COGNITIVE PHONETICS

Mark Tatham

Reproduced from: Tatham, Mark (1990) Cognitive Phonetics. In Ainsworth, W.A. (ed.) *Advances in Speech, Hearing and Language Processing,* Vol. 1 pp. 193-218. London: JAI Press

OUTLINE

INTRODUCTION: THE NEED TO EXTEND PHONETIC THEORY

There is a serious problem facing researchers who wish to relate phonological and phonetic theory. Not only is it the case that the aims and objectives of the two theories are different, but in addition the investigatory material, methods, and underlying *metatheories* differ significantly. On the one hand we have phonology which fits within the general meta theoretical framework of linguistics as a whole, but on the other hand, as explained later, current phonetic theory does not square with current linguistics. Because of this major difference direct interface is therefore precluded. The difference can be simply characterised by stating that linguistics, including phonology, is almost entirely mentalistic or cognitive in conception, whereas phonetics is not.

Despite the necessarily narrow perspective adopted by linguistics, the scope of a comprehensive theory of language remains enormous, ranging from philosophy, through cognitive science to experimental physics. Chomsky (1957) was of the view that ultimately a model characterising language and speech production would consist of descriptions of how thought could be mapped to sound. If so, the derivation process itself must take in the most abstract and the most concrete of observable phenomena. This has opened the possibility and perhaps the necessity for many viewpoints from which to model language.

Chomsky of course had a good point in wanting initially to restrict the domain of linguistics so tightly, since the ultimate model could not yet be formulated. In 1957 the subject needed a fresh start, and he accomplished just that. But times have changed, linguistics is unrecognisably more sophisticated than it was before Chomsky came along, and many of the needs for a model of language have changed.

Since the 1950s linguists have for the most part been treating language as a cognitive phenomenon. It is not difficult therefore to understand why phonetics-a study connected with acoustics, aerodynamics, motor control, and so on-would be thought of as being non-cognitive by both linguists and phoneticians alike. At most a phonetic model might have an input derived somehow from a cognitively oriented phonology.

An equal contributory factor to the treating of linguistics and phonetics differently has been the rapid development of the technology available for examining phonetic phenomena in the laboratory, together with the widespread acceptance of modern phonetics as a discipline more suited to the approach of the hard sciences because of those experimental possibilities. It is not possible for cognitively oriented linguistics to be carried out experimentally in any kind of laboratory other than the psychology laboratory because the object under investigation is abstract. The phenomena we measure in the phonetics laboratory are not abstract in this way.

Many linguists, especially phonologists, have felt nevertheless uneasy and tried to account for the detail revealed in the phonetics laboratory in the latter stages of their cognitive models. After all, not all of this detail is simply the result of low level involuntary co articulatory effects. For example, voiced obstruents are observed in English to lack vocal cord vibration in final position, whereas in French the vocal cord vibration continues throughout - surely, then, the presence or absence of vocal cord vibration is the result of a decision, and should therefore be treated in the component (phonology) set aside for dealing with the cognitive aspects of speech production. So in phonologies of English we regularly find a rule like the following:

$$\begin{vmatrix} \text{C} \\ +\text{voice} \end{vmatrix} \rightarrow \begin{vmatrix} \\ -\text{voice} \end{vmatrix} \; / - \#$$

'Voiced consonants lose their voice when they occur before a word boundary.'

All that is now required of the phonetics is that it take the consonant as specified without voice and execute it. Such a view is simplistic and turns out to be wrong, as we shall see later.

Some phonologists have used phonetics to provide an explanatory basis for some processes in phonology. Again using the obstruent devoicing rule as our example, a phonologist might 'explain' the rule in invoking the fact that in final position there is a tendency for sub-glottal air pressure to fall, thus destroying the balance between air pressure and vocal cord tension needed to ensure vibration. Notice, though, that you can't have it both ways: either the rule is cognitive (in which case the tendency to lose vocal cord vibration is neither here nor there-you decided to stop the vibration) or it is not (that is, the aerodynamic constraint dominates), in which case the rule is physical and phonetic. But if the latter is the case then what about French which does not devoice final obstruents in the same way?

Both approaches to the handling of phonetic detail fail seriously because they make no genuine or lasting contribution toward bridging the gap between abstract and concrete. On the one hand the phonologist collects together anything in speech production which might seem to be cognitive in origin and on the other the phonetician deals only with the physical phenomena. There has been little attempt by phoneticians to adapt their models to the new cognitive linguistics.

One way of holding the two together has been to establish linguistic hypotheses whose concrete correlates are amenable to properly conducted investigation in the phonetics laboratory. We find some experimentalists in phonetics going out of their way to research hypotheses which are phonologically derived. For example there have been several studies of the phonetic correlates of abstract phonological distinctive features.

At best in the contemporary study of language a dynamic phonology outputs a string of abstract objects which are captured by a phonetics whose job it is to render these objects as sounds. Phonetics is just tacked on to phonology.

There are however many phenomena which a purely physical phonetics cannot satisfactorily account for. Besides the vocal cord vibration in French final obstruents mentioned above, there are examples of other coarticulatory phenomena which seem to vary from language to language (Ladefoged, 1967). The timing of vocal cord vibration onset in vowels following voiceless stops, affricates and fricatives varies considerably (Lisker and Abramson, 1964). The degree of nasality spilling into inter-nasal oral vowels varies between dialects of English (Morton and Tatham 1980a). The range of coarticulation induced displacement of the tongue position for lingual stops, affricates and fricatives preceding various vowels varies depending on the number of such segments any particular language has in its inventory. The variability in the precision of the articulation of a given segment in a particular segmental context varies from language to language dependent once again on the language's phonological inventory. A purely physical phonetics can accurately note such phenomena but it is quite unable to properly account for their occurrence. Why should the articulation of a particular segment vary in precision from language to language, or in different contexts in the same language when there is no neuro-physiological, anatomical or aerodynamic explanation?

Since the decision in the mid-1950s that linguistics should for the time being propose a static model of language characterising the underlying knowledge base, new applications of the discipline have come about. Thus for example there is currently a major preoccupation among researchers with speech and language processing, exemplified in the design of speech synthesis and recognition systems and artificial intelligence systems centering around the use of language. In its purest form this preoccupation is with computer *simulation* of the corresponding human processes: speech production and perception, and linguistic cognition. It is understandable that researchers in these areas of simulation have found comparatively little in contemporary linguistics to underpin their engineering work, since neither linguistics nor phonetics has seriously addressed what it means to develop simulation models or a corresponding theory.

