Speech Synthesis and Models of Speech Production - I

Mark Tatham
Katherine Morton

Reproduced from: *Interim Report* (July 1969 – September 1970), Science Research Council Award B/SR/6733 (July 1969 – March 1971).

CONTENTS

1. INTRODUCTION

This interim report on the Current Status of the present project is intended to support an application (submitted for 1st October 1970) for a major extension to run from April 1971 for a period of three years. We shall try to show that the line of investigation we are pursuing is productive and relevant to current problems in linguistics and indeed any work involving the understanding of speech.

Our terms of reference are different from those outlined in the original proposal considered by a Special Panel of the Council in April 1969, in that we decided, in view of the shorter period available (twenty-one months instead of three years) that it would be best not to attempt any practical experimentation in speech synthesis itself. Instead we have devoted our time (July 1969 to September 1970 is covered in this report) to developing mainly the theoretical aspects of a speech production model, bearing in mind that this was to be a working model suitable for later implementation as a control strategy for speech synthesis.

We believe — and this point will be developed below — that speech synthesis has reached a state of the art now when it can be relied on as a most suitable tool for experimental linguistics. By the latter term, we mean not only experimental work of interest to phoneticians, who are concerned with the production of speech, but also to linguists in general — especially psycholinguists concerned with the perception of speech — and engineers concerned with the problems of man-machine communication. Indeed it is significant that we have been able to contribute substantially to the thinking of a parallel project being conducted n the Department of Electrical Engineering Science and sponsored by the Ministry of Technology. The broadening of use of speech synthesis however seems on1y

of importance if it is regarded not as a clever trick of making a machine speak, but can incorporate a genuine speech production model specifically designed for the purpose and enabling artificial speech to be generated in a way seen to be analogous to that of a human being.

The present report is divided into several sections, the first of which is concerned with the history of the project — and our reasons for choosing the particular areas we have covered. Later sections deal with our work on the development of phonetic theory and in particular with the formulation of our speech production model. Possible implementation of the model is discussed and an outline of the remaining instrumental work to be conducted before the termination of the present project. An appendix is added which lists our publications relating to the project, together with abstracts of those papers.

2. HISTORY

The project was conceived following work which both team members had been involved in at the University of California at Los Angeles and in the Phonetics Department at Leeds University. We had been dissatisfied at the state of phonetic theory which seemed to us to be little more than an elaborate collection of data at the time. One or two researchers had formulated proper more or less formal models concerning specific links in the speech production chain, but very little work had been done on a complete model.

Accordingly we proposed to the Council during 1968 and 1969 a research project which would bring together the work in phonetics for an attempt at the formulation of a theory suitable for synthetic speech research to continue on a footing more theoretically sound than had hitherto been the case. We were to work on the model and to collect a certain amount of data which would serve a dual purpose: it would enable hypotheses arising during the model building process to be confirmed or refuted and at the same time provide us with useful data for the tables which necessarily term a part of any speech synthesis by rule program.

The Council, following a meeting of a Special Panel in April 1969, agreed to a grant of £11,000 which would enable the present team to investigate the groundwork of the proposed project for a period of twenty-one months. This period terminates in March 1971. Moreover, the grant would be insufficient to employ more personnel than a single research officer, and would not provide us with synthesiser hardware and online computing facilities in the laboratory which would be essential for the completion of all of the projected aims.

It was therefore decided that the project should be limited to those areas from the original proposal which could be tackled with the available resources. Practical speech synthesis was rejected immediately since, although it would have been possible for us to get together the hardware, we would not have had the engineering expertise for electronics design work. The online computing facility would not have been available either.

What remained, and what has proved reasonable ground to cover with the available financial resources and time, was the concentration of effort on the theoretical side of the research, together with a certain amount of experimental work to support the theory.

A further and ultimately extremely important task was for one member of the team to become thoroughly familiar with some of the computing techniques involved in the use of small computers for this kind of work. This has been accomplished with the aid of the Department of Electrical Engineering Science and the Computing Centre of the University. From October 1970 our own laboratory will have the most basic configuration of a Honeywell 316 which is ideally suited to our purposes. We have been able to process some of our experimental data by computer and the new installation (paid for by the University) will enable the team to have adequate computing expertise by the end of the project

The final terms of reference, therefore cover three areas:

- model-building;
- experimentation;
- gaining an expertise in computing.

