns of behavior (Pierrehumbert et al. 1990, van Gelder et al. 1995, Pierrehumbert, 2001). But the model must take responsibility for the properties that can be modeled with formal tools as well as the properties not amenable to formal description.

24

5. Counter-arguments and Rejoinders.

In earlier sections we described several situations that create problems for any theory of phonetics that is formal or symbolic: the problem of temporal features and the problem of nondiscrete grammatical distinctions (i.e., incomplete neutralization). These may not immediately strike linguists as showstopper arguments against the entrenched and comfortable view that language can be adequately represented with formal graphic symbols, but we think that, when seriously considered, these cases present insurmountable difficulties for the classical symbolic view of phonetics and phonology. One linguistic response to our arguments might be:

> (4) `Your data are only about surface facts, but the formal elements that linguistics studies lie deeper than this. They do not need to be audible on the phonetic surface. Thus, e.g., the incomplete neutralization phenomena may show that neutralization does not occur in some contexts where we thought it did, but the correction for this is simply to postulate one or several new underlying distinctions that happen to neutralize incompletely during the performance phase at phonetic output. So there need be no deep theoretical problem here.'

In reply, we point out two things. First, this move pulls the language upstairs and out of sight. In claiming the phonetic tokens are abstract and lack necessary acoustic or articulatory specifications, one relieves the symbolic phonology hypothesis of most testable empirical claims. What had once appeared to be an empirical hypothesis, justified by phenomena like those listed above in (2), is now protected from empirical refutation – as well as from empirical support. The claim that language is formal in this case threatens to become something more like a religious commitment: any incompatible data are dismissed as somehow irrelevant (as invalid tests of surface phenomena) and as revealing a lack of understanding of the nature of language. But this is not a scientifically respectable response. Something more substantive will need to be found to dispute our arguments.

The second problem is that the data we presented above are by no means the only data we might have presented to make the points (a) that phonologies employ distinctive ratio-based timing patterns, and (b) that the phonetic parameters are, for all practical purposes, infinitely divisible. As for ratio-based timing patterns, there is also the Japanese mora, a unit of speech timing (Port et al. 1987, Han 1994). As for other examples of the unconstrained divisibility of speech, we can look at vowels. Chomsky and Halle suggested 4 or 5 binary features for coding vowel types. This is probably a sufficient number of features when looking only at any single language. But Labov has shown that many historical sound changes in vowel pronunciation take place gradually by a seemingly smooth shift of mean target location within a community of speakers (Labov 1963). The vowel targets of various languages and dialects appear to fall in many places in the F1×F2 plane (Maddieson 2002, Disner 1983, Bradlow 1995). Although Ladefoged and Maddieson (1996: 4-6) hope it will be possible to specify some universal set of continuous parameters for vowel description, they do not suggest there is only a fixed set of possible vowels which is what Chomsky and Halle must insist on. Disner (1983) and Maddieson (2002) showed that speakers of two languages with 7-vowel systems placed their vowels in very different locations in the space. Similarly, the IPA Handbook (1999) speaks of the `continuous nature' of the vowel space and offers the cardinal vowel system to provide useful reference points for locating other vowels. So far as we are aware, no one who studies phonetics (as opposed to phonology) has ever suggested that there is a fixed set of vowels across human languages. Despite this, phonologists continue to

behave as though there is a fixed universal set of possible vowel types – as though [High, Back, Round] or [Voice, Obstruent, Tense] mean the same thing in every language. But they do not.

There are many other examples as well. Certainly intonation shows no sign yet that there might be a discrete set of values on any phonetic dimension – whether static tones or contours. Ladefoged has pointed out that even looking at a feature like implosiveness for stops, there is a gradient between languages in the degree of negative oral air pressure in their production (Ladefoged 1968: 6). The general observation here is that anywhere that you look closely at phonetic phenomena, the cross-language identities evaporate. Clinging to the view of a fixed alphabet will require expansion of the alphabet by many orders of magnitude. Almost the only way one could believe in a discrete universal phonetic inventory is if one refuses to look too closely at the data!

A generativist might respond:

> (5) `But none of these observations disprove that discrete features underlie these phenomena. Maybe there are more discrete vowels than we realized. Perhaps we need dozens of VOT values, oral air pressures, intonation contours, etc., rather the few mentioned in the Chomsky and Halle feature set. So what? You still haven't disproven our general claim.'

We agree, of course, that we have not disproven the possibility of a large but finite set of discrete apriori features. The issue is whether the relevant constraint is only that there be a finite number of features. If one postulates a very large and expandable phonetic inventory, then one risks being adhoc since one now permits uncontrolled expansion of the set of premises. But further, the use of the phonetic alphabet to account for children's rapid acquisition of language loses plausibility since, as noted, repetitions of a single word become unrecognizable if the tokens record very minute differences. Recording far too much detail in a transcription for a language learner seems almost as bad as recording too little. And finally, according to *SPE* phonetics, the theory in effect claims for each infant the ability to phonetically identify (not merely discriminate) all sounds in all languages in the world at an age when nothing else about the infant's nervous system would seem to justify confidence in such capabilities.

