THE ROLE OF SPEECH SYNTHESIS
IN THE DEVELOPMENT OF PHONETIC THEORY

MARCEL A.A. TATHAM

Building an adequate phonetic theory is a task difficult enough in itself, but yet more difficult when faced with the need to incorporate or even, at worst, reconcile the abstractions of linguistic theory in general and the more concrete notions of acoustic and neuro-physiological theory. Clearly most progress on any complete model of speech production as of language in general (Chomsky 1968) derives from an abstract-to-concrete direction — rather than the opposite — although, of course, all theories are constructs incorporating a measure of verifiable fact.

It is clear that would-be builders of speech production models, whatever they say, do not help but provide a central key for their task from the notion that they are 'taking over' in some linear sense at a particular stage in what, if I may be forgiven, is tacitly acknowledged to be a somewhat grander task — the development of a theory of language in general.

Now, a large amount of the data forming the basis of phonological theory derives from phonetics — but notice that it seldom derives from the theory of speech production but from some kind of informal survey of the data of phonetics. Let me suggest that phonology would be the better off if it proceeded from phonetic theory rather than reorganised phonetic data. Possibly the use of synthetic speech could help in phonological research (Mattingly 1970), but it may be said that its use in serious *phonetic* research is limited by the minimal specification of the sound output. It is arguable however (Tatham 1970a) that this is the least important consideration and that the most benefit to be derived comes from the construction of the control system. There are then two things, at least, wrong with any claim that speech synthesis is a useful tool. The first is that if I am clever enough to write an algorithm that will be able to produce an output of convincing running speech, there is no guarantee that I can take this to be a suitable model of speech production; and the second is that if I have a model for speech production which is complete enough to implement in speech synthesis, then I don't need to waste time implementing it — it stands on paper. It is naive to suggest that it would be 'nice' to actually make a machine run through the logic — I hope we have better things to do. It is equally naive to suggest that the speech output from the synthesiser will be that much the better in

quality for having been produced from this or that particular program, where the programs simply replicate alternative models.

As an aside, let me stress that I am talking about building a speech PRODUCTION model. There are quite different reasons for a fact ADVOCATING well-derived synthetic speech for perceptual experiments — witness among others the work of Haggard (1969 seq.).

It happens that phonetics researchers regularly seek a *raison d'être* for their work in attempting to throw light on the abstractions of linguistics. Thus we have a by now large quantity of literature describing, for example, progress in the quest for the physical correlates of this or that abstract phonological unit. But when it is argued that linguistics could proceed best from abstract to concrete, this cannot be simply what is meant. It is perhaps this attitude that has provided one of the bases of the predisposition for regarding phonetics as interpretive of phonology. Notwithstanding the obvious sub-theories of say, acoustics of speech, or neuro-muscular control of articulation, the role of a production model has been seen as passive, amounting often to no more than answering the question: how do we get this or that observed sound from this or that abstract phonological unit?

The theory of phonetics, though, is not the stringing together of these sub-theories which themselves are only taken over from much larger disciplines. It is THE THEORY OF PHONETICS, embodying the power not only to stand in its own right but supplying of necessity predictions as to expectations in other areas of linguistics and explaining phenomena which might just otherwise be noted in the form of systematised facts.

Let me give two examples aside — the data here being derived from electromyographic experiments. It is well known that linguists have usually not felt the need to note contextual variations in the articulation of running segments. One theory of speech production (Tatham 1970b) rather than simply noting that phonology does not need to list these variations, explains how they are the result of peripheral mechanisms and so not part of the high-level generating of language at all, and, contrary to the suggestion of at least one researcher (Wickelgren 1969), not to be accounted for in any treatment of phonology (Tatham and Morton 1970). Another example might be the handling of the geminate consonants which occur in some languages. Preliminary electromyographic studies of Finnish (Tatham and Morton, forthcoming) indicate that there is probably not a simple overlapping repetition of a single consonant motor program to produce the geminate version — although acoustically the geminate may well often have roughly twice the duration. Such data would establish within the theory of phonetics that motor control proceeds differently in each case and explains why identification in the phonology would be in error.

Synthetic speech studies have largely contributed to the interpretive notion of phonetics by proclaiming success in constructing programs that one way or another, while simulating the acoustic and articulatory sub-theories, achieve their aims by providing a more or less adequate sound output from the simplest possible version of a systematic phonetic input. Surely the goal should be the derivation of a more

adequate input to obtain the given and trivial output *via* what is known of the production process. There is not just no point in providing speech from a minimal input — except as the elaborate demonstrations we have had in recent years — positive harm can result from not attempting to supply explanations of facts outside the immediate domain of phonetics.

The aim of this paper has been to stress that phonetic theory should play less of an interpretive role and more of an explanatory or predictive role alongside phonological theory. I have suggested that contrary to what was once believed, synthetic speech studies as they are for the most part conducted today under the guise of helping build that theory, hinder rather than aid.

Language Centre
University of Essex
Colchester

REFERENCES

Chomsky, N.
    1968   *Language and Mind* (New York).
Haggard, M.
    1969-71   *Speech Synthesis and Perception* (= *Reports on Speech Synthesis and Speech Perception*) (Psychological Laboratory, Cambridge University).
Mattingly, I.
    1970   "Speech Synthesis for Phonetic and Phonological Models", in *Status Report on Speech Research, Haskins Laboratories SR-23*.
Tatham, M.A.A.
    1970a "Speech Synthesis, a Critical Review of the State of the Art", *International Journal Man-Machine Studies* 2.
    1970b "A Speech Production Model for Synthesis-by-Rule", *Working Papers in Linguistics* 6 (Computer and Information Sciences Research Center, The Ohio State University).
Tatham, M.A.A. and K. Morton
    1970   *Explaining Some Apparently Context-Sensitive Effects in Speech* (= *Occasional Papers* 9) (University of Essex, Language Centre).
    forthcoming   "/p/ and /pp/ in Finnish — Some Audio, Air-Pressure and Electromyographic Data".
Wickelgren, W.A.
    1969   "Context-sensitive Coding, Associative Memory, and Serial Order in (speech) Behavior", *Psychological Review* 76.1.

DISCUSSION

NOOTEBOOM (Eindhoven)
1. What are the consequences for your own work on speech synthesis?

2. Speech synthesis normally is taken to be the acoustical output of a speech synthesiser. It might also mean, however, the simulation of articulator movement by e.g., computer. This kind of synthesis could be very useful in studying compensatory

articulation, which in its turn could lead to more insight into the flexible ways our brain uses the vocal organs for realizing semi-constant perceptual goals. This could help us in understanding the relation between phonology and speech production. Would you reject this kind of synthesis also?

TATHAM

My object in Tatham 1970b was to outline a possible approach to speech synthesis-by-rule, employing the control program as much as possible of a speech production model, and I believe it is still possible to test certain parts of the model using synthesis methods. In the present paper, though, I am stressing the dangers inherent in such an approach and particularly in advocating the use of synthetic speech as an aid in constructing phonologies (cf. Mattingly 1970).