We can see then that it is necessary to extend and develop phonetic theory in three important directions:

1. To improve compatibility between phonology and phonetics.
2. To explain some factual observations.
3. To make it usable to support simulations.

2. PHYSICAL PHONETIC THEORY

2.1 The Traditional Perspective

Phonetic theory is traditionally about the physical aspects of speech production. It assumes that cognitive processes at the phonological level have decided on what is needed at the phonetic level to produce the right sound wave. Phonology here is not the pre-transformational phonology which simply described the patterning of sounds at the surface. Nor is it standard transformational phonology which is simply concerned with the static knowledge base supporting such cognitive decisions. Phoneticians have always assumed a dynamic phonology, sometimes mistakenly supposing transformational phonology itself to be dynamic.

Figure 1 shows the layout of the traditional phonetic component. The input to the phonetic process consists of a string of *extrinsic allophones* (Ladefoged, 1967; Tatham, 1969): objects derived by phonological processes and which embody all cognitive processing needed to allow a non-cognitive phonetics to proceed. The extrinsic allophones are abstract objects.

A phonetic knowledge base details how these allophonic objects are to be executed physically, and specifies such things as target configurations of the vocal tract which would ideally achieve the objective of producing appropriate sounds, given the aerodynamic system involved. Phonetic processes draw on the knowledge base to realise physically the cognitive requirement.

PHONOLOGY	underlying level ↓ derived level	←	representation of underlying string of minimally specified segments
		←	fully specified representation of segments
	↓		
PHONETICS	underlying level ↓ derived level	←	extrinsic allophonic string (abstract)
		←	intrinsic allophonic string (abstract)

Figure 1. Phonetics following on from phonology. Note that the input to phonetics is an abstract extrinsic allophonic string, and that the output is an abstract intrinsic allophonic string.

Along the way the execution of the phonologically specified extrinsic allophones is degraded by artefacts of the physical system itself. Thus the neat boundaries separating segments in the input are lost as mechanical and other processes blur segments into one another. Further mechanical and aerodynamic constraints cause the desired articulatory targets to be undershot or overshot, such that what emerges is an almost continuous sound wave in which the original discrete segments are barely recognisable. In an abstract description of the resultant sound wave we refer to *intrinsic allophones* extrinsic allophones which have become degraded by physical constraints.

That the system nonetheless works as a satisfactory encoding of the original string of abstract objects is explained by the human perceptual system's ability to repair the degradation and in some sense recover the original abstract intention.

This version of physical phonetic theory emerged fairly clearly between 1965 and 1975 and has since become known generically as Translation Theory, although there are several variants. It follows the idea in linguistics that language encoding proceeds as a cascade of layered processes, each re-encoding the output of the previous layer until finally the original thought has been translated into an acoustic event. The metatheoretical shortcoming of the speech production sections of the theory is the failure to properly address the problems which arise because of the abrupt interface between wholly abstract phonology and wholly physical phonetics.

The same disjuncture is apparent in contemporary theories of speech perception which was modelled separately from production. Although some theories of perception involve drawing on a knowledge base of facts about production (e.g. the motor theory of speech perception (Liberman *et al.,* 1967) and the analysis by synthesis theory (Stevens and Halle, 1967)) speech production theory rarely refers to the perceptual process.

2.2 Experimental Support for Traditional Phonetic Theory

The phonetic theory outlined is supported by experimental work designed to discover what speech is and how it is made. That is, experimental work has usually concentrated on what goes on within the component, rather than investigate the nature of its input. A sample of the kinds of questions which have been addressed in phonetic studies over the past few decades might include:

a. What is the nature of the sound wave resulting from the speech production process?

b. How is the sound wave produced aerodynamically?

c. What articulators are involved in setting up the appropriate vocal tract configurations to produce the sound wave?

d. What muscles are involved in producing particular vocal tract configurations?

e. What is the general nature of motor control?

f. Is there anything special about the motor control of speaking?

g. How is the motor control mechanism organised, and what is the role of feedback, if any in the motor control of speech?

h. What are the mechanical, aerodynamic and other artefacts which degrade the desired articulation of separate segments?

2.3 The Shift in Research Objective

But as there comes about a change of emphasis from simple surface description to a proper *explanatory* model of the processes involved in speech production, and in particular to the formulation of a model oriented toward *simulation* of those processes, there is a shift in the type of question being asked. The vagueness of the question ‘*What* is *speech?’* turns into the more pointed question ‘*What underlies the speech we observe and describe?’*, and specifically for the purposes of this article: *‘What thinking* is *involved in speech production, how* is *it organised, what are the constraints on the cognitive processes in speech production and what are their sources?’*

It is necessary at this point to distinguish between cognitive *representation* and cognitive *processes*. For our present purposes cognitive representation refers to the way in which real phenomena are abstractly represented either in the mind of a speaker or in a model of speech production (or perception). Cognitive processes are the mental operations which manipulate those representations; some of these will be general cognitive processes not specifically speech oriented, while others may be dedicated processes.

Hitherto it has been usual to regard cognition as occurring only in phonology; or rather phonology has been defined as including all cognitive processes associated with speech production. Figure 2 illustrates the relationship between the cognitive and physical components.

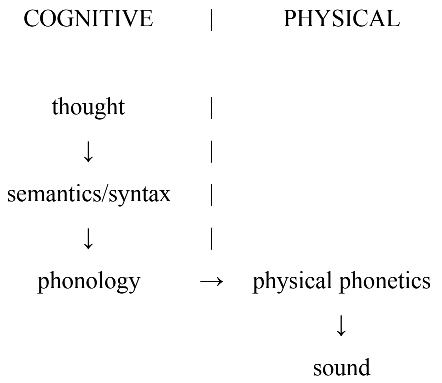

Figure 2. The usual relationship between cognitively oriented linguistics and physically oriented phonetics.

The awkward translation in the model occurs where the arrow crosses the gap between cognitive and physical at the juncture of phonetics and phonology. Physical phonetics has an *abstract* input, *physical* processes and a *physical* output. Figure 3 illustrates how the model looks if we set up parallel cognitive and physical models. Notice that language processing now occurs on both sides of the gap between cognitive and physical, with no attempt to cross it. There is no modelling of any interaction between cognitive and physical. Notice also that cognitive phonetics cannot have sound (a physical phenomenon) as its output, just as the neural processes involved in language encoding cannot have thought (a cognitive phenomenon) as their input.