Altogether then, it is clear that several unavoidable predictions of the symbolic phonology hypothesis have clear counter evidence. Back in the 1960s, it might have been reasonable to hope that phonetics research would gradually converge toward a fixed universal inventory of features, e.g., a limited set of vowel types, for example, that would be combinable into all words in all languages. But it is clear instead that 40 years of phonetics research has provided absolutely no suggestion of convergence on a small universal inventory of phonetic types. Quite the opposite: the more research we do, the more phonetic differences are revealed between languages. So the hypothesis of a universal phonetic inventory should have been abandoned long ago on the basis of phonetic data but phonologists have not been paying attention. Any idea of a universal phonetic alphabet should be completely abandoned as a premise for phonology. Some parts of the word inventory of apparently all languages exhibit regularities suggesting symbol-like discreteness, but that is as far as discreteness goes. Linguistics simply cannot make the convenient assumptions of timelessness and digitality for linguistic phonetic units – or for any other linguistic units either, for that matter. This means that rather than simply assume that language is formal, we need to determine the degree to which it does and does not have the properties of a discrete formal system.

26

6. Consequences for Research Issues in Phonology.

Our primary conclusion thus far is that *there is no discrete universal phonetic inventory and thus phonology is not amenable to formal description*. If a reader were to be persuaded of this, what implications are there for phonological research? It seems that some mainstream research programs in phonology would appear misdirected, but there is also considerable research that is quite compatible with the phonetics implied here. The key change is that discrete phonetic and phonological symbols must be replaced with categories and parameters that are rich in detail. In addition, of course, a vast range of new research topics appear. There are a number of consequences for phonological research.

First, some traditional research strategies seem inappropriate and unlikely to be productive. For example, the strategy of comparing related properties of two or more languages at once in order to choose between alternative analyses of one of them (e.g., interpreting English syllable structure in the light of constraints on, say, Korean syllable structure) seems to be risky since the phonological systems of different languages are generally incommensurable unless the languages are closely related historically. Phonological analysis is most easily done on a single language at a time. Of course, many analytical issues will turn up for which the data from the language in question will be inconclusive. A better approach is to look much closer at the details of the phenomenon of interest by doing laboratory investigation of multiple speakers or statistical speech analysis of appropriate databases. We cannot assume that the Voice feature in English is the same as the Voice feature in any other language. They may exhibit enough similarities that the same graphic symbols are satisfactory for many purposes including academic communication. But, of course, they will still manifest many differences in phonetic detail (e.g. Port et al. 1980, Flege et al. 1986, Local 2003). Some of these differences may be auditorily fairly obvious, such as the spirantization of some voiced stops in Spanish or the aspiration of syllable-initial voiceless stops in English. But others will be subtle, requiring laboratory study of minimal word sets to see the differences (e.g., the V/C duration ratio invariant in English voiced stops, or the absence of vowel lengthening in Arabic before voiced obstruents, etc.). Certainly, there are still some generalizations to be drawn across languages, such as, say, the tendency of [ki] to evolve historically into [či]. But stating these cross-language tendencies should be organized around articulators (since we know for sure they are universal[9]) and organized around specific gestures, rather than in terms of abstract universal phonetic features. We cannot assume that any general description, like [+Voice] → [–Voice] or a constraint such as *[Voice], will have any universal meaning.

Second, the basic distinctions in phonology between a feature and a segment, and between phonemics and phonotactics need to be rethought. Neither distinction can be maintained consistently. Our view is that speakers can record in memory and control in their production far more detail than traditional linguistics supposed. Speakers are not restricted to the use of any particular abstract and `efficient' linguistic description of the units of language. (In fact, it may be primarily the linguist who needs to represent languages with a very small number of abstract symbol tokens, not speaker-hearers.) At the time scale of speech production and speech perception,

---

[9] The speech articulators are as universal as our gross anatomy. But human groups differ in the shape and size of noses, skulls, necks so there might be some differences in at least the economy of particular speech gestures. That is, we should not rule out the possibility that some speech sounds might be easier for some racial groups than for other groups.

it does not matter whether the [t] in *stop* is the `same unit' as the [t^h] in *top,* or whether /tr/ is a consonant cluster or singleton stop. Such phonological identities are relevant if your purpose is to write words down on paper with an efficient set of graphemes. But such issues are quite irrelevant for speaking and hearing words in real-time. All the basic issues of phonological description need to be looked at afresh.[10]

For example, what happens to the issue of `phonotactics' from this point of view? The phonotactics of a language is a description of how segments are distributed in a language. It is the description of the segmental contexts a given segment occurs in. Thus, one might note that English has a syllable-onset cluster like /str-/ (as in *strong, street, strew*) but that Swahili has no such syllables. But this formulation gives priority to the segmental, letter-like description. There are constraints on the speech patterns in each language, but there are much better ways to describe the patterns than by use of letter-like, serially ordered tokens. Every language has its own constraints on the space of probable speech producing gestures. These can be described with probability distributions of gestures (Pierrehumbert 2001, Bod et al., 2003).

Third, another change demanded by the new view is that we must take seriously all of what used to be called `external evidence.' It is time to be serious about experimental psychology and to incorporate experimental research techniques into phonology. These could employ behavioral measures such as those in (6) plus others.

(6) Experimental Evidence Regarding a Tentative Phonological Analysis

a. probability of one response or another in an identification task of experimentally manipulated speech samples

b. response time for cognitive phonological tasks (e.g., monitoring for a sound type, word confirmation, choice of a preferred pronunciation, judgment of naturalness, etc.)

c. accuracy of response to a binary listening task (e.g., same vs. different, heard previously or not, which member of a minimal pair, etc.)

d. measurements of time or spectrum from audio or video recordings of speech production (e.g., word and gesture durations, vowel formants, voice-onset time, interstress intervals, etc.)

Forms of evidence like these help reveal aspects of the cognitive representation of words.

