Some Problems in Phonetic Theory

Mark Tatham

Reproduced from: Amsterdam Studies in the Theory and History of Linguistic Science IV: Current Issues in Linguistic Theory 9 *Current Issues in the Phonetic Sciences*, eds. H. and P. Hollien, Amsterdam: John Benjamins B.V., 93-106 (1979).

In this paper I want to highlight some of the difficulties which arise when one considers phonetics as a part of linguistic theory. Several researchers have argued that phonetics should be viewed as following on from the phonological component of a transformational generative grammar, accepting as input the output from that component. Others have sought to relate phonology and phonetics by making suggestions about details of the phonology where consideration of such details is phonetically motivated — as, for example, attempts to provide a phonological feature set relating ore closely than that proposed in, say, *The Sound Pattern of English* [Chomsky and Halle 1968], to alleged phonetic facts. Indeed, many see phonology as in danger of being too abstract and seek to relate its abstractions to the real world by incorporating phonetic observations. For myself, I don't want to do anything to phonology — or at least not very much: I would rather make proposals concerning phonetic theory. As I see it, phonetic theory is at odds with phonological theory: not only are the data sets of each different in type, but the formats of the current theories are so incompatible as to at least run the risk of technically rendering vacuous any 'corrections' to the phonology which might come from phonetic theory. I shall try to show why I think there are problems of mismatch between the two theories and make some suggestions to improve compatibility

Let us consider some data: In French there is a vowel /u/. This vowel is characterized in the phonology by being high and back and at the same time by being round. There is another vowel /y/ which is characterized by being high and front, whilst also being round. There is also the vowel /i/ which is high and front like /y/, but differs from /y/ by being non-round. Notice that /u/ differs from no other vowel along the rounding parameter alone: all other vowels in French have a different high/low—front/back characterization. In the case of /y/, however, there is distinction from /i/ only along the rounding parameter.

Now, moving to phonetics, suppose we examine the lip rounding of /u/ and /y/. This is a relatively simple task using the technique of electromyography which enables us to examine the degree and duration of contraction of muscles — in this case the *orbicularis oris* which is a sphincter muscle running around the lips and responsible, in one of its modes of contraction, for lip rounding. As with all physiological parameters, measurements of the contraction of a muscle over repetitions of a phonologically 'same' utterance show a certain degree of variation. That is, if, for example, I repeat the word in French *doux* (/du/) several times, some variation from example to example will occur in the contraction of *orbicularis oris*. However, if we compare the range of variation exhibited by this muscle over repetitions of *doux* with the range of variation in repetitions of, say, *du* (/dy/), we notice that the range is significantly less for /y/ than for /u/. This difference in range exists notwithstanding the overall increased rounding of /y/ and the lip protrusion associated with the articulation of that sound. It has been suggested that this narrowing in range can be interpreted as reflecting an increase in the precision with which any one of the repetitions was articulated. If this is the case — that is, if a native speaker of French articulates /y/ on the lip rounding parameter with more precision than when he articulates /u/, then we might well ask ourselves the question Why? A possible explanation might be that since the difference between /y/ and at least one other vowel, /i/, is carried only by the lip rounding feature it becomes crucial to get the lip rounding right to avoid confusion at a perceptual level between different morphemes; for example: /dy/ and /di/ in *j'ai dû* — 'I had to,' and *j'ai dit* — 'I said.' It may be that the difference in meaning between these two phrases rests entirely on the precision of contraction of a single muscle.

Another piece of data: take the minimal pair in English *cap* and *cab*. At an underlying level in the phonology these two morphemes are distinguished only in the third segment and on the feature [voice]. They may also be differentiated on the feature [tense] if that feature is not vacuous: /p/ being [+tense] and /b/ being [-tense]. A phonological rule operates, lengthening the vowel segment before the [+voice] consonant in cab, and this is followed at some stage in the phonology by a rule which changes [+voice] to [-voice] in voiced obstruents in final position. Thus, *cap* and *cab* are differentiated at the underlying level on the final segment, but (and still in the phonology) at the derived level of systematic phonetics, on the penultimate segment. The phonetic data is in accord with the surface phonology in this case (but see below): vocal cord activity is indeed prolonged during that portion of the utterance which may be said to correspond to the vowel segment in the phonology and is noticeably absent during the part of the utterance corresponding to the final phonological segment — this is in the word *cab*. In we also note the absence of vocal cord activity at the end of the utterance and that its duration during the identified vowel section is less. Certainly in my English when either of these words is in sentence or phrase final position the final bilabial occlusive is often unreleased, and it would follow that at least from an acoustic viewpoint the morphemic difference between the words is carried by the relative duration of the vocal cord activity. I imagine that one would note in this example a relatively high degree of timing precision on this parameter.