COGNITIVE	|	PHYSICAL
thought	|	
↓	|	
semantics/syntax	|	neural processes involved in language
↓	|	↓
phonology	|	physical phonetics
		↓
		sound

Figure 3. Parallel models of language.

The extreme dualism of the parallel model is in fact unhelpful. One objective of the study of language is surely to investigate the relationship between cognitive and physical, or abstract and real. Whilst recognising the metatheoretical difficulties associated with attempting such an investigation *Cognitive phonetics* (Morton, 1986; Tatham, 1984, 1986) proposes the composite model shown in Figure 4. The gap between abstract and real is crossed at a lower level than in Figure 1. The new Cognitive Phonetic component in the model includes some of what hitherto was included in phonology simply because phonology was defined as including all cognitive processes involved in speech production, but which could not be justified on independent grounds as being other than phonetic, and additional non-phonological processes being modelled to account for some apparently non-physical behaviour in phonetics.

COGNITIVE	|	PHYSICAL
thought	|	
↓	|	
semantics/syntax	|	
↓	|	
phonology	|	
↓	|	
cognitive phonetics	→	phonetics
		↓
		sound

Figure 4. A single model relating cognitive and physical processes at a lower level than the model shown in Figure 1.

3. PHONOLOGY

3.1 Transformational Generative Phonology

Transformational Generative phonology and its derivatives envisage a set of rules for deriving an output level from an input level. Broadly speaking an input consists of a string of abstract objects minimally specifying the potential sound shape (not actual sounds) of sentences. The input string is not intended to convey anything but this minimal specification, but simply to characterise the sound shape underlying what will later be an acoustic signal.

Since it turns out that speakers of particular languages idiosyncratically re-encode or elaborate on these minimal sound shapes (although for effective communication purposes they need not) and that the design and operation of the actual speaking apparatus constrain what the physically unfettered mind might ideally do, additional adjustment to this input level is required to produce a final output string of abstract objects specifying how a sentence is to be pronounced. This process of adjustment is the phonology proper with constraints some of which are psychologically and some physically based. The linguist's phonology consists of a static knowledge base, and does not usually model the acquisition, storage, accessing or usage of the rules it contains.

The output level of phonology embodies all cognitively derived information necessary for the phonetic component to proceed to produce a corresponding sound wave for transmission to another individual. Because of the way phonology has accepted constraints from phonetics, phonetic production is guaranteed. It is the idea that the phonological output includes *everything* of cognitive derivation concerning speech production that places contemporary phonetics outside the cognitive domain.

The objects which are manipulated by phonological processes are abstract sounds or segments. They are parametrically specified using what phonologists call "distinctive features" to enable processes to be generalised between segments. So, for example, it is observed that all members of the subset of segments called consonants change the specification of the voicing parameter from presence to absence if they occur finally in a word or phrase. There is thus no need to write a separate rule for each occurrence of segments. Such processes are given in phonology as context sensitive productions.

3.2 Anomalies in the Output of Phonology

There is clear evidence that physical actualisation of phonological output is by no means automatic. That is, there is evidence at the phonetic level of systematic variation of actualisation which cannot be ascribed to the usual type of phonological process. Two examples illustrate the point:

- For a particular segment the phonological features are given a binary specification indicating whether it does or does not have a given feature. It is often the case however that phonetically a segment is rendered with systematic manipulation of the physical correlate(s) of the abstract feature going beyond binary representation.
- There are systematic changes in the sound wave when the rate of delivery of an utterance is varied, or has it rhythm altered in some way, where these changes cannot be accounted for by any known physical constraints but do seem to be occurring at the speaker’s will.

These and many other examples make it hard to maintain the notion of an automatic or purely physical phonetics without a good many adjustments to standard phonology-adjustments which are non-phonological in nature. So, for example, is it an automatic effect that dialects of English vary as to the amount of nasalisation of otherwise oral vowels they permit when these vowels occur sandwiched between nasal consonants? True, the occurrence of nasalisation is an automatic-a coarticulatory- effect but the occurrence of different degrees of nasalisation in different dialects could not be automatic, or it would be the same for all dialects.

Although it is the case that some low level physical constraints dominate phonological processes, apparently limiting the abstract sound encoding of sentences, others which are often severe seem to cause no adjustment within phonology. Many coarticulatory effects fall into this category. For the most part (but there are exceptions) there are seldom phonological rules which anticipate coarticulation and avoid or completely counteract the effect. The explanation usually given for this is that whereas it may be expected that phonology will avoid requiring phonetics to produce impossible sounds or combinations of sounds it will not avoid coarticulatory degradation of idealised sound in those cases where subsequent perception can *repair* the damage. There are however systematic cognitively based *manipulations* of coarticulatory effects taking place at the phonetic level, an example being the variable nasalisation mentioned above.

It is understandable that the early formulation of modern phonology did not take account of these phenomena-most have been noticed since that time. So we now have a choice: should cognitively originating manipulation of physical constraints be placed within phonology (since a purpose of phonology is to account for everything cognitive in speech production), or should we introduce a new cognitive dimension to phonetics?

3.3 Revising Phonology

The criteria for classifying a phenomenon as phonological need revision. There is a difference between the systematic substitution of one abstract segment for another or the systematic changing of the sign on an abstract feature (processes already characterised by phonology) and the systematic reduction or enhancement of a *necessary physical artefact* (which is what manipulation of coarticulation is). In other words, the prime distinguishing mark between phonology and phonetics is not that the one is cognitive and the other is not, for they can be each modelled cognitively or physically depending on which side of the vertical line (Figures 2-5) you stay.

What is important is the type of change which the rules specify. On the one hand what is being changed is abstract and comparatively free of physical constraint, on the other physical constraint dictates either a limiting or enhancing of something which is bound to occur anyway. Although in the phonetics the rule represents a potential cognitive act, that act is not psychologically free; it is utterly dominated by a physical inevitability.