Of course, experimental phonologists, phoneticians and psychologists have been doing such research for many years. High-quality research in this tradition is recorded in the biennial series *Laboratory Phonology,* by Cambridge University Press (1992-2003) and appears in *Journal of Phonetics* and *Journal of the Acoustical Society of America* as well as other journals covering language development and experimental psychology. Acoustical analysis and physiological measures of articulatory gestures can (1) reduce uncertainties about a phonological analysis, (2) contribute to understanding variations across speakers and time, (3) discover patterns that are not easily perceivable in conscious auditory terms (e.g., spectral and temporal details, durational ratios, and much more yet to be discovered, see Hawkins, 2003) and (4) can suggest models of the linguistic control of speech production (cf. Browman et al. 1992, Saltzman 1995, Kelso 1995).

---

10 We agree, then with Faber (1992) that our lifelong experience using letters as representations of speech sounds and words has biased our intuitions strongly toward segmental description of language.

28

The fourth implication of this new approach to phonology is that, since continuous time is fully incorporated into the theory of language, phonologists can now join the search for explicitly temporal patterns characteristic of various languages. By taking gestures as the basic units rather than states (Browman et al. 1992) we are no longer restricted just to `rhythms' defined in terms of the serial order of symbol types (e.g. Hayes 1995). Several kinds of temporal patterns have been discussed: V/C ratio, Japanese mora timing plus the periodic timing of stressed syllable onsets in certain chant-like styles of speech (Cummins et al. 1998, Tajima et al. 2003, Cummins 2003, Port 2003). Periodic timing is also exploited for artistic and social purposes (e.g., religious chant, street-caller chants, song, etc.). It seems likely that many other kinds of temporally defined structures will turn up in languages of the world as soon as linguists begin to look for them.

Finally, the most general implication is to encourage reconsideration of whether any area of linguistics should identify itself with the assumption that language is a pure symbol system – a system that is definable in discretely contrastive terms at every level of cognitive structure. There is now a great deal of evidence from the psychological literature that human cognition employs categories that are definable in many ways. Psycholinguistic research reveals that most **cognitive categories** have little resemblance to the kind of Aristotelian categories linguists are committed to (Lakoff et al. 1999, Kruschke 1992, Goldstone et al. 2003, Johnson 1997, Pierrehumbert 2001, 2003, Hawkins 2003).

We linguists have tended to assume that any linguistic categories are abstracted away from concrete details and specified by a minimal number of discrete degrees of freedom (like distinctive features) (see Bybee 2001). But much evidence now shows that memory for auditorily presented words also includes minute phonetic details such as information about a particular speaker's voice. We can even remember the collocation of specific words spoken by a specific voice. The evidence also suggests that this information tends to remain in storage for days and weeks (Goldinger 1998, Pisoni 1997, Hawkins 2003). Data like these suggest that phonological categories (or symboloids), like syllable-types, segments and segmental features, may have both an abstract representation and also something like a cloud of specific examples or episodes of concretely specified events. The episodes are sequences of events in time. But one can simplify intuitions by conceptualizing them as points in a very high dimensional space (Hintzman 1986, Nosofsky 1986). The kind of linguistic category that might be involved may have a fairly low-dimensional description (along the lines of an allophonic phonetic description) but also be accompanied by a large set of episodes that are specified by a vast number of descriptive parameters including many temporal and speaker-dependent properties. The members of the equivalence class are bound together by some (probably learned) measure of similarity (Nosofsky 1986, Goldstone 1994, Werker et al. 1984).

Because of the amount of detail that is stored for each of the exemplars, the system adapts (that is, learns) a little from each presentation of a member of the category (Goldinger 1998). Words and other units can be primed for activation to varying degrees. Behavior is typically sensitive to the frequency of occurrence of the categories (Phillips 1984, Bybee 2001, Pierrehumbert 2002). Some (and maybe all) linguistic categories, like words, morphemes and sound types, appear to have properties like these (see Bybee 2001, Pisoni 1997, Pierrehumbert 2001, 2003). Categories that are phonetically rich and temporally detailed are quite compatible with the view of phonetics that is suggested by the research discussed above.

If linguistic categories do not have the properties of discrete abstract symbols as we had thought, then our phonological research target should be `Figure out what enables speakers to talk and understand each other using speech.' This is intended as a minimally biased way to seek understanding of the form of language in long-term memory – the form in which words are

available for a speaker and hearer to call up for production or perception at the moment of speaking. The great hypothesis of 20th century structural linguistics, starting with de Saussure and the Prague School, was that the speaker's solution to the problem of remembering words would have an unmistakable resemblance to the written language with hierarchical data structures resembling those of various orthographic units: segments, words and sentences. But for that to be true, there must be a basic-level alphabet of crisp letter-like tokens, suitable for discrete combination in building larger structures. Since there apparently is no basic symbolic alphabet for cognition (at least not a phonetic one), despite the obvious existence of higher (i.e., temporally longer, more abstract) structures of the phonology and lexicon, we must keep our minds open and employ whatever models work best to explain relevant phenomena. The evidence reviewed in this essay suggests that whatever code is used for remembering words and phrases is probably quite different from language to language. Human memory for words is phonological in the sense that the code is unique for each language. But memory for language is apparently also rich in phonetic details. Memory also includes some information that is not what we think of as linguistic: the speaker's voice, the speaking rate, temporal details, emotional setting and so on. The phonological categories studied by linguists, such as phonemes, syllable onsets, features, etc., are only part of the story.