Yet a third piece of data: consider in my English the utterance /man/ — *man.* Phonologically at the underlying level we have a nasal consonant /m/ followed by a non-nasal vowel /a/, followed by nasal consonant /n/. At the systematic phonetic level, specification of the nasal feature would not have changed — though some may argue with this. An examination of the actual articulation of this nasal—non-nasal—nasal sequence, however, reveals a certain degree of phonetic nasality in the vowel — that is, the velum does not quite close at any time between the two consonants. Several dialects of English have rather more phonetic nasalization of the vowel segment than I do — and they have this greater degree of nasalization consistently. Since there is no contrastive nasal feature on vowels in any dialect of English — that is, since the feature [nasal] is always marked minus at the underlying level in the phonology in vowels, and since there is no phonological rule transferring any contrast from an adjacent segment onto this feature of vowels, precision at the phonetic level with respect to opposing oral vowels and nasal vowels is unnecessary.

Now, let us look back at this data and try to fit it into a theoretical framework which integrates both phonology and phonetics. The output of the phonology — that is, at the level of systematic phonetics — can be taken as a characterization of the requirements to be imposed on the phonetics. As such there clearly is going to be nothing there which is utterly impossible for the phonetics to carry out on the one hand, and on the other it is said that all information will be there for the phonetics to proceed as an automatic, that is, non-linguistic component: nothing linguistically determined can be added during the phonetic realization of the phonological requirement.

Now, note that phonetically /y/ is realized with more precision than /u/. How does the phonetics know that more precision is required?

Take the second piece of data: vowels before voiced obstruents are to be phonetically lengthened and voiced obstruents in final position are not to have vocal cord activity. Let me say that last part again: if the systematic phonetic level characterizes the phonological requirement to be imposed on the phonetics then a marking of [-voice] on the final obstruent should mean 'do not vibrate the vocal cords'. I have some further data (courtesy of Tom Shipp, VA Hospital San Francisco) derived from an experiment observing electromyographically the behaviour of the laryngeal muscles during these phonologically devoiced segments. The data clearly shows that although the vocal cords were not vibrating the laryngeal muscles were doing all the right things to bring about vibration — the vocal cords did not vibrate because of the fact that the air pressure differential above and below the vocal cords was incorrectly balanced to allow spontaneous vibration to take place. In other words the data suggest *not* that there was no intent to voice but that on the contrary there was every intent to voice. This is a demonstration of the classical mistake of making incorrect deductions from surface observations.

If we revise the phonology and do away with the devoicing rule on the grounds that, whatever happens phonetically, voicing is signalled as a requirement, we give ourselves quite a problem. Voicing is required: it does not happen. A coarticulatory influence has prevented vibration, and this is of no consequence because phonologically the morphemic contrast is redundantly signalled by the vowel length. But look at final phonologically voiced obstruents in a language like French: there is vocal cord vibration. The phonetic action taken in French to obtain vocal cord vibration for these [+voice] obstruents must be different from that taken in English or the results would be the same in both languages.

On to the third example: the phonological requirement for the vowel segment in *man* is that there should not be nasality — there usually is, however — but in different dialects there is more or less. And this more or less is perfectly consistent. It must surely follow that what is regarded often as an artefact of coarticulation is nevertheless systematic in a way which is not determined by the mechanical or other considerations which determine that the artefact shall occur. In other words, it is clear that control is being exercised over the artefact — just as it is clear that in French control is being exercised over the air pressure balance artefact — and just as it is clear that control is being exercised over the variability artefact in the case of French /y/.

I hope it is by now obvious where I'm going in my argument. What may not be obvious is why I chose those three examples: there are several other which could have been chosen. I selected these because they are indicative of the control of different artefacts for different reasons. My lip rounding example is about maintaining precision to preserve morphemic contrast and shows phonetic attention to the most fundamental use to which the features of a segment can be put. My voiced obstruent example — besides showing that the inclusion of a devoicing rule in the phonology of English is probably wrong — is about maintaining at the phonetic level a different interpretation of a phonological requirement seemingly identical across two languages as far as the linguistic contrast is concerned. My oral-vowel-between-nasal-consonants example is similar to the voiced obstruent one, but is about different interpretations of the phonological requirement in different dialects of the same language.

In each of these cases we are talking about adjustment of a property which is essentially phonetic for a specific linguistic aim: to maintain contrast dictated from a high level, to maintain a low-level not crucially contrastive variant operating differently in different languages, and to maintain a variant operating differently in different dialects of the same language. To include these operations in the theory we must decide at what levels they are introduced.