Extending the earlier nasalisation example, consider the difference between *systematic enhancement* of the co articulatory nasalisation phenomenon and the decision to have a set of nasal vowels as part of the phonological inventory. The former is a cognitive phonetic act, the latter a cognitive phonological act. Internasal nasalisation of vowels is a phonetic coarticulatory artefact and as such might well fit within some current automatic phonetic theories. Nasal vowels as abstract phonological objects are not physical artefacts and we can choose entirely whether to have them or not. There is a major difference here. Inventory selection for phonology involves choice from within the available set of phonetically realisable abstract objects (or segments); inventory selection for phonetics is choice from possible modifications, within the limits set physically, of the physical realisation of the phonologically selected objects. It is choice which makes both cognitive, free choice which makes one phonological and choice only to manipulate the inevitable which makes the other phonetic.

4. GAPS IN PHYSICAL PHONETIC THEORY

4.1 Creating the Phonological Inventory

A major gap in current phonetic theory is the lack of an explicit mechanism for creating a phonological inventory. Human beings can make a very large number of sounds with their vocal apparatus. A subset of these sounds can be repeated on demand and perceived to be different, and it is from this subset that the sounds of any particular language or dialect are drawn. But how does phonetics make these known to phonology, or round the other way, how does phonology choose from what is available phonetically for systematic phonological usage in a language? There is nothing explicit in standard phonology or phonetics about the creation of phonological inventories.

What is really interesting about the nature of the inventory of sounds is the fact that there are numerous instances of enlargement of inventories by employing the systematic manipulation of phonetic artefacts discussed earlier.

Take the well-known aerodynamic artefact of aspiration and the corresponding measurement called voice onset time (VOT) (Lisker and Abramson, 1964).[1] We would expect that for a given stop the artefact would always produce, within a given range, the same effect. That is, for example, for a voiceless stop the VOT might be 45 ms ±10 ms; the 45 ms being a measure of the intrinsic aerodynamic artefact and the 20 ms range being a measure of the intrinsic variability of the effect. For many languages though this is not the case. In some languages we may find that voiceless stops are followed by no aspiration (that is, the VOT is 0 ms), in others the aspiration may be comparatively short-say, 5 ms. In yet other languages we find that phonetically identical voiceless stops may be followed systematically by up to four distinct zones of VOT rather than the one of other languages. In such cases e also find that what looked phonetically like identical stops turn out to be phonologically distinct, the phonetic correlate of the distinction lying not with different stop articulations but with systematically differing VOTs which have narrower zones which do not normally overlap.

What is clear from these observations is that the phonetically undeniable aerodynamic effect (undeniable because its existence and explanation have nothing to do with language) is somehow under linguistic control-is being manipulated for linguistic purposes. There are many examples of linguistic manipulations of phonetic artefacts in languages (Morton and Tatham, 1980a) and it is interesting that they can assume phonetic or phonological status. Sometimes they can be modelled as phonetic: the apparent curtailing of the nasalisation of vowels between nasal consonants in some dialects of English, or its enhancement in others. Sometimes they can be modelled as phonological: the limiting of aspiration in French and even more so in Italian are examples. The nasalisation example is phonetic because no contrast promoting perceptual identification of a segment is involved. The aspiration example is phonological because it signals contrast with the phonetic rendering of the opposing voiced stops which have a negative VOT in those languages (that is, vocal cord vibration is in evidence before the stop is released).

Yet more interesting are those languages which have several zones of VOT which phonetically convey phonological contrast (Korean is an example). Here, though, the status of the control of the intrinsic artefact is raised to the highest level in phonology-that of morphemic contrast-whereas in the Italian and French example the contrast is an implementational one at the lowest phonological level.

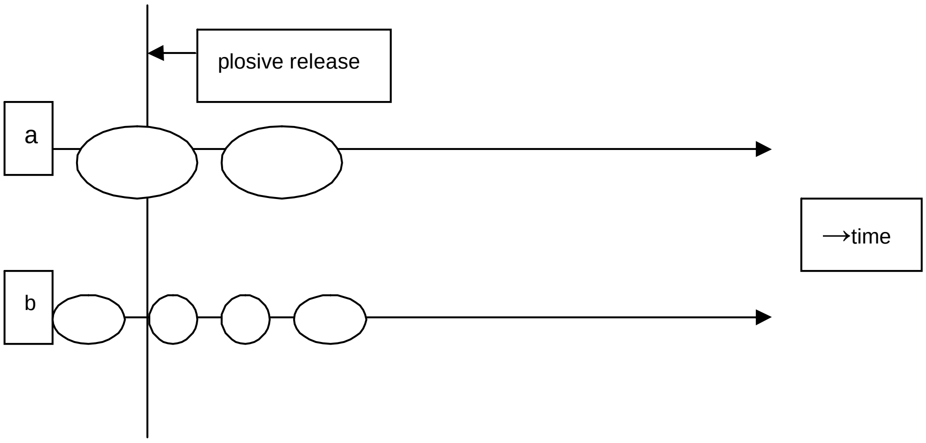

Figure 5. Voice onset time comparison between two different languages. 5(a) A language with only one voiceless plosive contrasted with a voiced plosive. 5(b) A language with three voiceless plosives contrasted with a voiced plosive (note the smaller zones of variability for the VOT).

The reason for dwelling on the control of phonetic artefacts is that we have here the means of increasing the phonological inventory available to languages: the phonetic system is able to produce not a set of directly controlled articulations with unhindered variability and artefacts limiting their use because of the need to avoid overlap, but is additionally able to control the variability to improve discrimination within the available space.

Figure 5 illustrates the effect of control. In Figure 5(a) we see a language which has just the voiced/voiceless opposition for stops in its phonology. Notice that the range of variability of VOT is quite wide, with the two zones all but overlapping. But in Figure 5(b) we see a language with not only a voiced/voiceless opposition, but also discrimination between the three voiceless stops. In total there are four zones of VOT which necessarily do not overlap: to prevent overlap the zones are narrowed down; that is, variability is restricted.

Current phonology has little to say about this phenomenon, and awards equal status to the different types of segment: the fact that the control systems in the phonetic rendering of such segments are different is of no importance. It might be as well, though, for phonology to recognise the phonetic differences since the risk of perceptual confusion among them is surely greater. There is a sense in which such segments are more risky both articulatorily and perceptually – the production control might be more likely to fail giving rise to acoustic degradation beyond the perceptual system's repair capabilities. At the very least a phonological marking of such segments would signal phonetics to be careful when it produces them. This is not so strange as it might seem, since we shall see later that this is one of the conditions in which we observe an increase in articulatory precision-and in the model it is necessary to identify the trigger for this.