It is not clear just how linguistic categories will be defined or how many kinds of them there are. But linguistics cannot afford to ignore what is being learned in experimental psychology and cognitive science about the kinds of generalizations and cognitive models humans construct regarding language. Humans learn to create linguistic categories of many kinds – sound units, motor patterns, perceptual categories, concepts, etc. Most of them are not very close to the discrete, sharply-defined notion of a `formal symbol.' The main exception, of course, is that people educated in literate societies also have a large set of alphabetic and other orthographic concepts tied to specific graphical patterns: letters, words as graphic patterns, etc. So obviously alphabetic (or other) orthographies provide many symbol-like tokens to influence the intuitions of literates.

We suggest using mathematics quite differently than it was used in generative phonology. Rather than assume that linguistic cognition obeys the postulates of idealized mathematical systems, we should use mathematics to develop explicit models of cognitive activity but without assuming discrete time. Generative phonology seems to have overlooked that, in order for there to be a formal system, there must be something or someone to execute the rules, whether a computer, a linguist or the subconscious brain. Unfortunately there is no evidence whatever that human brains automatically and unconsciously implement any formal system at all. And the fact that linguists can do so at a conscious level (leaning on paper and pencil as memory aids) does not offer much support to the proposal for unconscious formal implementation.

In a linguistics committed to the physical world (rather than to some Platonic heaven), language needs to be naturalized so as to fit it into a human body. That implies, first of all, casting it into the realm of space and time. It requires changing our focus of attention from our preconceived views of the form of linguistic knowledge toward the study of linguistic behavior and performance. We should study behavior simply because speech and language take place in time, so linguistic `knowledge,' whatever it is like, must be dynamic as well. Temporal information is sure to be needed to discover how the whole system really works. If we want to speak of `linguistic knowledge' then we should do so in a way that includes both static knowledge and the processor, both steady-states and dynamics and perhaps the abstraction as well as the specific episodes. If the cognitive system for language is something designed to run in time, then it will only be understood in such terms. The attempt to separate the static aspect from change-of-state is Procrustean and

30

does violence to understanding of the entire system. Language will never be understood by insisting on the distinction between Competence and Performance

What is universal about phonology is not any fixed list of sound types, but rather the strong tendency of human language learners to discover or create sound categories out of what they hear. Human infants seek categories of sound types in the speech around them – even before they learn their first word. Children will discover patterns in various size ranges. And different members of a speech community may easily learn approximately the same patterns. This process, continued over time, yields a vocabulary suggesting the tables of minimally contrastive words that are often found – the ones that are taken to be evidence for discrete features. Because the learning and the analysis take place at the level of individual speakers, there is little to prevent small interspeaker differences and gradual changes within speakers over time. Thus, gradual changes may occur in the mean phonological structure of the community of speakers. The sound system of each language does exhibit some discrete features but there is also much that is not discrete at any point in time and much that is not static.

7. Progress in Modeling.

Our talk of memory for articulatory and acoustic detail along a huge number of dimensions may seem to place unreasonable demands on a speaker's perceptual system and memory. But humans are able to learn from many detailed features of the environment. Present-day modeling of neural processes suggests ways in which this learning can be done. There are now neural network models of perception that seem to exhibit the kind of behavior that is required: systems that learn to direct their attention to the input dimensions that are most informative and which can then learn to categorize inputs discretely (e.g. Kruschke 1992, Grossberg 2003). Current models can exhibit categorical perception and the `perceptual magnet' effect (e.g., Guenther et al. 1996). Other model systems can parse word-like categorical units from overlapping patterns presented in continuous time (Grossberg et al. 2000, Grossberg 2003) and can model recognition of vowels without a normalization process (Johnson 1997). There are also model systems that produce speech-like articulatory patterns at a range of speaking rates and degrade in ways that human speakers degrade using segmental and feature-like memory representations of words (e.g. Saltzman 1995, Browman et al. 1992). And there is a rich tradition of research and modeling of word memory (Hintzman 1986, Kruschke 1992, Shiffrin et al 1997, Goldinger 1998). These and other models are typically implemented as computer programs to demonstrate their principles of operation. They learn to categorize inputs presented in continuous time and some produce continuous output (see Grossberg 2003). The design features of the network models permit them to learn, for example, whatever recurring frequency-by-time patterns there may be in a rich and detailed speech environment and to exhibit temporal behavior resembling human performance.

Of course, all these demonstrations are still primitive and incapable of anything like learning the phonology of a real language. It remains to be seen whether these models will scale up to effectively model full-scale human performance. But at the very least the rationalist framework for linguistics, one that has symbols from bottom to top, is no longer the only theoretical approach imaginable. Linguists can contribute to understanding human linguistic skills by describing, with as little bias as possible, the phonological patterns to be found in languages of the world. The descriptions should be supported both by traditional distributional data of phonological research and by experimental results that clarify the category structure at issue. There is much to learn about phonological systems: about the physical and neural equipment that supports them as well as how they are shaped through time both in children and in language-learning adults.

Obviously, what is being endorsed here is a program to bring linguistics into the domain of conventional cognitive science. The assumption that language was somehow unique among animal skills and could be understood only by applying a large dose of formalist modeling may have had some plausibility in the 1960s but is difficult to justify today. There is simply no reason to continue the formalist effort to the exclusion of a broader range of approaches to understanding linguistic cognition.