These are discussed below and at this stage of the project only a. is adequately covered.

3. THEORY AND BUILDING THE SPEECH PRODUCTION MODEL

a. The Basis of Phonetic Theory

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 description 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).

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 largely 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. Phonetics is our centre of focus because we can see at least 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 (Jakobson *et al*. 1951; Chomsky and Halle 1968) with distinctive features of articulation or soundwaves (Fant 1967).

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 sound-waves which are also accessible to 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; McNeilage 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 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 haven't 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 ii, 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 neuromuscular 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.

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 electro-physiological analysis (MacNeilage *et al*. 1968; Tatham *et al*. 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 neuromuscular 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 1961); 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 (Sliss 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 for 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 final 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 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 Each 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 /1/ in English) can be inter-related 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 neuromuscular system employed to produce articulatory configurations resulting in that acoustics can be viewed as autonomous, and used in the production of autonomous neuromuscular 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 we are developing however does do just that and seems capable of indicating such a relationship throughout (see Tatham 1970a).*

> [**footnote:* This section has been adapted from a paper: 'Defining the Bases of Phonetic Theory' read at the July 1970 meeting of the Linguistics Society of America, and appearing in University of Essex Language Centre *Occasional Papers* No. 8, November 1970.]

b. Acoustic *vs.* Articulatory Model

The question arose whether we should devote a lot of effort to the construction of m model having the acoustics of speech as its main concern. The decision was taken against acoustics in favour of an articulatory based model because the acoustic theory as it stands at present is very well developed (Fant 1960; etc.). The model we have been able to construct is therefore primarily intended to throw light on articulation, but specifically on the neuromuscular processes involved in the linguistic control of speaking. Linguistic control, because, of course, we cannot be sure that the neuromuscular mechanism functions similarly when speech is not involved (as with swallowing or sucking). There is evidence from dichotic listening experiments, for example, that indicates that human perceptual Strategies differ for speech and non-speech sounds. Furthermore a model which accounts for the observed results of brain damage distinguishes linguistic and non-linguistic control of neuromuscular processes (Whitaker 1969). Therefore we decided upon this qualification to our model.

The question of an acoustic model vs. an articulatory model is further important when the use to which the model is to be put is considered. Most of the work to date in speech synthesis (see Tatham 1970b) has been concerned with the control of acoustic analog synthesisers ('terminal analogs'. 'formant synthesisers' or 'resonance-synthesisers'), and the work has centred around target values and transition rules for synthesis of the acoustic waveform. We decided at an early stage that our model would enable the control of a synthesiser to be based on the control of articulation. This approach has lately been adopted with some success by Haggard at the Experimental Psychology Laboratories in Cambridge (Werner and Haggard 1969).

c. The Motor Control of Speech

For the purposes of the present theory we have assumed that all or the majority of the muscles of speech contain a system of feedback known as the gamma loop system. This means that muscles contain particular fibres (the muscle spindles) which have the property of signalling the rate at which they are being stretched — thus providing running information on the contractile state of the muscle in question. Such an assumption is very convenient for our purposes, but although a number of neuro-physiologists support the existence of this mechanism in the muscles involved in speech, there are a number who do not.

The existence of muscle spindles has been known for a long time, but only comparatively recently (see Tatham 1969, and Hardcastle 1970) has an adequate model been formulated in neuro-physiology which could be assumed for the purposes of phonetic theory. Indeed, one of our constant problems in establishing a theory of speech production has been the current controversies in neuro-physiology — even to the extent that recently (Partridge and Huber 1967) the relationship between electromyography (one of the modern phonetician's basic experimental tools) and movement has been questioned. A brief resume of the possible role of the gamma loop system is set out in Tatham 1969.

The model as it is at present formulated assumes that there is a one-to-one correlation between linguistic units of phonemic size and commands to the muscles responsible for moving the articulators. Such an assumption was itself the basis of a controversy between 1965 and 1969 in the phonetic literature. It was the case that researchers could not agree as to the existence or significance of minute but consistent variations in the electromyographic signal associated with a particular consonant in varying vowel environments.