4.2 Variable Phonetic Precision

An important gap in phonetic theory is the lack of an account of some areas of cognitive processing in speech production. Phonology takes care of the grosser, higher level cognitive processes: the language has this or that contrast between segments, segments can be modified at will in particular contexts, and so on. But what about the fact that we observe continuously variable precision in articulation?

I am not referring here to the observation, for example, that the duration of the stop phase of consonants in all languages seems remarkably precise and lacking in variability, or the fact that lip rounding in, say, English is remarkably variable even within the speech of a single speaker. Those are general observations about the overall behaviour of particular groups of segments or parameters. What I am referring to is the fact that the precision of anyone of these parameters varies itself during the course of an utterance: that is, intra-parameter rather than inter-parameter differences. Why is it that the precision (the inverse of variation) itself varies under certain circumstances? Answering this question begins with a systematic examination of those circumstances.

What we find is that the physical production of speech is sensitive to potential perceptual error (Tatham and Morton, 1980). If two sounds differ in only one parameter then that parameter is precisely executed if there is possible ambiguity (but not necessarily if there is not). A simpler example is the way in which increased precision is introduced into production when ambient noise increases--clearly a phonetic (not phonological) attempt to improve perception, involving knowing that perception will suffer under such conditions.

Thus a gap in phonetic theory is the failure to account for the relationship between production and perception where production must embody knowledge of perception. Several models of perception have included models of some aspect of production (either articulation-the Motor Theory-or the acoustics, the Analysis by Synthesis Theory), but no physical model of production includes a model of perception. Yet clearly perception is taken into account in speaking.

One reason the theory does not permit physical models of phonetics to consult their own models of perception is that such a process is necessarily cognitive, the output of such a process resulting in adjustment to the physical parameters of speech-a contradiction within a theory which places all cognitive processing in a prior (and separate) phonological component.

Much then of what is missing in a phonetic theory which is purely physical is an account of processes which must be cognitive but which are not properly phonological. The nature of these cognitive processes, their scope and their relationship with other cognitive components in language encoding (not just phonology) is the subject matter of Cognitive Phonetics.

5. THE THEORY OF COGNITIVE PHONETICS

5.1 The Relationship between Phonology and Phonetics

There is much to be gained by setting up a cognitively oriented phonetic component in linguistics which can parallel the more usual physically oriented phonetics. At the very least this would put on a proper footing the formal relationship between phonology and phonetics. At best it would set up a basis for a model which would account for much of the data unable to be handled by more traditional approaches to phonetics.

A traditional view of the relationship between phonology and phonetics is to think of phonology as concerned with assignment and phonetics with implementation. That is, the cognitive processes of phonology are about the assignment of this or that sound shape to a sentence, whereas the physical processes of phonetics are about the concrete implementation of the requirements of such an assignment. However the leap between the two is metatheoretically unsound, since they are two different types of models.

It is quite simply not satisfactory to attempt to interface physical phonetics with cognitive linguistics or specifically with cognitive phonology. Phonologists should realise that the theory of physical phonetics has little to say to them. We could, of course, devise a non-cognitively oriented phonology-a kind of neuro-phonology-as part of a more comprehensive neuro-linguistics, but for the present linguists are not inclined this way.

This is to take, for the moment, an extreme dualist view, presenting physical and cognitive as though there were some opposition or gulf between them. The position is, of course, arguable: why not take a monist view and solve the problem by bringing cognitive and physical, or abstract and real, together in a single unified theory which would make the problem disappear? The fact however that we could argue about which view to take means that we have as yet no obvious or agreed metatheory on which to build a new unified linguistics. There are signs that if we adopt the right framework for our model we might begin to bring the two together.[2] But for the moment let us stay with dualism, but move our phonetics into the other camp to make it cognitive.

5.2 The Attack on Translation Theories

Switching at the end of phonology between cognitive representation of speech and a physical one is the basis of a major criticism which can be levelled at linguistics. Theories which translate in this way from one sphere to another are inherently unsound, and the arguments have been lengthily and convincingly rehearsed in the literature over the past two decades.

A popular solution is that put forward by the proponents of Action Theory (Fowler *et al.*, 1980). Adopting from neuro-physiology ideas new to phoneticians they suggest that much of phonological assignment is redundant, and that a great many of what were hitherto thought to be cognitive processes are in fact intrinsic properties of much lower level physical systems. In constructing a physical phonetic model which is more satisfactory than its predecessor, action theorists merge into a unified model of speech production much of the earlier incompatible assignment and implementation. In effect they push the divide between abstract and real higher up the system, giving less scope to a now much reduced phonology. Their argument is that assignment is less free than previously thought since properties inherent in the peripheral physical mechanisms of speech production perform automatically operations earlier handled at a higher cognitive level.

Unfortunately Action Theory falls into the dualist trap also: '. . . linguistic segments as known and uttered must have context-free or invariant properties' (Fowler *et al.,* 1980, p. 375).

There is no sense whatever, of course, in which a segment in linguistics could be uttered-since the segment is an abstract descriptive construct of the discipline itself. But if linguistic segments means *language* segments (in the *mind* of the human being) then the statement that they must have context-free or invariant properties is correct, though they still could not be uttered directly. Once again, the only way we have of bridging the gap between abstract and real is to skirt the problem by saying that the cognitive unit somehow triggers the physical utterance. I agree that at its earliest stages the physical specification of an utterance unit (a segment) must be context-free and have invariant properties. Physical context sensitivity comes later. But within cognitive phonetics physical context sensitivity is known about, and can form the basis of a counteracting strategy.

Whilst action theory pushes the gap between abstract and real higher, cognitive phonetics in effect pushes it lower. Cognitive phonetics was developed independently of action theory, but in response to the same observation concerning the inadequacy of the translation relationship between phonology and phonetics. Cognitive phonetics is not however in any sense opposed to action theory, and is well able to take in the idea that assignment has less freedom than we previously thought. But there are one or two areas in which the surface data of speech is inadequately explained by the action theory model. Cognitive phonetics, by bringing down a cognitive strand from phonology to the lowest phonetic levels can complement action theory to lead to a better overall account of the data.