Let us backtrack a little. At the input to the phonology morphemes are correctly strung together to provide a linguistic encoding of a high level concept or idea otherwise non-transferable between human beings. I am saying something quite obvious and uncontroversial. Language is an encoding/decoding system enabling the copying of ideas between brains. These ideas are encoded into soundwaves which exit from one human being and transfer to another human being where a decoding process takes place. Such an encoding/decoding system must have several properties, one of which is that the encoding and decoding algorithms must be complementary in some sense. One unfortunate property of this type of encoding/decoding system is the spurious introduction all along the line of noise, distortion or error. In the abstract various techniques are in principle available to minimize the introduction of errors or to detect and inversely filter — that is, negate — errors when they occur.

In principle when designing an encoding/decoding system the designer predicts the occurrence of such errors and either avoids those situations where they will occur or takes steps to minimize or cancel them as completely as possible — or rather, as completely as *necessary*: the distinction between 'as completely as possible' and 'as completely as necessary' is an important one as I shall try to show in a moment.

Up to the input to the phonology there have been constraints at work dictating in a sense some of the format of the syntactic strings able to be generated. These constraints can often prove extremely revealing a to the nature of the brain and its workings, and linguists go to considerable lengths to highlight classes of constraint and hierarchies of constraint. The constraints to this point — the underlying level of the phonology — have been essentially neurological or psychological and have been revealed to the linguist by adopting a metatheoretical stance which begins by setting up a too-powerful model and successively limiting that power.

The constraints imposed on a phonology are often, I though, of an essentially different nature. Not only are we dealing with neurological and psychological limitations on the encoding process — but also with phonetic constraints: that is, motor, mechanical and acoustic constraints at least. And, picking up a phrase I used earlier, the phonology is not going to require a phonetic impossibility.

Phonological encoding is about transforming strings of morphemes into strings of phonetic requirements — that is, into strings of objects (in the abstract sense) which, when phonetically encoded into soundwaves, will sufficiently enable a decoding device to extract all the relevant information from the signal. The phonology is, however, doing a double task. Whilst attempting error minimization (and we shall look at at least one way it does this in a moment) it is in addition required to introduce variants into underlying strings — merely for the hell of it. The classical example of this is the introduction of the palatal and velar alternates of an underlying /1/ in English. Some researchers claim that this serves to aid perception, but I doubt if that is the reason.

One major error minimization device is redundancy. A given piece of information is stated and restated in different ways. An essential property of this redundancy in phonology is that it is patterned — that is, predictable. The values of certain features in the specification of segments may be predicted from the values of other features, or indeed whole segments may be predicted from their context.

At any rate, we exit from the phonology at the derived or systematic phonetic level with a string of segments designed to take account of the possibilities of phonetic realization and which has as accurately as possible encoded the input ready for conversion to soundwaves, and which has already built in the possibility of correcting some errors which might arise later in the encoding/decoding process, and which has catered for any idiosyncratic alternations. All of these phenomena we may call linguistic and all transformations occurring during phonological encoding are systematic or rule governed.

Now, and only now as far as we are concerned today, do the real problems begin, and central to those problems is the one of *precision*. Neural control of the organs of speech is not entirely precise even under ideal conditions — or rather perhaps I should say the results of neural control (the actual positioning and movement of the organs of speech) are not precise. The degree of precision varies — but is particularly vulnerable to the constraint of time, and vulnerable also to the constraint of context: that is, the precise achievement of a desired configuration of the vocal tract may be particularly difficult (or impossible) given preceding and following configurations. When a human being constructs a mechanical system where precision is of importance he generally builds in some kind of monitoring device which can adjust the system if it begins to run wild. As you know, there has been much discussion in the phonetics literature as to whether there are such devices in speaking, what their exact role is, and so on. I don't wish to extend that particular discussion here, except to say that I believe that such devices do exist but that they have a limited role to play inasmuch as they often operate too slowly to be effective on a segment-by-segment basis.

Time is our big problem. Notice that phonology seems to have forgotten about time. Notice also that it works perfectly well without it (since it is abstract) — or rather, in the human being, if it has to take account of time in its own operation, then the temporal constraints *it* suffers are those imposed by the time required for the transmission of neural impulses and for computations *within* the brain. When we're talking about transmission of neural impulses to muscles, the time it takes them to contract and the time it takes the articulators to move — then we have entered a quite different area of problems.

