Defining the Bases of Phonetic Theory

Mark Tatham

Paper read at the summer 1970 meeting of the Linguistic Society of America, The Ohio State University.
Reproduced from *Occasional Papers* 8, Language Centre, University of Essex, 1970.

The task of any phonetic theory is to determine the form of a phonetic component by establishing the internal and external constraints on that component. The phonetic component itself converts linguistic knowledge of the structure of the speech act into time-varying commands suitable for control of the articulatory mechanism. Performing involves knowledge, and this knowledge must be expressed in a form accessible to the speaker operating in time. Knowing how to use knowledge of performance constraints involves manipulation of the conversion from segmental notional time embodied in simple sequencing to timing of muscular control. A solution to the handling of this time conversion is discussed.

The task of any phonetic theory is to determine the form of a phonetic component for a grammar. The function of the theory is to relate linguistic descriptions with the facts of speech (Ladefoged 1965): to do so it must be expressed in the simplest, most explicit form possible and in a way which enables transparency of this relationship no matter whether the theory is approached from the phonological angle or the articulatory/acoustic angle. A statement of the theory in this form enables testing to take place — a prerequisite of any model or theory building operation Fromkin (1968).

Providing an adequate theory of phonetics is proving extremely difficult, and the question might well be asked: why is this the case? Certainly inroads have been made in the area of phonetic description from the classical approaches of the late l9c and the early 20c to the exciting developments of today when we may be finally cracking the problem of the motor control of speaking. Inroads have likewise been made into phonological and syntactic theory, but it seems, even to those engaged in the development of phonetic theory, that in these areas the recent contributions have been somehow more productive.

The principal difficulty lies in the form of the projected phonetic theory itself and the extreme opposing nature of the input and output constraints which must be applied to the resulting model. The theory has as its function the relating of linguistic descriptions with the facts of speech and it is patently obvious that linguistic descriptions with respect to their abstraction in formulation are by and large incompatible with the facts of speech. The solution to the problem of establishing phonetic theory hinges on the breaking of the incompatibility.

Linguistic descriptions are of course highly abstract even at the phonological level. Explicit input/output relationships are set up to account for data, the selection of which is constrained by decisions as to the domain of linguistic theory and more specifically the domain of any particular component of the grammar. Notice that we could put abstract syntax and phonology of the kind we have now into the same undesirable position of phonetic theory by requiring that it relate itself directly and explicitly to actual observed neural functioning. This demand is not made because the most basic constraint on the form of this side-stepping procedure — namely an empirical model of brain function in language — is lacking (but see Whitaker 1971 forthcoming); or because, as linguists, most of us don’t know enough about it anyway. One or two attempts have been made to set up syntactic or phonological descriptions using the types of operations (or form of rules) known or assumed to be typical of brain processes (Reich 1968), but these, though possibly adequate for some abstract linguistics, fall far short of satisfying the present demand — the demand that the facts of linguistics be related to the facts of human beings operating linguistic behaviour.

Phonetics is the centre of focus because we can see in principle ways of relating sounds or articulations (existing in the real world) to the abstractions of phonology. Some researchers have provided more or less rigorous algorithms for example for deriving a particular sound segment from a particular phonological segment with the usual environmental constraints, and so on (Halle 1959a). They have also had a measure of success relating abstract distinctive features with distinctive features of articulation or soundwaves (Fant 1967; Chomsky and Halle 1968) — hardly surprising if we remember that historically the distinctive features were worked out that way (Jakobson *et al.* 1951; Chomsky and Halle 1968).

We can go even further than this. The phonetic component itself converts linguistic knowledge of the structure of the speech act into time-varying commands suitable for the control of the articulatory musculature. It then relates the resulting articulations which are accessible to instrumental investigation to soundwaves which are also accessible to instrumental investigation. Recent developments in descriptive phonetics have resulted in the formulation of models capable of doing this: the input to these speech production models is considered as the output of a suitable phonology, where that output consists of a string of segments that possess no time other than the notional time associated with the simple linear sequencing of segments (Tatham 1970a). By utilising discoveries (Kozhevnikov *et al.* 1965; Fromkin 1968; MacNeilage 1968; Tatham 1969; Ohala 1970; Lehiste 1970) which indicate that the intuitively felt syllabic structure of speech is a function of the mechanism of speaking (i.e. innate) rather than of a higher-level requirement in, say, the phonology, a true time dimension can be added to the concatenated segments to simulate in a more or less adequate way the temporal arrangement of those segments in the neural control of the vocal tract to produce speech (Tatham 1970a).