In action theory relationships between objects, such as muscles, at a low physical level are modelled as being intrinsic to systems of groupings of such objects. These groupings are called '*coordinative structures'*, and are responsible for the organisation of the movement of objects within the system as well as the relative timing of the individual movements of those objects. The relationships are characterised in the theory by *equations of constraint* which detail all relationships intrinsic to a system. The equations of constraint could be said to be a statement of the knowledge a system has of how it is to work. Physical knowledge held by the physical system should not be concerned with mental knowledge about the properties and effects of the system.

Cognitive phonetics characterises knowledge of these systems and their intrinsic properties. The knowledge is available to enable recruitment of the appropriate system at the correct time, and to provide the basis for the manipulation of the involuntary artefacts described earlier. The physical basis of the adjustments is the mechanism referred to in Action Theory as *tuning.*

5.3 Some Details of Cognitive Phonetics

The overall encoding system of language is designed to have an output which reflects differences between objects at the input; that is, thought is transformed into sounds, and there is a relationship between the two. The output differences must be systematic (or systematically derived) so that a decoding process can recover (within the limitations of the system) the original input objects; that is, the perceiver can reconstruct a copy of the original thought with recourse only to the sound wave and what he/she knows of language in general and the encoding system in particular.

Within cognitive phonetics there is a set of rules which have been called *'production instructions'* (Morton and Tatham, 1980b). It is these which organise the detailed strategy for manipulating the otherwise neutral or free-running physical system described by action theorists. They initiate tuning to provide ongoing adjustment of the system. Particular instructions are called because of what is known about the motor, aerodynamic and acoustic possibilities of the system. Calling production instructions is event driven; that is, the occurrence of a particular segment in a particular segmental or other context causes certain production instruction routines to be evoked. The knowledge required for such a system to operate is held within the mind, and so is treated cognitively-in this dualist view it cannot be physical.

The facts of the motor system are described within cognitive phonetics only because their detail needs to be known in order to be manipulated. Action Theorists have not paid sufficient attention, it could be argued, to this distinction. There is a difference between physical facts, how they are described abstractly by the scientist and their abstract mental representation in the human being.

6. MODELLING COGNITIVE PHONETICS

6.1 Knowledge of Articulatory and Acoustic Space

Both the sound and the articulation of a particular segment can be modelled as existing within domains allocated to that segment within an overall sound space (Tatham and Morton, 1980; Tatham 1984). Thus it can be determined experimentally by examining different vowel articulations, for example, that particular vocal tract configurations are associated with particular vowels. Looking closer and repeating each vowel on many occasions (but within one speaker for the moment) it can be seen that although there is often considerable variation in vocal tract shape for a given vowel segment there is in fact very little overlap between the ranges of shapes associated with the different vowels. The same is true for examples of vowel articulations and their acoustic equivalents taken from many speakers in the same linguistic community (Peterson and Barney, 1952).

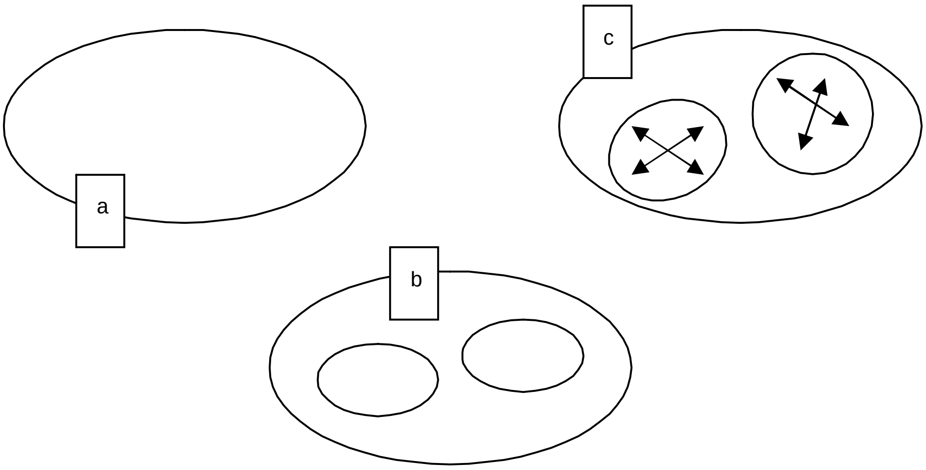

Figure 6. The relationship between spaces, domains and articulatory shapes. (a) Overall space in which articulations occur; (b) two domains within the space; (c) articulatory shapes within their allotted domains.

Thus we model a shape domain for a particular vowel, and speak of the system aiming to produce a shape within the given domain when required to give an instance of the vowel sound. We model the boundary size of the domain by reference to some abstract point within the space and departure from that point in different directions. We explain the existence of the domain by reference to the various neural, physiological and mechanical constraints within the system which contribute to failure to place the articulation at the idealised point on each repetition. Figure 6 shows, in the abstract, the overall space, domains and points. We note that domains for different but adjacent vowels often touch, or sometimes slightly overlap each other. We further note that domains vary in size, and that this size directly correlates with the distance between different vowels within the space. That is, the smaller the distance between two vowels the smaller their respective domains, or the less the variability in placing the articulation for each with respect to the abstract point in the vowel space. The larger the distance the larger the domains. But in all cases it is as though variability were being constrained to avoid overlap between adjacent domains. An explanation for this strategy might be the avoidance of perceptual confusion between the resulting sounds. The mechanism for this domain limiting strategy needs modelling within the overall articulatory model.

We see an example here of the coming together of linguistic (and therefore cognitive) assignment and phonetic implementation. Sounds exist within a space which can be determined and delimited by experiment. Particular sounds occupy domains within that space. Each sound can appear anywhere within its respective domain, but must not encroach significantly (determined by perceptual criteria) on the domain of any other sound. In an entirely abstract system the domains would not overlap at all. These conditions are linguistic since they are imposed by system requirements-that is, the linguistic encoding system is such that overlap is not permitted if the system is to function properly. Obviously such conditions of system requirements are dominated (or limited in scope) by the properties of the mechanism implementing the system: if it can't be done it can't be used.

6.2 The Phonetic Inventory

The next question to be addressed is *'How many such sounds are possible'*. That is, given a fixed space in which all sounds occur, how many sounds can be packed in and still satisfy the non-overlap condition? A secondary question here might be '*How many sounds are necessary within a language?'*. Obviously the number of sounds possible constrains the number of sounds available for any particular language. The number of sounds in a language must be equal to or less than the number of non-overlapping domains.