The organs of speech assume different configurations for different phonological requirements, and the accuracy of these configurations is constrained firstly by the fact that, all other considerations (including time and context) being optimum, precision is not 100%, and constrained *in addition* by time and segmental or spatial context.

But notice — now back to my original data — that degree of precision varies. Lip rounding for /y/ is more precise than for /u/; spontaneous vocal cord vibration during final phonologically voiced obstruents is more carefully controlled in one language than in another; nasalization of phonologically non-nasal vowels is allowed more in one dialect of a language than in another. So the precision *is* controlled and is therefore controllable. Two questions immediately spring to mind: *how* is it controlled? and *why* is it controlled? And two further questions: at what stage in the encoding process does this control take place? and what are the limits of the control?

To the question 'why is it controlled?,' I have in a sense already provided some kind of answer. It is controlled to maintain morphemic contrast, or to maintain some idiosyncratic surface output. The other questions are more difficult to answer. Firstly, *how* is it controlled? For short-term precision, say within a segment, I don't believe the degree of precision. is controlled by monitoring followed by subsequent adjustment: the known monitoring systems, from the slow auditory feedback through to the comparatively fast gamma loop servo system, are just not fast enough — in any case there is every evidence that we go to a particular configuration directly without a semi-oscillatory onset which would be a property of a servo controlled system. And also in any case the servo system would have to be set, and setting the system would depend on knowing just what setting is required.

Now, that last remark leads me straight into a suggestion as to how we control precision we control precision by prior computation, and of necessity the prior computation must involve consideration of the imprecision which is going to occur unless steps are taken. So, spontaneous vocal cord vibration will not occur unless the right balance is obtained between supra- and sub-glottal air pressures, and this fact must, of course, be available for inclusion in the computation to obtain vocal cord vibration. What also must be known is that in obstruents, as opposed to, say, vowels, the air pressure balance will be disturbed because the free flow of air out of the mouth will be interrupted. In other words the system must have knowledge of firstly that [+voice] as an abstract phonological requirement involves vibration of the vocal cords, and secondly that vibration of the vocal cords will not occur if other features specify interruption of the airflow (the abstract phonological features are mutually exclusive if the 'normal' algorithm for obtaining vocal cord vibration is followed), and thirdly what to do about it if necessary.

But that is not all. We see from the data that sometimes we do not get vocal cord vibration with phonologically voiced obstruents and that sometimes we do — so that it must be the case that a decision has been taken as to whether or not to overcome the effects of the predicted constraint. Such a decision occurs with predictable outcome — in French it goes one way, in English the other. The decision is therefore principled.

There is an alternative model possible. I have assumed so far that the phonological requirement for vocal cord activity is signalled by marking the voice feature as [voice] at the level of systematic phonetics in both languages. I have done this on the basis of similarity of function of the segments in question in English and in French. I went on to assume that despite similarity of function an idiosyncratic (or contrastively unmotivated) decision was taken to disregard the predictable non-occurrence of vocal cord vibration in English, but not to disregard it in French. The alternative model would suggest that the phenomenon was akin to the velar/palatal /1/ alternation in English and that somehow the voiced obstruents in French were marked at the end of the phonology in a special way allowing the automatic realization of those segments *with* vocal cord vibration. I do not believe this alternative model to be appropriate — at least not on those grounds. Surface variants accounted for in the phonology seem to me to be properly those which do *not* involve the containing within limits of a phonetic constraint.

Now I seem to have contradicted myself, for I said earlier that I believed that the phonology took into account phonetic constraints. If it does not seem too much like playing with words, let me explain that I mean the phonology takes account of phonetic constraints on the phonology, whereas the phonetics takes account of constraints on the phonetics. Some constraints cannot be overridden or modified in any linguistically useable way: obviously the phonology must take account of this. But some constraints can be modified and the extent to which they can be modified will determine their phonological 'usability' — but the decision as to just what degree of overriding is to be executed is taken by the phonetics. To do this the phonetics would need to know the phonology.

So you could imagine a series of phonetic questions: what is the phonological feature to be executed? what does this involve? what constraints will be encountered? do these matter? if yes, then how are they overcome? Notice that if the answer is 'yes,' then by definition they are able to be overcome — or the phonology would never have made such demands on the phonetics.