The accepting, though, of this highly abstract input derived from present-day phonologies which have not even yet attempted with any measurable success to constrain themselves with neurological considerations is itself highly dubious. It is not the business of phonology to concern itself with neural processes — at least it is not in the discipline we understand as phonology at the present time. Phonology is concerned with identifying, describing and accounting for the sound patterns of language or languages (Halle 1959b): it does this in an explicit and explanatory fashion. It is not and should not be involved in at present inaccessible considerations of brain function which might lead to wild speculation. Phonetic theory *is*, on the other hand, highly involved in these considerations — if you take them away then you have no phonetics, except in a really crude and theoretically non-productive way.

Present models of speech production, whether they have been derived from work in understanding the human process (MacNeilage 1968; Wickelgren 1969) or from work in trying to make and operate speech synthesisers (Kelly *et al.* 1961), all share one property: they are properly generative Holmes *et al.* 1964; Tatham 1970b). That is, they assume that from a comparatively small inventory of items and rules an infinite or very large number of utterances can be produced: no proper phonetic theory would now assume the storage of complete utterances. Generally these items are listed and indexed, in a way analogous to the theoretical justification behind similar strategies in the syntax.

These lookup tables, as they are called, are static in nature as are the rules of syntax, and as such embody, theoretically at least, the speaker's knowledge of the phonetic (rather than phonological) pattern of language and/or his language. They embody one extra dimension — the dimension that I have been arguing is not present in syntax or phonology — namely, information or knowledge of neural and neuro-muscular mechanisms and functions. I have pointed out recently (Tatham 1970c) that hitherto these two dimensions — the one accounting for the phonetic patterns derived from linguistic considerations, and the other accounting for the external a-linguistic constraints — have been subject to confusion. A system of composite rules of the kind sometimes proposed (Ohman 1967a) merely obscures the important interplay between the two dimensions which can be understood to express the use the linguistic system makes of the available speaking mechanism. The crudest example I can think of is that it cannot be the case that any language would or could employ more sounds than the human vocal mechanism is capable of making — a statement which seems so obvious, yet a principle which has not yet been adequately accounted for in phonetic theory.

The best way I can elaborate on the constraints which might underlie phonetic theory is to discuss a specific section of a typical speech production model, in this case my own (so that I do not run the risk of misinterpreting anyone) (Tatham 1970a).

It is not necessary for the construction of a model of speech production for the input to be temporally indexed. That is, relative timing of segments and timing within segments can be established within the speech production model itself as part of the mechanism dominated by the sheer physical requirements of setting up and organising motor-commands to the musculature responsible for moving the articulators.

A psychological reality to the sequencing of segments is all that need be posited. Recent observational and descriptive studies in phonetics using techniques of electrophysiological analysis (MacNeilage and Declerk 1968; Tatham and Morton 1968) are revealing that in, for example, C[onsonant]V[owel]C[onsonant] monosyllables there is a programming or control cohesion between the initial C and the V of such utterances. By this I mean that analysis indicates that neuro-muscular control for the C and the V are not completely independent at the highest level of the motor system: that is, the C and the V exhibit interdependent properties which defy explanation in terms of what we know of lower level reflex feedback loops and similar mechanisms. The actual motor command for each segment could be viewed as context sensitive (Wickelgren 1969; but see MacNeilage 1970, MacKay 1970, Whitaker 1970); alternatively we could assume that in terms of motor control this initial C and the following V constitute in some sense a motor control *unit* exhibiting many of the properties of those individual segments, yet at the same time possessing properties dictated by their mutual context (Ohman 1967b; Tatham 1969).

Furthermore, other studies (Slis 1968; Lehiste 1970) indicate that in cases of strain on the overall rate of utterance of a CVC monosyllable there is a compensatory effect in time between the V and the *final* C, as though an effort were being made to maintain the length of the complete utterance — the CVC. This temporal compensation is much less apparent between the first two segments, at least as observed in data from English (but cf. Kozhevnikov *et al.* 1965, where temporal compensation was inferred to be between the first two segments in Russian).

Knowledge of typical motor programs for segments in isolation coupled with knowledge of typical durations for those individual segments can easily be integrated, at least in theory, with the principle of cohesion at the motor level between the initial and medial segments and with the principle of compensation at the temporal level between the medial and final segments, to produce, within the desired overall time for the complete CVC group, a motor program which would result in an articulation consistent with the observed data. In other words, interrelating the way in which the motor control of speech seems to operate — that is syllabically in terms of CV plus an optional C — with the temporal compensation effects which occur seemingly to maintain rate in utterances, can enable us to add a time dimension by rule to a string of input segments not phonetically context related. It furthermore enables us to predict motor programming effects other than durational ones.

Such tables and rules have not yet been worked out: the principle appears valid however. What I want to make clear is that a highly abstract input expressed in the form of segments solely derived from morpheme structure considerations together with a few idiosyncrasies (like the distribution of clear and dark /l/ in English) can be interrelated with a model based on posited mechanisms in the *actual* or real workings of the human being, to generate a time varying speech output.