8. Conclusions.

We began this discussion of the phonetic and phonological building blocks of human speech by contrasting the Chomsky-Halle tradition with that of modern experimental phoneticians. Both groups of scientists gave too much priority to segmental descriptions of speech, but phonologists went further in presuming there is a fixed inventory of speech sound types that are available for all languages. This fixed inventory purportedly provides an invariant vocabulary of speech sounds for comparison between all languages. But we have seen that phonetics can not provide such a vocabulary, and thus that cross-linguistic phonological comparisons have no place to stand, no standard basis for comparison. The actual comparisons are currently based on transcriptions by linguists and phoneticians. These segmental transcriptions are generally biased toward segmental description and lack any account of temporal patterns. The assumption that language is a formal symbolic system seems so obvious to many linguists as to scarcely require any justification. And it does have a little truth to it. But if letter-sized units are taken as a foundational premise about linguistic cognition, then we are misled to interpret all the aspects of language that are *not* symbolic as illustrating a need for increasingly arcane symbolic description.[11]

One form of evidence against the assumption of static building blocks came from studies of speech timing showing that some phonologically significant patterns are describable only in temporal terms (such as durational ratios in English voicing and Japanese moras). So-called temporal implementation rules cannot provide a reasonable account for them. These cases violate the premise that phonetic features are symbols representing articulatory states. A second form of evidence is that some phonological features are not discretely different from each other even in ideal speaking and listening conditions. It is impossible for native speakers to be certain which sound category they hear. Such situations are incompatible with the premise that phonology or phonetics constitute formal symbol systems. These cases imply that, at both the `phonetic' and `phonological' level, discrete phonetic atoms may sometimes exist but do not always exist.

Generative phonology, like any symbolic phonology theory, is based on the idea that linguistic structures are made by assembling small letter-like atoms into larger structures. If you are building a house, you do need a pile of bricks apriori. But the many kinds of structures in human cognition need to be explained very differently – in terms of self-organized components that run in time, not apriori static ones. These structures can be learned, but they will not necessarily be discrete in the conventional way. A universal phonetic alphabet once seemed inevitable as an account of phonological discreteness, but it is important to explore other ways to think about how sound structures come into being. Many ways are now understood for constructing stable systems using continuous parameters. Dynamical systems with many degrees of freedom can coordinate all

11 We note that our most fundamental criticism – that language is not a formal system – seems to be the same as that expressed by Hockett (1968) who concluded that `language is not well-defined'.

32

these variables and exhibit completely discrete states (see Port et al. 1995, Kelso 1995, Thelen et al. 1994, Clark 1997, Large et al. 1999, Guenther et al. 1996, Grossberg 2003). A simple example is that a plucked guitar string will oscillate, not just at one frequency, but at several discretely different frequencies (see Port 2003 for discrete timing based on harmonic ratios).

Today the premise that language must be completely formal is impeding progress in phonology, and probably in other parts of linguistics as well. We have tried to show that it presents linguists with a serious problem because the formal theory of language is not permitted to have realtime characteristics. We have shown here that rich behavioral details are essential to describe linguistic behavior – in word recognition processes, in the gestures of speech articulation, for speech memory and so forth. One must conclude that whatever symboloid structures of language there might be are not the only representation exploited in linguistic behavior. Linguistics cannot stand by and deny the relevance of continuous time if it is to seriously address aspects of human cognition.

There is only one route left to justify doing traditional generative phonology or for studying only the abstract sound structures of a language and to deny the relevance of articulatory, acoustic and auditory details. It is to claim `We don't care about linguistic behavior, only about linguistic knowledge.' But there is no assurance that a coherent static description of knowledge exists just because that is what one wants to study. There is a risk that, for methodological purposes, this mission may be implemented as `We care about how to write down a description of a language.' But this is a very questionable goal because it reflects, at least partly, our very high level of comfort with alphabetic descriptions of language. If it is linguistic behavior that we want to account for, then we must let go of the requirement that we also be able to write our linguistic descriptions down.

References

1999. Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet. Cambridge, England: Cambridge University Press.

Bannert, Robert. 1974. Temporal organization and perception of vowel-consonant sequences in Central Bavarian. Working Papers, Department of Linguistics, Lund, Sweden, 12.47-59.

Bod, Rens, Hay, Jennifer and Jannedy, Stefanie. 2003. Probabilistic Linguistics. Cambridge, Massachusetts: MIT Press.

Boersma, Paul. 1998. Functional Phonology: Formalizing the Interaction between Articulatory and Perceptual Drives. The Hague, Holland: Holland Academic Graphics.

Bradlow, Ann. 1995. A comparative acoustic study of English and Spanish vowels. Journal of the Acoustical Society of America 97.1916-24.

Browman, Catherine and Goldstein, Louis. 1992. Articulatory phonology: An overview. Phonetica 49.155-80.

Bybee, Joan. 2001. Phonology and Language Use. Cambridge, UK: Cambridge University Press.

Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. Phonetica 22.129-59.

Cho, Taehun and Ladefoged, Peter. 1999. Variations and universals in VOT: Evidence from 18 languages. Journal of Phonetics 27.207-29.

Chomsky, Noam. 1964. Current Issues in linguistic theory. The Structure of Language: Readings in the Philosophy of Language, ed. by Jerry Fodor and Jerrold Katz, 50-118. Englewood Cliffs, New Jersey: Prentice-Hall.

Chomsky, Noam. 1965. Aspects of the Theory of Syntax. Cambridge, Massachusetts: MIT Press.

Chomsky, Noam. 2000. Linguistics and brain science. Image, Language, Brain: Papers from the First Mind Articulation Project Symposium, 13-28: MIT Press.