Now, even if you do not agree with me, bear with me because if all this *is* an adequate model then one or two interesting questions arise. How *does* the phonetics decide whether or not overcoming constraints is necessary? In what sense could the phonetics be said to know the phonology? — to know, for example, that precision in the lip rounding of /y/ is more important that in /u/? Well, some writers have suggested a sort of summary of the phonology — particularly of redundancy — somewhere around the end of the phonology or beginning of the phonetics. Their reasons have varied, sometimes concerned with accounting for diachronic sound change, or sometimes accounting for syncope (i.e., segment deletion), etc. Postal (1968) was rather emphatically concerned to avoid any what he called 'independent level of autonomous phonological representation' on the grounds that such a level would involve spurious duplication of rules and perhaps loss of generalization. However, I believe that to involve the phonology proper in anything non-phonological in the strictest sense would be ill-motivated, and this leaves us with a non-automatic phonetics. And a non-automatic phonetics means a phonetics which is *linguistically sensitive*.

I have been treating the phonetics as an encoding system. One kind of encoding system can be identified as a passive system — that is, for any given input a well defined and invariant algorithm is operated to give the output. The passive encoding system is, however, a special case of what might be called *active* systems. An active encoding device is sensitive to its input in a rather interesting way. In the particular case of phonetics a passive encoding would entail segment-by-segment (or feature-by-feature) scanning of the input and segment-by-segment execution of the encoding. An active phonetic encoding would scan more than one segment (or feature) at the input, and, as a consequence of seeing this segment (or feature) in a particular context would adjust the encoding algorithm. Note that the trigger for adjustment has not come from the input segment or feature itself, and note further that the same segment or feature introduced n a different environment would be encoded by a different algorithm. When scanning the input what the phonetics would be looking for is difficulty of execution of individual segmental requirements. How a particular phonological feature is to be encoded varies with the marking of other features associated with that segment and with the marking of those features in surrounding segments. Clearly, such an active encoding would be unnecessary if the phonology had anticipated all the problems and specified each systematic phonetic segment exhaustively — and that is a position some researchers may care to take. I believe, however, that a solution of *this* kind does not accord with the facts or provide as much insight into speech as the former solution.

I favour the active encoding solution for a number of reasons. Let me give just one example. It is clear that on many occasions we can choose alternate phonetic encodings. I can say, for example, *cap* with a released /p/ or [cap'] with an unreleased /p/ — the most the phonology might signal is that it is linguistically unimportant whether there is release of the obstruent or not. If release is either to occur or not, then, despite the fact that it doesn't matter which, a decision must be taken as to which to do, since both involve active motor control of the articulation. Such a decision could hardly be phonological if there is no phonological interest in the outcome. Some researchers have suggested that degree of phonological interest might be signalled to the phonetics by arranging the features in segments at the systematic phonetic level hierarchically, with crucial features higher than less crucial features. Note that how crucial a feature is will vary with phonological context.

I am trying to show, of course, that we cannot have an autonomous automatic phonetics. And if we are to have phonetics integrated with linguistics then we are up against the metatheoretical problem of format. One cannot legitimately integrate two differently formatted theories. I propose a reappraisal of the format of phonetic theory and adjustments to bring it in line with phonological theory, or with linguistics in general. Linguistics is a statement of what a native speaker/hearer knows about language in general and about his language in particular. Syntax, for example, therefore characterizes classes of sentence — rather than produces individual one-off sentences; and phonology characterizes classes of abstract phonetic shapes. Formatted along these same lines phonetics would be a statement of what a speaker knows of the implementation of phonological intentions and as such would characterize classes of articulatory or acoustic outputs — rather than describe the wave shape or articulatory configuration of any particular utterance.

With such compatibility phonetics would more transparently relate to phonology. Phonetic explanation of certain phonological phenomena would be more meaningful and self-evident, and problems thrown up by phonological theory more readily investigated by phonetic methods.

There is another aspect to linguistic theory which I alluded to much earlier. Formally linguistics proceeds by successively constraining a too powerful device. Constraints are identified according to type and hierarchically imposed on the unconstrained device. This formal technique might usefully be adopted into phonetics and, in a sense, we have already begun to do so. We talk of mecanico-inertial constraints, of temporal constraints, of missed targets and the like. I do not see any difficulty in principle in formally organizing these ideas with a view to optimizing insights into how human beings work phonetically.

I began by quoting some relatively simple data at you, and I have ended by talking about metatheoretical problems, formalism and a phonetic theory that is as abstract as phonology — the way between the two may not have come out as clearly as I would have liked: I excuse myself by saying that I set myself a non-too-easy task. I do firmly believe, however, that linguistics and phonetics are not to be treated as distinct, but that they have a mutually revealing role to play. I further believe that currently phonetics is lamentable in its inability to capture significant generalizations about speech, and I hope that even if my tentative proposals prove valueless in themselves, they will nevertheless have indicated that a different approach from the one we have now is worth considering.