This observation, however, while seeming sensible is in fact too simple. It presupposes that the encoding and decoding systems are entirely linear in nature: but this need not be the case, and in fact is not the case. Trade-offs can occur to increase the number of possibilities for encoding, or which can decrease the demands made on the mechanism, or both.

The question which leads to realising that trade-offs between abstract requirements and concrete possibilities can operate derives from a closer examination of the word 'necessary' used earlier. It seems necessary to keep distinct phonological objects separate, and their related sounds separate in order to decode copies of those objects which are also distinct. This 'obvious' constraint holds though only when sounds are encoded and decoded in complete isolation-something that never occurs under normal conditions. If something of the occurrence of a phonological object is predictable from the occurrence of related contextual objects, then it may be possible to relax the distinction condition on the given object.

Such a system is an active one which actually 'considers' whether context might alter one-to-one object encoding (and decoding) to enable improvement in the system's efficiency. One option that stems from improved efficiency is the possibility of processing a greater number of objects than would be allowed were encoding and decoding not able to extend beyond the one-to-one consideration of single objects. The addition of context which must itself be recognisable, describable and repeatable *enlarges* the range of encoding possibilities in the sense that, provided what happens is adequately kept track of or indexed, it permits relaxation of the constraint that phonological objects and phonetic sounds should remain as optimally located within their domains.

6.3 The Varying Relationship between Encoding and Decoding

It is often the case that the division of labour between encoding and decoding is adjusted. Under conditions of high ambient noise, for example, less complex encoding is undertaken in order to reduce the loading on the decoder. That is, the speaker will articulate more precisely or speak more loudly to reduce the probability of decoder error. The trigger for the adjustment is the detection of the high ambient noise: an auditory signal requiring cognitive interpretation and calculated reaction. Similarly when the communication environment is optimal we find a higher level of encoding taking place, with the decoder working to a complementary higher level.

Though not perhaps of direct relevance to the study of speech under varying environmental conditions, it is interesting to note that similar variations take place in semantic and syntactic encoding. But what is of direct relevance here is the fact that, for example, if the speaker judges that he is about to use an unfamiliar word then he will momentarily reduce his speech rate and increase precision of articulation.

7. THE RELATIONSHIP BETWEEN COGNITIVE AND PHYSICAL PHONETICS

7.1 The Theory of Action

If we believe that motor operations in articulation are parametric in nature then we can model speech production at this level in one of two major ways.

1. We can assume that there is complete independence of the parameters, each controllable independently of all others, and that the constraints on control of each parameter are independent from the constraints on the others.
2. We can assume that the dependence between parameters can be characterised.

Before considering the data which might force us to choose one or the other model, it is worth reviewing the possibilities of each.

7.1.1 Completely Independent Control of Parameters

There is linguistic dependence between the physical parameters of speech. Features, whether considered in articulatory or acoustic terms, come together as a physical representation of abstract phonological objects. The reason for viewing the system parametrically is that we can observe the different parts of the system coming together in different combinations to achieve their objective. The question is whether the physical parameters have independent control.

If we assume independent control we are forced into the situation of establishing a dependence between the parameters at a relatively high level in order to reflect their linguistic interdependence. Intrinsic independence of parameters means that any apparent cooperation between them for encoding purposes must have been intended deliberately and calculated at a relatively high level in the system. It also means there is little or no constraint on the possible combinations of parameters. To a large extent phonology assumes this to be the case.

For those high level calculations to have taken place every detail of the individual motor control possibilities for each parameter must be known to the part of the system which calculates how they are to interact. The processing device needs very complete knowledge of the general properties of the system. It also needs complete knowledge of its current status. So, for example, suppose one of the parameters is the contraction of the lip musculature. The processor needs to know what the contractile possibilities are and what sorts of signals to send to the musculature to achieve a particular lip configuration. But suppose that we are dealing with a vowel segment requiring a certain degree of lip rounding, and that the vowel has been preceded by a bilabial plosive: the lip musculature was already contracted to a certain extent. We do not need now to send the same signal for vowel rounding as would have been necessary if the preceding segment had been, say, a velar plosive without lip involvement.

Independence of parameter control leads to maximum versatility in configuring the vocal tract, but necessitates a huge computational loading on that part of the system charged with working out the control values for each parameter.

7.1.2 Interdependent Parameters

If on the other hand parameters are not independent, but somehow have a built-in interdependency such that one parameter cannot respond without a correlating and well-defined response in an associated parameter then the situation is very different.

In such a case higher levels in the encoding system are constrained by the lower level's interdependence of parameters. Features within the phonology no longer have the abstract independence usually given them. There is a very real sense in which the physical system dominates abstract possibilities.

However, since much of what in the alternative system needs to be calculated and programmed is now an intrinsic property of the low level system the computation loading required for calculating control signals is significantly reduced and the actual control signals themselves made simpler.

Before Action Theory, independence of physical parameters was by and large assumed, accompanied by a correlating freedom of parameter specification at the phonological level with a heavy computational load assigned to the phonetic motor control system. Since Action Theory we now understand that it is a fact that there is interdependence intrinsic to low level physical systems and phonetics has moved toward the idea of simpler motor control. The idea has not yet filtered through to phonology which in turn will be forced to

- recognise that there may be severe constraints on what is phonologically possible, and
- introduce some characterisation of those constraints at an abstract level.

The problem here is the one discussed in some detail earlier: the constraints imposed by the intrinsic properties of the low level physical structures seem to predict a smaller number of distinct vocal tract configurations than is actually observed in speech. The solution adopted in Cognitive Phonetics has been the introduction of the notion that *via* the tuning mechanism the intrinsic dependency constraints which are not absolute can be modified on demand.

Hence a compromise. Totally independent control of parameters gives versatility and high computational load-this is now rejected. Action Theory proposes interdependence of parameters giving rise to lack of versatility, simplicity and low computational load-this does not entirely accord with the observed facts. Cognitive Phonetics proposes interdependence of parameters which is not absolute and which can be, and is, modified on a continuous basis. This gives rise to the proposal for a dual control system: on the one hand a simple general signal for control relying on intrinsic mechanisms for achieving fine action, but on the other had a parallel signal directed on continuously variable tuning of the intrinsic properties of the low level structures.