Chomsky, Noam and Halle, Morris. 1968. The Sound Pattern of English. New York: Harper and Row.

Clark, Andy. 1997. Being There: Putting Brain, Body, and World Together Again. Cambridge, Mass.: Bradford Books/MIT Press.

Clark, Andy. 2004. Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence. Oxford, England: Oxford University Press.
Clark, Michael and Hillenbrand, James. 2003. Quality of American English front vowels before /r/. Journal of the International Phonetic Association 33.1-16.
Cummins, Fred. 2003. Rhythmic grouping in word lists: Competing roles of syllables, words and stress feet. Paper presented at Proceedings of the 15th International Conference on Spoken Language Processing, Barcelona, Spain.
Cummins, Fred and Port, Robert. 1998. Rhythmic constraints on stress timing in English. Journal of Phonetics 26.145-71.
Dinnsen, Dan and Garcia-Zamor, Maria. 1971. Three degrees of vowel length in German. Journal of Linguistics 4.111-26.
Disner, Sandra. 1983. Vowel Quality: The Relation between Universal and Language-specific Factors. UCLA Working Papers in Phonetics 58.
Eimas, Peter D., Sequeland, E. R, Juszyk, Peter W. and Vigorito, J. 1971. Speech perception in infants. Science 171.303-06.
Eisen, Barbara and Tillman, Hans Guenther 1992. Consistency of judgments in manual labeling of phonetic segments: The distinction between clear and unclear cases. Paper presented at Proceedings of the International Conference on Spoken Language Processing 1992, Banff, Alberta, Canada.
Elert, Claes-Christian. 1964. Phonological Studies of Quantity in Swedish. Stockholm, Sweden: Almqvist & Wiksell.
Fetzer, James H. 2001. Computers and Cognition: Why Minds are not Machines: Studies in Cognitive Systems. Dordrecht, Netherlands: Kluwer Academic Publishers.
Fischer-Jorgenson, Eli. 1952. On the definition of phoneme categories on a distributional basis. Acta Linguistica 7.8-39.
Flege, James and Hillenbrand, James. 1986. Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. Journal of the Acoustical Society of America 79.508-17.
Flege, James, Munro, M. J. and Skelton, L. 1992. Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin and Spanish. Journal of the Acoustical Society of America 92. 128-43.
Flege, James. 1993. Production and perception of a novel second language phonetic contrast. Journal of the Acoustical Society of America 93. 1598-1608.
Flemming, Edward. 1995. Auditory Representations in Phonology: Doctoral Dissertation: University of California, Los Angeles.
Fodor, Jerrold. 1975. The Language of Thought: Harvard University Press.
Fodor, Jerry and Pylyshyn, Zenon. 1988. Connectionism and cognitive architecture. Cognition 28.3-71.
Fourakis, Marios and Iverson., Gregory. 1984. On the `incomplete neutralization' of German final obstruents. Phonetica 41.140-49.
Fowler, Carol A., Rubin, Phillip, Remez, Robert and Turvey, Michael T. 1981. Implications for speech production of a general theory of action. Language Production, ed. by B. Butterworth 373-420. New York: Academic Press.
Fox, Robert and Terbeek, Dale. 1977. Dental flaps, vowel duration and rule ordering in American English. Journal of Phonetics 5.27-34.
Goldinger, Steven D. 1998. Echoes of echoes? An episodic theory of lexical access. Psychological Review 105.251-79.
Goldstone, Robert L. 1994. The role of similarity in categorization: Providing a groundwork. Cognition 52.125-57.
Goldstone, Robert L. and Kersten, Alan. 2003. Concepts and categorization. Comprehensive Handbook of Psychology, ed. by Alice Healy and Robert Proctor, 599-621. New Jersey: Wiley.
Green, David M. and Swets, John A. 1966. Signal Detection Theory and Psychophysics. New York: Wiley Publishers.
Grossberg, Steven. 1999. How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. Spatial Vision 12.163-86.
Grossberg, Steven. 2003. The resonant dynamics of speech perception. Journal of Phonetics 31.423-45.
Grossberg, Steven and Myers, C.W. 2000. The resonant dynamics of speech perception: Inter-word integration and duration-dependent backward effects. Psychological Review 107.735-76.
Guenther, Frank H. and Gjaja, Marin. 1996. The perceptual magnet effect as an emergent property of neural map formation. Journal of the Acoustical Society of America 100.1111-21.
Halle, Morris. 1985. What speakers know about words. Essays in Honor of Peter Ladefoged, ed. by Victoria Fromkin: MIT Press.
Halle, Morris and Stevens, Kenneth N. 1980. A note on laryngeal features. Quarterly Progress Report, Research Laboratory of Electronics 101.198-213.

