The Phonetic Component

Katherine Morton
Mark Tatham

Paper read at the spring 1970 meeting of the Linguistics Association of Great Britain, Manchester. Reproduced from *Occasional Papers* 8, Language Centre, University of Essex, December 1970.

This work was supported in part by grant B/SR/6733 from the Science Research Council.

The output of the phonological component of a transformational-generative grammar requires a phonetic specification which should be in line with the facts of speech production. This paper is concerned with what properly belongs in the phonology and what in the phonetics and the criteria underlying the decisions to account for one phenomenon in one component and another phenomenon in the other component.

Total abstraction seems to be on the wane as far as the phonological component goes and attempts are being made at the present time to incorporate a certain degree of naturalness into phonology — where 'naturalness' means in a loose definition 'that which seems to be a common occurrence in many languages', and, more rigorously, 'that which has underlying physiological or psychological motivation'. Thus, as Schane (1969) and others have shown, any consideration of the concept of naturalness in the formulation of phonological rules must take into account these two major recurrent underlying bases for a criterion of naturalness: physiological (or neuro-physiological) factors and psychological factors. The psychological category governs rules concerned with perceptual or psychological phenomena such as the tendency for maximum differentiation (where categories or functions of categories are polarised as in the maximal opposition between a voiced vowel segment and a voiceless consonantal segment), or certain syllable structure constraints (where segmental patterning in terms of consonantal and vocalic segments is established). We do not intend to treat the psychological aspect in this paper.

The other category — rules whose formulation is physically or physiologically motivated — is easier to handle, and we propose to discuss some of the criteria for establishing broad axial divisions for accounts of these phenomena: phonological/phonetic, competence / performance.

An important problem facing the natural phonologist is just where to stop: the level of systematic phonetics as the output of the phonological component has become even more vague than it was earlier, now that parts of the phonology are beginning to be based on physical criteria. We suggested earlier that systematic phonetics might be a level established where knowledge governing voluntary linguistic aspects of phonology gave place to involuntary phenomena, such as co-articulation (see Tatham 1969). This solution, however, is too simple for two reasons: firstly, it confuses competence with performance by implying (though not assuming) that the final stages of phonetics were the subjects only of performance, and secondly, although in phonetics there are involuntary phenomena they seldom occur in random, ungoverned ways. If it is the case that phonology should stop where the simple automatic non-linguistically governed phonetic rules take over, then possibly certain rules in the phonology need a clearer theoretical justification than has hitherto been the case. We are thinking here of some of the assimilation rules in particular.

Assimilation is a common phenomenon and it usually follows some easily predictable lines: for example, one might guess that a candidate for assimilation is a voiceless intervocalic obstruent; it very often in many languages becomes voiced. The phonologist, armed with a good understanding of the neuro-physiological facts of speech production, cannot fail to

establish assimilation rules as a good natural category: in the intervocalic voiceless obstruent example there is a great articulatory tendency for voicing to continue through the stop. But what is it that the phonologist should be interested in?

Assimilation is very common, as we have pointed out and is by no means consistently operational: the same speaker assimilates this or that amount on different occasions. Assimilation in the phonology needs to be systematic and shown to be different from casual phonetic assimilation.

As we have argued before it is not the assimilation which is linguistically relevant, but the *limits* to which assimilation is permitted. It can be contended, though, that any assimilation rule in phonology is *trivial on its own* — that is, unless hierarchies in the natural rules are established accounting for degrees of assimilation and preferred assimilatory environments: unless this is done, nothing *phonological* has really been captured. We are arguing against a phonology which restates phonetic trivia.

For example, there is probably going to be some devoicing in a consonant in absolute final position: this may or may not be a *phonetic* fact of little interest to phonology, It has phonological relevance in German, where there is total neutralisation of the voiced/voiceless distinction in final position, compared with English where this is not the case. Schane, noting the neutralisation and apparent optimisation of the final consonant in German has pointed out that in this case we cannot simply regard the devoicing as a case of assimilation (that is, assimilation with the following silence) and cannot even treat it is a physiologically motivated rule. This rule for German is psychologically motivated to establish maximum differentiation from the optimal vowel. In English however there is partial devoicing — but this is nothing more than phonetic — and neutralisation does not occur (hence such pairs as bad/bat). In German therefore we properly place the devoicing rule in the phonology; in English in the phonetics.

Let us now take a look at some of the facts of coarticulation and reduction — factors responsible for assimilation at the phonetic level, which may also provide the physiological motivation for a rule in the phonology. During articulation, from time to time a steady state can be noted in the movement of the speech organs; but this is comparatively rare at normal rates of utterance. A mass, say the lips or tongue, must assume one position or configuration at one moment in time, a different one a very short time — milliseconds — later and yet a third a short time later, and so on — including occasions when the configuration of the particular articulator in question is irrelevant and it could be doing anything.

In practice these target positions, as they are called, are seldom, if ever, achieved, no matter what the speaker's intention, because there just isn't the time available for the desired movement. Mechanical inertia and limitations in the neural system prevent a succession of fixed steady-states with only short transitions; very often the transitions are longer than anything that could be called steady-state, since so much integration or smoothing has taken place. In other words no matter what the form of the motor-commands issued to the muscles controlling the articulators, we don't achieve segmental speech or hit targets corresponding to the positions taken up by the articulators if any one segment is uttered in isolation. BUT, we could if we wanted to — or rather we could do better than we do.