The current theory supports the view that such variations do exist, and following Ohman (1967b) and MacNeilage (1968) points out that variations in EMG can be seen as a one-to-one variation in muscular innervation which is directly the result of two effects lower in level than that stage of the processing where motor commands are computed. The effect, it is hypothesised is due either to the gamma loop feedback circuit having the effect of changing the innervatory signal (and therefore the degree of contraction, hence the different EMG) at the periphery, or to the fact that two types of innervatory nerve fibres exist: alpha and gamma fibres. Both types can be regarded as receiving a control impulse simultaneously, but because the gamma fibres transmit the impulse faster (although they are less in diameter than the alpha fibres they incorporate less inhibition) they have the effect of adding to the innervation associated with the previous segment. Theoretically the gamma system can be used to account for two distinct effects observed In EMG variations: (a) right-to-left effects — gamma loop feedback; (B) left-to-right effects — gamma-fibre fast transmission of innervatory impulse The gamma system is further incorporated in to the control model because it has the property of being 'set' to a required level — this ascertains that a muscle will contract (and therefore an articulator move) to the required position and hold that position. The basis for the muscle control portion of the model is fully set out in Tatham 1969 — 'The Control of Muscles in Speech'.

d. The Nature of the Input

As mentioned above under a., it is possible to assume an input for the model which is similar to the segmental output of the phonology known as 'systematic phonetics'. This level, however, is under question at the moment, since it does not solve the phonetically motivated question of the relationship between intrinsic and extrinsic allophones (ace Tatham 1969 — 'Classifying Allophones' in *Occasional Papers* No. 3) The problem can be quite simply stated: there are peripheral variations (i.e. articulatory and acoustic) where linguistics assumes identity of segments — that is, the phonemes which are assumed in the morphology turn Out to have variants at the periphery. Some of these variants can be easily established In the phonological component (such as the distribution of palatal and velar /1/ in English); others are easily established in the phonetic component (such as the distribution of fronted or retracted [k] dependent on the front/back feature of a following vowel). What is difficult to account for is the fact that although the phonetic component variations would normally be

held to be involuntary it nevertheless is the case that non-involuntary variations can be detected at this level. This caused some researchers (for example Ladefoged in his 1967 *Linguistic Phonetics*) to assume that all allophones were directly attributable to linguistic control. This is an unsatisfactory explanation.

We have been able to suggest however that a model which assumes both the phonological variations (linguistic and voluntary) and the phonetic variations (a-linguistic and involuntary) and posits a third parameter of extrinsic control over the intrinsic phonetic variations provides a more explicit explanation. Such a control has its base in the phonology (since it is dependent on phoneme inventory and perceptual crowing of features — see Tatham 1970a) , but operates in the phonetics. The introduction of this concept in the model considerably clarifies the problem of allophonic variation., and at the same time explains the fact that phonetic (intrinsic) variation of any particular articulatory target varies from language to language. We are able therefore to retain the original universal concept of rule-governed variations dependent on the articulatory mechanism (assumed to be similar for all speakers of all languages), yet explain language dependent sub-variations.

The input to the model therefore should properly consist of a string of extrinsic (totally phonologically determined) target allophones. At the same time input information a required for limiting linguistically the later intrinsic allophonic variations.

The problem of the introduction of time into the phonetic model constitutes a further input problem. Hitherto all models (including and especially those working models for speech synthesis) have assumed that the extrinsic allophonic segmental input is not time indexed. In speech synthesis by rule programs time is generally added by rule modification of values obtained from a lookup table which expresses the target (or normal) duration for individual segments. At the present time it is highly debatable whether this is a satisfactory procedure. We have chosen not to use this method of adding time (see next Section).

e. Adding Time

The adding of time to a simple sequential (itself notional time) input constitutes one of the most difficult problems in the construction of a speech production model. As mentioned above, models forming the basis of synthesis control usually incorporate a lookup table of typical durations associated with particular phonemic segments. These are then modified by rules according to context. Thus for example, an intrinsically short vowel such as [a] will be lengthened when it occurs immediately preceding a voiced consonant.

A simple model such as this will generate only speech having always the same time parameter values none of the temporal variations observed in human speech will be produced, let alone explained (— explanation is of course our primary aim). Variations in the timing of segments in like contexts are not random (Kozhevnikov *et al*. 1965; Sliss 1968; Lehiste 1970.