34

Han, Mieko. 1994. Acoustic manifestations of mora timing in Japanese. Journal of the Acoustical Society of America 96. 73-82.
Haugeland, John. 1985. Artificial Intelligence: The Very Idea. Cambridge, Mass: Bradford Books-MIT Press.
Hawkins, Sarah. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. Journal of Phonetics 31.373-405.
Hayes, Bruce. 1995. Metrical Stress Theory: Principles and Case Studies. Chicago, Illinois: University of Chicago Press.
Hintzman, Douglas L. 1986. `Schema abstraction' in a multiple-trace memory model. Psychological Review 93.411-28.
Hockett, Charles. 1968. The State of the Art. The Hague: Mouton.
Jakobson, Roman, Fant, Gunnar and Halle, Morris. 1952. Preliminaries to Speech Analysis: The Distinctive Features. Cambridge, Massachusetts: MIT.
Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. Talker Variability in Speech Processing, ed. by Keith Johnson and John Mullenix, 145-65. San Diego, California: Academic Press.
Joos, Martin. 1948. Acoustic Phonetics. Language Monographs 23 (Supplement 24)
Keating, Patricia. 1984. Phonetic and phonological representation of stop consonant voicing. Language 60.286-319.
Kelso, J. A. Scott. 1995. Dynamic Patterns: The Self-Organization of Brain and Behavior. Cambridge, Massachusetts: MIT Press.
Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America 59.1208-21.
Kruschke, John. 1992. ALCOVE: An exemplar-based connectionist model of category learning. Psychological Review 99.22-44.
Kuhl, Patricia and Iverson, Paul. 1995. Linguistic experience and the ``perceptual magnet effect''. Speech Perception and Linguistic Experience: Issues in Cross-Language Research., ed. by Winifred Strange, 121-54. Timonium, Maryland: York Press.
Labov, William. 1963. The social motivation of a sound change. Word 19.273-309.
Ladefoged, Peter. 1968. A Phonetic Study of West African Languages: An Auditory-Instrumental Survey. London: Cambridge University Press.
Ladefoged, Peter. 1980. What are speech sounds made of? Language 56.485-502.
Ladefoged, Peter and Maddieson, Ian. 1996. Sounds of the World's Languages. Oxford, U.K.: Blackwell.
Lakoff, George and Johnson, Mark. 1999. Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought: Basic Books.
Large, Edward W. and Jones, Mari R. 1999. The dynamics of attending: How we track time-varying events. Psychological Review 106.119-59.
Lehiste, Ilse. 1970. Suprasegmentals. Cambridge, Massachusetts: MIT Press.
Liberman, Alvin M., Delattre, Pierre, Gerstman, Louis and Cooper, Frank. 1956. Tempo of frequency change as a cure for distinguishing classes of speech sounds. J. Experimental Psychology 52.127-37.
Liberman, Alvin M., Delattre, Pierre, Gerstman, Louis and Cooper, Frank. 1968. Perception of the speech code. Psychological Review 74.431-61.
Liberman, Mark Y. 1975. The Intonational System of English. Department of Linguistics: Indiana University Linguistics Club.
Lieberman, Phillip. 1965. On the acoustic basis of perception of intonation by linguists. Word 21.40-54.
Lindblom, Bjorn. 1990 On the notion of ``possible speech sound.'' Journal of Phonetics 18.135-152.
Lisker, Leigh. 1984. `Voicing' in English: A catalogue of acoustic features signalling /b/ vs. /p/ in trochees. Language and Speech 29.3-11.
Lisker, Leigh and Abramson, Arthur. 1971. Distinctive features and laryngeal control. Language 44.767-85.
Lisker, Leigh and Abramson, Arthur. 1964. A cross-language study of voicing in initial stops: acoustical measurements. Word 20.384-422.
Lisker, Leigh and Abramson, Arthur. 1967. Some effects of context on voice-onset time in English stops. Language and Speech 10.1-28.
Local, John. 2003. Variable domains and variable relevance: Interpreting phonetic exponents. Journal of Phonetics 31.321-39.
Logan, John, Lively, Scott E. and Pisoni, David B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. Journal of the Acoustical Society of America 89.874-86.
Maddieson, Ian. 2003. Vowel spacing in four-vowel systems. Journal of Acoustical Society of America 113.2331.
Manaster Ramer, Alexis. 1996. A letter from an incompletely neutral phonologist. Journal of Phonetics 24.477-89.
McCarthy, John. 2002. A Thematic Guide to Optimality Theory. Cambridge, England: Cambridge University Press.