Under normal circumstances, of course, it isn't critical that a target should be reached. Even though a great many segments become blurred, and even though target are missed, there is still sufficient information there for ambiguity to be avoided — indeed the very manner of the integration or transitions may itself provide perceptual cues. Whether disambiguation or decoding is performed by a motor-perception strategy, whether the listener recruits syntactic and semantic knowledge for this purpose is neither here for there for phonetic theory. If disambiguation or simple de-coding can be performed (as in, say, nonsense words) without recourse to semantic or syntactic knowledge, then this fact must be accounted for.

Now, although as speakers we may or may not succumb to a particular co-articulation effect, there remains the question of how it is that not all co-articulation tendencies are given in to equally. It is a fact that some languages permit a greater degree of co-articulation in a particular area than others. That we need not wholly permit a particular co-articulation effect to take place presupposes knowledge of two kinds: firstly of just when the effect will take place — that is, when in both time and segmental context; we can call this contextual or environmental knowledge or the environmental condition. Secondly, knowledge of just what the effect will be (that is, knowledge of the mechanics of the effect — which might also involve knowledge of neural limitations in addition t the mechanical limitations).

We are not talking here of any decision on the part o the speaker to speak 'more deliberately' at any one time, but of problems associated with what might be called 'phoneme overlap' — that is, overlap of articulatory space that any sample of allophones might occupy. Let us consider the theory that speech is a broad vocalic gesture on which are superimposed short consonantal gestures. Imagine, though, a stretch of speech consisting only of vowels or vowel-like segments: an example might be the English 'how are you', in which there are no stop or fricative consonants during the utterance and voicing continues throughout. To make this a bit clearer, suppose we have a low back vowel [a] followed by a high front vowel [i], followed immediately by another low back vowel [a]: that is, [i] sandwiched between two [a]'s. Let us further suppose that this is a language with only four vowel segments to differentiate morphemes: /i/, /æ/, /o./ and /u/.

Now, if we further add a temporal constraint that the central segment [i] is comparatively short (perhaps because it is unstressed), then given only a short time for moving the mass of the tongue for this segment we would expect the target to be missed. At the point of reversal of direction of movement of the tongue from forward to backward (that is, at the point where the attempt to hit [i] must be abandoned for a return to [o.]) an articulatory position approximating to [t] or [e] might be the result.

Given the principle of maximum differentiation which predicts that in a four-vowel system there will be little possibility of perceptual confusion from missed targets, we expect that [L] or [e] will be satisfactory substitutes for [i]. In other words if the number of members in the set of vowels is small then, if the principle of maximum differentiation — a perceptual principle — has applied, a great deal of freedom can be permitted in the articulation of any of the members of the set before the system breaks down.

But now suppose a language with twelve vowel-segments differentiating morphemes, and further suppose that members of this set include [i], [I] and [e]. Now there is a problem. Clearly the phenomenon of phoneme overlap is more likely to occur than in the previous example. We would predict that on this occasion the permitted area of missing the [i] target will be much smaller. There is target-missing according to rule, but the amount by which the rule has been effective is less. In other words, *knowing* that the target will be missed and *knowing* that this will result in incorrect acoustic encoding of the segment and therefore perceptual error, the speaker attempts to do something about it.

There are therefore two distinct limits to be set on any co-articulation rules: mechanical or similar limits expressing just what can or can't be done with particular articulators (or better muscles) in a particular time- governed environment, and also linguistic or psychological or perceptual limits, expressing just what any particular language will tolerate in the way of missed targets before confusion arises. Of necessity the linguistic limits must lie within the perceptual limits: there is no language we know of which expects the physiologically impossible to solve a linguistic segmental differentiation problem.

Now, these limiting factors must be known in some sense to the speaker and while programming his muscles for speech he must give them values. The knowledge exists and must be accounted for; that we do not invent new words demanding mechanical impossibilities is an important fact to be stated formally somewhere; that we do not tax

perception by missing targets willy-nilly, even when there is a tendency to do so, is a further important fact to be captured. Notice again that the intrinsic nature of the facts is different.

If we apply the blanket term ‘competence’ to these two sets of rules, then we are faced with a problem of definition. Competence is about knowledge of linguistic facts — not of mechanical ones. Yet clearly the two are interrelated: one presupposes the other; the limits of one are contained by the limits of the other, and so on. One of the requirements of a model of linguistic competence is that it should generate only grammatically correct forms, but we must be careful when using this requirement in phonetics. Not only must the phonetics be grammatical — that is, obey rules governing both universal and language-specific rules about forms and their sequencing — but it must WORK — that is, be capable of articulation. The units and categories for statements of phonetic competence are, of course, crucial in a way that perhaps they are not at higher levels. It is for this reason (namely, that regard should be paid to the real world as physical reality is approached) that physiological and psychological considerations are being taken into account in phonological theory. Although the mechanical facts of speaking may not be in the province of linguistics, linguistics must account for the employment of knowledge about the *limits* of the mechanical constraints.

References

Schane, S. (1969) Natural Phonology. Paper read at the spring 1969 meeting in York of the Linguistics Association of Great Britain

Tatham, M. (1969) Classifying Allophones. Essex University Language Centre, *Occasional Papers* 3