The task is even grater when we consider that a speaker may decide to speak slowly or fast over a range which is continuously variable. An utterance lengthening of, say, 20% in overall duration (for speaking slowly) never results in a 20% increase in the duration of all segments. The present working model would be forced to proceed in this way.

We hold that the incorporation in the model of typical durational values for segments is a position that can be improved by the rule governed temporal linking of these segments according to syllabic constraints (interpreting the data of Sliss and Lehiste) MacNeilage has established that there is a cohesion between the initial C and the V of a CVC monosyllable and our own data supports this (Tatham and Morton 1968b) . Apart from resultant qualitative cohesion in the obtained target variations there is also a durational cohesion. Durational compensation of the V and the final C have been observed by Sliss and Lehiste (at least for Dutch and English) . It is quite possible that rules can be established which are context dependent for providing durational variation which have more explanatory power than the traditional rules if they transparently reflect that this durational variation is not solely dependent upon segment-type context, but upon *syllabic positional context*. Full formalisation of this position has not yet been accomplished, but in principle it should provide the ability to

generate correct variations of relative segmental durations dependent upon an input constraint along the dine of slow/fast.

f. The Construction of the Syllable

Our work on the addition of time within the model has forced us to incorporate the notion of syllable. Researchers are pretty well agreed at the present time that speech is syllabic in nature. We see no reason to postulate this as a linguistic phenomenon at the phonological level — indeed the phonology is more satisfactory if syllables are entirely neglected. We postulate that the syllabic cohesion of individually identifiable extrinsic allophones (Tatham and Morton 1968b), observable in EMG studies (MacNeilage 1968) and audio studies (Lehiste 1970) is an involuntary (and therefore a-linguistic and phonetic) phenomenon associated with the manner of function of the motor control of speech (see Tatham 1970a).

4. THE PROJECT'S SPECIFIC CONTRIBUTION TO THE FIELD

Our specific contribution has, during this first year, been mainly theoretical, and this is reflected in the progression of ideas clearly visible in our published work (see **Appendix A — Summaries of Publications**). Briefly we have contributed the following theoretical points.

i. Theoretical

Establishment of a complete model of speech production especially designed for implementation in speech synthesis (which is seen as a test situation for the model) . We have incorporated all current theory on the neuromuscular production of speech.

For the first time a model has been produced which is explicitly linguistic in origin — thus contributing to the ultimate goal of current linguistic theory — an explanatorily adequate performance model.

In detail we have provided a theoretical basis for the incorporation of specific feedback mechanisms into the complete model (earlier researchers had dolt with parts of the model).

We have provided a theoretical solution to the problem of treating these inertia based allophones (intrinsic allophones) which nevertheless exhibit higher level (i.e. phonological and/or perceptually determined) control. Treatment of this problem had hitherto been confused.

We have incorporated recent empirical data showing Consistent inter-syllable segmental timing compensation into the model in an attempt to produce syllabic units automatically. This has been coupled with...

The use of notions of initial syllabic segment motor-cohesion to underline the theoretical standpoint that...

Syllabic grouping of segments is a reality, but comes about. as a property of the meter-encoding mechanism. That is, we have postulated that syllabic grouping is an innate property of tic motor cortex. Notice that this standpoint has been arrived from the point of view of theory- and model-construction based on linguistic and not neuro-physiological evidence.

ii. Practical

We expect that our final experimental work in the remaining months of B/SR/6733 will contribute empirical data on:

- the relationship between EMG and movement,
- the problem of the motor control of plosives (specifically an answer to the question: is the timing of the plosion related to a pressure threshold or to a syllabically-based segment duration constraint (we expect the latter)),
- the timing compensation now known to exist between the final segments of a CVC syllable when that syllable comes under constraints imposed by overall utterance rate variation — this will enable the model to predict local duration variations to achieve synthesis-by-rule of fast, medium and slow speech,

- timing of segments within the syllable based on physiological parameters (EMG, air pressure, airflow) where hitherto this data has been available only from acoustic experiments,
- the variation of the EMG signal for particular segments which is constrained by the segmental context. Such data is already available in limited quantities, but more is needed for the model to be made complete. This data will enable us to write rules which reflect the extrinsic control of intrinsic allophones.