8. CONCLUSION

The abstract cognitive phenomena described by cognitive phonetics are linguistic, though they are not phonological (in the way we normally understand the word).

The phonetic system is balancing three considerations:

1. its phonological input;
2. system constraints;
3. output decodability.

This results in an output which varies under a number of conditions including input variations, idiosyncracy, system error, ambient noise, varying decoding ability and conditions.

Establishing the spatial domains within which phonetic objects are to be located is crucial. The distribution of those objects is critical to the efficiency of the encoding/decoding system. The abstract patterning of the phonetic object is handled by phonology. Details, such as the precision of the boundaries of the spatial domain of an object is phonetic, and although as abstract as phonology, should no longer be thought of as idealised in the meaning of that word in linguistics. It is abstract in the sense that it is cognitive, involving knowledge, manipulation of that knowledge, the gathering of knowledge and its use in active decision making.

Cognitive phonetics deals with the mental processes involved in encoding and decoding the final stages of the transformation of thought to sound. It involves what needs to be invoked when inputting an idealised phonological requirement with a view to outputting a sound wave which has to be decoded back to some copy as little degraded as necessary of the original thought. It is about considering the manipulation or recruitment of implementing mechanisms, on a long term, short term or on-going basis. It is about the cognitively dominated manipulation of the control of these mechanisms.

Such processing requirements necessitate access to data or a knowledge base. The knowledge has to have been gathered, stored, retrieved and processed. It concerns physical mechanisms and how they are controlled, as well as information about the range of performance requirements that might arise. In addition, there is the acquisition of data monitoring on-going performance, which may need treatment separate from the stored long term data.

Cognitive Phonetics also has a role distinct from that in on-going motor control. It is the source of information about available mechanisms, their effects, limits and constraints for cognitive phonology. Cognitive phonetics takes over the role hitherto held by non-abstract phonetics in its relationship to phonology. It is the component of the grammar following the phonology, and it is responsible for what in phonology is phonetically dominated.

In discussing in this chapter the need for cognitive phonetics, what it is about, how it works, and how it relates to phonology and physical phonetics, attempts have been made to show that much of speech production at a level below the necessarily idealised role of phonology is nevertheless cognitive. It has been claimed that the component as described is within the metatheoretical spirit of contemporary linguistics, while at the same time attempting to achieve a relationship with the physical world which is not required of linguistics. If anything, phonetics is an extraordinarily difficult area to deal with because it is here that abstract and real, cognitive and physical meet in language.

NOTES

1. VOT is the time between the release of a voiceless plosive or end of a voiceless fricative beginning a syllable and the start of vocal cord vibration in the following vowel. The delay in vocal cord vibration onset is often referred to as aspiration. The phenomenon is modelled as an aerodynamic artefact caused by a disturbance in the critical difference between intraoral air pressure and subglottal air pressure needed to induce vibration in vocal cords of a given tension. The rise in intraoral air pressure as a result of the stop or stricture (in the case of a fricative) is what has upset the balance.
2. Recently the parallel distributed processing approach (Rumelhart and McClelland, 1987) has been used by researchers in several areas of speech production and perception, particularly in simulation models aimed at solutions in speech synthesis and speech recognition. Neural networks (the programming implementation of parallel distributed processing devices) have been used to model both cognitive and neural processes satisfactorily, uniting the abstract and physical by way of a common mathematics. The neural network is particularly useful when set up as a learning device: it is able to establish for itself relationships or associations between a given input and a given output. Neural networks have been used with some success in the Advanced Speech Technology Laboratory at Essex University to learn relationships between abstract representations (such as strings of extrinsic allophones) and parameterised associated waveforms in either direction, simulating production (synthesis) or perception (recognition). The advantage of a device capable of establishing systematic relationships for itself is that the researcher does not need to program rules into the system. Thus in a situation where we know that two objects are related, but do not know how, the neural network can establish a relationship (set up rules) for us. There is, of course, no guarantee that these relationships are those a human being uses.

REFERENCES

Chomsky, N. (1957) *Syntactic Structures.* The Hague: Mouton.

Fowler, C. (1980) Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8, 113-133

Fowler, C., Rubin, P., Remez, R.E. and Turvey, M.T. (1980) Implications for speech production of a general theory of action. In B. Butterworth (ed.) *Speech Production,* Vol. 1. New York: Academic Press.

Ladefoged, P. (1967) Linguistic Phonetics. *Working Papers in Phonetics 6*. Linguistics Department, University of California at Los Angeles.

Liberman, A.M., Cooper, F.S., Shankweiler, D.S. and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review* 74, 431-461.

Lisker, L. and Abramson, A.D. (1964) A cross-language study of voicing in initial stops. Acoustical Measurements. *Word* 20, 384-422.

Morton, K. (1986) Cognitive phonetics-some of the evidence. In *In Honor of Ilse Lehiste.* R. Channon and Linda Shockey (eds.). Dordrecht: Foris Publications, 191-194.

Morton, K. and Tatham, M.A.A. (1980a) Devoicing, aspiration and nasality – cases of universal misunderstanding? *Occasional Papers* 23, 90-103. Department of Language and Linguistics, Essex University.

Morton, K. and Tatham, M.A.A. (1980b) Production instructions. *Occasional Papers* 23, 104-106. Department of Language and Linguistics, Essex University.

Peterson, G.E. and Barney, H.L. (1952) Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 182.

Rumelhart, D.E. and McClelland, J.L. (1987) *Parallel Distributed Processing.* Vols. I and II. Cambridge, Mass.: M.I.T. Press.

Stevens, K.N. and Halle, M. (1967) Remarks on analysis by synthesis and distinctive features. In *Models for the Perception of Speech and Visual Form.* W. Wathen Dunn (ed.) Cambridge, Mass.: M.LT. Press, 88-102.

Tatham, M.A.A. (1969) Classifying Allophones. *Occasional Papers* 3,14-22. Department of Language and Linguistics, Essex University.

Tatham, M.A.A. (1984) Towards a cognitive phonetics. *Journal of Phonetics* 12, 37-47.

Tatham, M.A.A. (1986) Cognitive phonetics_some of the theory. In *In Honor of Ilse Lehiste.* R. Channon and Linda Shockey (eds.). Dordrecht: Foris Publications 271-276.

Tatham, M.A.A. and Morton, K. (1980) Precision. *Occasional Papers* 23, 107-116. Department of Language and Linguistics, Essex University.