There are other parts of the current speech production model which could be cited as examples. They all exhibit the property of positing a strategy for the correct use of lookup tables. The strategy is triggered by the segment-sequencing required as a result of linguistic operations at some higher level and it results in the manipulation of static lookup tables whose function is two-fold: the storage of information concerning the properties of the vocal mechanism, together with the storage of information concerning the linguistic demands or strain to be put on that mechanism.

The facts of the acoustics of speech and of the neuro-muscular system employed to produce articulatory configurations resulting in that acoustics can be viewed as autonomous, and used in the production of autonomous neuro-muscular and acoustic theories. Such theories do not possess the property, though, that their simple integration or combination leads automatically to a general theory relating linguistic descriptions with those facts of speech. A theory of the kind I have been describing, however, does do just that, and seems capable of development to indicate such a relationship throughout.

References

Chomsky, N. and M. Halle (1968) *Sound Pattern of English*. Harper and Row, New York

Fant, G. (1967) Sound, Features, and Perception. *STL-QPSR*-2-3/1967 RIT, Stockholm

Fromkin, Victoria (1968) Speculations on Performance Models, *Journal of Linguistics* 4

Halle, M. (1959a) *Sound Pattern of Russian*. Mouton, The Hague

Halle, M. (1959b) Questions in Phonology. *Nuovo Cimento* 13

Jakobson, R., G. Fant and M. Halle (1951) *Preliminaries to Speech Analysis*. MIT Press, Cambridge, Mass.

Holmes, J. N., I. G. Mattingly, and J.N. Shearme (1964) Speech Synthesis by Rule. *Language and Speech* 7

Kelly, J. L. and L. J. Gerstman (1961) An Artificial Talker Driven from a Phonetic Input. *JASA* 33

Kozhevnikov, V and L. Chistovich (1965) *Speech: Articulation and Perception*. Joint Publications Research Service 30.543, Washington

Ladefoged, P. (1965) The Nature of General Phonetic Theories, *Georgetown University Monograph* 18: Languages and Linguistics

Lehiste, Ilse (1970) Temporal Organization of Spoken Language. *Working Papers in Linguistics* 4. Computer and Information Science Research Center, Ohio State University

MacKay, D. G. (1970) Spoonerisms: the Structure of Errors in the Serial Order of Speech. *Neuropsychologia* 8

MacNeilage, P. F. (1968) The Serial Ordering of Speech Sounds. *Project on Linguistic Analysis Reports* Series 2, 8:Ml-M52, Berkeley

MacNeilage, P.F. (1970) Motor Control of Serial Ordering of Speech. *Psychological Review* 77.3

MacNeilage, P. F. and J. L. Declerk (1968) On the Motor Control of Co-articulation in CVC Monosyllables. Haskins Labs: SR-12, New York

Ohala, J. (1970) Aspects of the Control and Production of Speech. *Working Papers in Phonetics* 15 UCLA Los Angeles

Ohman, S. E. G. (1967a) Numerical Model of Co-articulation. *JASA* 41.2

Ohman, S. E. G. (1967b) Peripheral Motor Commands in Labial Articulation. *STL-QPSR*-4/1967 RIT, Stockholm

Reich, P. A. (1968) Competence, Performance and Relational Networks. *Report CPN: Linguistic Automation Project, Yale University*, New Haven

Slis, I. H. (1968) Experiments on Consonant Duration Related to the Time Structure of Isolated Words. *IPO Annual Progress Report* 3.71-80, Institute for Perception Research, Eindhoven, Holland

Tatham, M. A. A. (1969) The Control of Muscles in Speech. Essex University, Language Centre *Occasional Papers* 3, Colchester

Tatham, M. A. A. (1970a) Articulatory Speech Synthesis by Rule: Implementation of a Theory of Speech Production. *Working Papers in Linguists* 6, Computer and Information Science Research Center, Ohio State University

Tatham, M. A. A. (1970b) Speech Synthesis — a Critical Review of the State of the Art. *International Journal of Man-Machine Studies* 2

Tatham, M. A. A. (1970c) Co-articulation and Phonetic Competence (abstract). *JASA* July 1970; also Essex University, Language Centre, *Occasional Papers* 8

Tatham, M. A. A. and Katherine Morton (1968) Further EMG Data towards a Model of Speech Production. Essex University, Language Centre, *Occasional Papers* 1

Wickelgren, W. A. (1969) Context-sensitive Coding, Associative Memory, and Serial Order in (Speech) Behavior. *Psychological Review* 76.1

Whitaker, H. A. (1970) Some Constraints on Speech Production Models. University of Essex, Language Centre *Occasional Papers* (forthcoming)

Whitaker, H. A. (1971) A Model for Neurolinguistics. University of Essex, Language Centre, *Occasional Papers* 10 (forthcoming)