A further practical contribution which we consider to be of importance has been the development of a computer program for the handling of EMG data. Whilst not presenting any considerable difficulty for computing science experts such a program is a comparative innovation in this country. The program enables the selection of various integration times for processing the EMG signal (derived principally from surface electrodes) and two or three methods of averaging recurring tokens of EMG signal from a repeated 'same' utterance. The program is also immediately suitable for processing audio signals (particularly with respect to amplitude), EMG , air pressure and airflow, and ether signals from biological transducers, including movement. It a capable, even on a small computer, of comparing several different channels of information from different transducers with respect to amplitude, general 'shape' and event timing. The program is briefly described by Katherine Morton in her paper 'Computer Processing of Electromyographic and Similar Data' in our *Occasional Papers* No. 5, October 1969.

5. COMPUTING

One of the terms of reference of the present project was to gain expertise in computing. We have been able to use computers in the University — specifically an ICL 1909, a Honeywell 316 and a Honeywell 516 for short and erratically spaced periods of time. Kate Morton has acquired the ability to use DAP16, the Honeywell 3/4/516 assembler language. The Language Centre has just purchased us the minimum configuration of a 316 for our laboratory and Katherine Morton will be spending a large amount of time in the remaining months on this machine. It will form the basis of an enhanced configuration ready for our practical work in synthesis by rule. Kate Morton has worked with computer experts during the grant period on our EMG data processing system (Morton 1969); this is working well now and will be sued extensively on a 516 for processing some of the remaining experimental data.

6. DISSEMINATION OF RESULTS

We have been fortunate in obtaining the use of the Language Centre's publication 'Occasional Papers' for fast dissemination of papers relating to our work. We believed right from the start that this was essential and have sought every means to communicate our work to other researcher Papers have been read at a number of conferences in the UK, and on four occasions at meetings in the United States. Dr Tatham was invited to lecture at Ohio State University this last summer (1970) on speech production models and has been invited to give a series of lectures in Holland in January 1971 based on the Project's work. Our impression is that our work has been received very favourably.

We have organised a gathering of experimentalists in this field, taking place at the end of September 1970. Called "The Essex Symposium on Models of Speech Production: Aerodynamic and Myodynamic Studies" it will promote dissemination of our work, particularly as both the formal papers and the discussion will be published.

APPENDIX A

Summaries of Publications

i. Project Pilot Stage (pre-SRC Grant)

1. 'Some Electromyography Data towards a Model of Speech Production'. M. Tatham and Katherine Morton, in *Occasional Papers* 1. Language Centre, Essex University, May 1968; also *Language and Speech* 12,39 (1969)

> An experiment in elementary electromyography is described; data on action potentials obtained from *m. orbicularis oris* is presented. There is every indication that the EMG signal is statistically insignificantly different in duration and amplitude for initial and final /p/ and /b/ in Context with several vowels in monosyllabic words. The date is linked to an embryonic theory of production.

2. 'Further Electromyography Data towards a Model of Speech Production'. M. Tatham and Katherine Morton, in *Occasional Papers* 1. Language Centre, Essex University, May 1968

> The present paper presents some electromyography data which may indicate that there is an intimate cohesion between the phonemic elements of a ClVC2 syllable It is suggested (following MacNeilage) that C1V- are linked more closely than -VC2 that the linguistic organisation of motor commands is syllabic in nature and composed of syllabically dependent but individually identifiable extrinsic allophones exhibiting two dimensions of cohesion: linguistic and voluntary (syllabic and extrinsic), non-linguistic and involuntary (coarticulatory).

3. 'Classifying Allophones'. M. Tatham, in *Occasional Papers* 3. Language Centre, Essex University, March 1969; also *Language and Speech* (forthcoming)

> Following Wang and Fillmore, Ladefoged makes an explicit between two types of allophone: extrinsic and intrinsic. This distinction is taken up and discussed in terms of voluntary and involuntary operation of the neuromuscular system It is suggested that extrinsic events do not occur unless under direct voluntary control and that uncontrollable intrinsic events are bound to occur when an extrinsic event takes place. A third category is established: controllable intrinsic events — that is, events which are bound to occur unless there is specific extrinsic resistance. The concept of the third category obviates the need, which is the logical outcome of Ladefoged's approach, for a concession that all events are deliberately programmed ultimately.