Moulton, William. 1962. The Sounds of English and German. Chicago: University of Chicago Press.
Newell, Allen and Simon, Herbert. 1976. Computer science as empirical inquiry: Symbols and search. Communications of the Association for Computing Machinery 19.113-26.
Nosofsky, Robert. 1986. Attention, similarity and the identification-categorization relationship. Journal of Experimental Psychology: General 115.39-57.
Peterson, Gordon E. and Lehiste, Ilse. 1960. Duration of syllable nuclei in English. Journal of the Acoustical Society of America 32.693-703.
Phillips, Betty. 1984. Word frequency and the actuation of sound change. Language 60.320-42.
Pierrehumbert, Janet 2000a. What people know about sounds of language. Studies in Linguistic Sciences 29.
Pierrehumbert, Janet. 2000b. The phonetic grounding of phonology. Les Cahiers de l'ICP, Bulletin de la Communication Parlee 5.7-23.
Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. Frequency Effects and the Emergency of Linguistic Structure, ed. by Joan Bybee and Paul Hopper, 137-57. Amsterdam: John Benjamins.
Pierrehumbert, Janet. 2003. Probabilistic phonology: Discrimination and robustness. Probability Theory in Linguistics, ed. by R. Bod, J. Hay and S. Jannedy. Cambridge, Mass.: MIT Press.
Pierrehumbert, Janet. and Pierrehumbert, R. 1990. On attributing grammars to dynamical systems. Journal of Phonetics 18.465-477.
Pind, J. 1995. Speaking rate, VOT and quantity: The search for higher-order invariants for two Icelandic speech cues. Perception & Psychophysics 57.291-304.
Pisoni, David B. 1997. Some thoughts on `normalization' in speech perception. Talker variability in speech processing, ed. by Keith Johnson and J. Mullennix, 9-32. San Diego: Academic Press.
Port, Robert. 1981a. On the structure of the phonetic space with special reference to speech timing. Lingua 55.181-219.
Port, Robert. 1981b. Linguistic timing factors in combination. Journal of the Acoustical Society of America 69.262-74.
Port, Robert. 1990. Representation and recognition of temporal patterns. Connection Science 2.151-76.
Port, Robert. 1997. The discreteness of phonetic elements and formal linguistics: response to A. Manaster Ramer. Journal of Phonetics 24.491-511.
Port, Robert 2002. Phonetics and motor activity. The Complete Linguist: A Collection of Papers in Honor of Alexis Manaster-Ramer, ed. by Fabrice Cavoto, 329-44. Munich: Lincom Europa.
Port, Robert. 2003. Meter and speech. Journal of Phonetics 31.599-611.
Port, Robert and Rotunno, Rosemarie. 1979. Relation between voice-onset time and vowel duration. Journal of the Acoustical Society of America 66.654-62.
Port, Robert and Dalby, Jonathan. 1982. C/V ratio as a cue for voicing in English. Perception & Psychophysics 2.141-52.
Port, Robert and Mitleb, Fares M. 1983. Segmental features and implementation in acquisition of English by Arabic speakers. Journal of Phonetics 11.219-29.
Port, Robert and O'Dell, Michael. 1986. Neutralization of syllable-final voicing in German. Journal of Phonetics 13.455-71.
Port, Robert, Cummins, Fred and Mcauley, Devin. 1995. Naive time, temporal patterns and human audition. Mind as Motion: Explorations in the Dynamics of Cognition, ed. by Robert Port and Timothy van Gelder, 339-71. Cambridge, Mass.: Bradford Books/MIT Press.
Port, Robert and Crawford, Penny. 1989. Pragmatic effects on neutralization rules. Journal of Phonetics 16.257-82.
Port, Robert F., Al-Ani, Salman and Maeda., Shosaku. 1980. Temporal compensation and universal phonetics. Phonetica 37.235-52.
Pye, S. 1986. Word-final devoicing of obstruents in Russian. Cambridge Papers in Phonetics and Experimental Linguistics 5.1-10.
Saltzman, Elliot. 1995. Dynamics and coordinate systems in skilled sensorimotor activity. Mind as Motion: Explorations in the Dynamics of Cognition, ed. by Robert Port and Tim van Gelder, 149-73. Cambridge, Mass: MIT Press.
Sampson, Geoffrey. 1977. Is there a universal phonetic alphabet? Language 50.236-59.
Scheutz, Matthias. 1999. When physical systems realize functions. Minds and Machines 9.161-96.
Shiffrin, Richard and Steyvers, Mark. 1997. The effectiveness of retrieval from memory. Rational Models of Cognition, ed. by Michael Oaksford and Nicholas Chater, 73-95. Oxford, United Kingdom: Oxford University Press.
Shriberg, Lawrence D. and Lof, Gregory L. 1991. Reliability studies in broad and narrow phonetic transcription. Clinical Linguistics and Phonetics 5.225-79.
Sievers, Edouard. 1901. Grundzüge der Lautphysiologie. Leipzig.

36

Sigurd, Bengt. 1965. Phonotactic Structures in Swedish. Lund: Berlingska Boktryckeriet.
Slowiaczek, Louisa and Dinnsen, Daniel A. 1985. On the neutralizing status of Polish word-final devoicing. Journal of Phonetics 13.325-41.
Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. Papers in Laboratory Phonology 5, ed. by Michael Broe and Janet Pierrehumbert, 313-35. Cambridge, UK: Cambridge University Press.
Stevens, Kenneth N. 1989. On the quantal nature of speech. Journal of Phonetics 17.3-46.
Strange, Winifred. 1995a. Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research. Timonium, MD: York Press.
Strange, Winifred. 1995b. Cross-language studies of speech perception: A historical review. Speech Perception and Linguistics Experience: Issues in Cross-Language Research., ed. by Winifred Strange. Timonium, Maryland: York Press.
Tajima, Keeichi and Port, Robert. 2003. Speech rhythm in English and Japanese. Papers in Laboratory Phonology VI, ed. by John Local, Richard Ogden and Rosalind Temple, 317-334. Cambridge: Cambridge University Press.
Tajima, Keeichi, Port, Robert and Dalby Jonathan. 1997. Effects of temporal correction on intelligibility of foreign-accented English. Journal of Phonetics 25.1-24.
Thelen, Esther and Smith, Linda. 1994. Dynamics of Cognitive Development. Cambridge, Mass: MIT Press.
van Gelder, Tim and Port, Robert. 1995. Its about time. Mind as Motion: Explorations in the Dynamics of Cognition, ed. by Robert Port and Timothy van Gelder, 1-44. Cambridge, Mass.: MIT Press.
Van Santen, Jan P.H. 1996. Segmental duration and speech timing. Computing prosody: Computational Models for Processing Spontaneous Speech, ed. by Yoshinori Sagisaka, Nick Campbell and Norio Higuchi, 225-49. New York, New York: Springer-Verlag.
Warner, Natasha, Jongman, Allard, Sereno, Joan and Kemps, R. Rachel. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. Journal of Phonetics 32.251-76.
Werker, Janet and Tees, Richard C. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 7.49-63.
Zue, Victor W. and Laferriere, Martha. 1979. Acoustic study of medial /t,d/ in American English. Journal of the Acoustical Society of America 66.1039-50.