4. 'The Control of Muscles in Speech'. M. Tatham, in *Occasional Papers* 3. Language Centre, Essex University, March 1969

> A brief survey of the anatomy and control of muscles is followed by an extended examination of the general function of muscle spindles and the role they might play in speech. Muscle spindles and the gamma system are capable of 'setting' the control system to the desired amount of muscle contraction (and therefore desired articulation) and, by means of the gamma loop feedback mechanism, of maintaining the required stability. Surveying the theories of ether speech researchers, it is argued that the gamma system accounts for (at least some of) the contextual variations observable in EMG signals.

5. 'Control Organization in Speech: Preliminary Report'. Katherine Morton, in *Occasional Papers* 3. Language Centre, Essex University, March 1969

> An outline of the preliminary interests of the present research project is presented. These arc divided into statements of problems surrounding the notions of i. voluntary and involuntary actions, ii. basic speech posture and its modulation, iii. timing of motor commands. The preliminary studies are concerned with confirming or refuting the hypothesis that motor commands closely correlating with the phonological phonemic segments interrelate in a rhythm-governed CV(c) repetition pattern.

ii. Publications since July 1969

6. ‘Experimental Phonetics and Phonology’. M. Tatham, in *Occasional Papers* 5. Language Centre, Essex University, October 1969

> Because phonological operations are not directly observable while phonetic ones often are, this fact does not automatically place one component in one plane (competence for phonology) and the other in the other plane (performance for phonetics). Just as our phonology is a systematic characterisation of the facts of phonological operations which must be known to the speaker, so phonetics can be a systematic characterisation of the facts of phonetic operations. The claim is made that by using the competence notion for handling underlying systems of the phonetics it is possible to generalise more widely and provide a direct link between the phonology and the phonetics something that has defeated attempts so far.

7. ‘On the Relationship between Experimental Phonetics and Phonology’. M. Tatham, in *Occasional Papers* 5. Language Centre, Essex University, October 1969

> This paper discusses the possibility that phonological performance could be dominated by lower level (i.e. phonetic) necessities. It is important to establish the role of late phonological rules and in particular to have a clear statement of just where the phonology should stop. It would be as well to bear in mind in any natural phonology that rules about voluntary event rules about involuntary events and rules about voluntary counteraction of intrinsic tendencies must not be confused together if any universality is to be apparent. Specifically that above all rules which purport in the phonology to be natural, but which span extrinsic and intrinsic levels in one condensed jump are likely to be inappropriate As has been advocated by several researchers a close look at the control of speech and its operation provides data which enables these errors to be avoided.

8. ‘Speech Synthesis — a Critical Review of the State of the Art’. M. Tatham, *in International Journal of Man-Machine Studies* 2, 1970

> This paper is divided into three parts: (a) the synthesiser, discussing various forms of apparatus; (b) control of the synthesiser — notions of speech synthesis by rule and the development of true generative capabilities in producing a theoretically infinite number of utterances from a store of a finite set of segments and a finite set of combinatory rules, (c) use of synthesisers. The final section discusses how, for the use of linguists, more attention should be paid to the implementation of a model of speech production which is linguistically oriented than to the goal of producing natural sounding synthetic speech — no matter how this is achieved.

9. ‘Articulatory Speech Synthesis by Rule: Implementation of a Theory of Speech Production’. M. Tatham, in *Working Papers in Linguistics* 6. Computer and Information Science Research Center, Ohio State University, Columbus Ohio, September 1970

> This paper discusses in detail the present state of the project’s speech production model, in particular our notion of the handling of syllabification and the introduction of time onto a nominal segmental input. A flow chart diagram is presented of the model and ways are discussed of implementing this model in a strategy for articulatory speech synthesis. The difficulty of constructing an articulatory synthesiser is solved (or postponed) by assuming its existence, but in fact using an acoustic synthesiser with a software input based on the conversion of articulation to acoustics (this approach has been adapted from Haggard).

10. ‘Coarticulation and Phonetic Competence’. M. Tatham, in *Journal of the Acoustical Society of America,* July 1970 (abstract); and appearing in *Occasional Papers* 8. Language Centre, Essex University, November 1970

> Possible criteria for establishing competence/performance distinctions in phonetic theory are examined. It is emphasised that, for the competence aspect, rules should take a form compatible with those of the phonological component and it is discussed

whether the coarticulation phenomenon can be handled adequately In the phonetic competence model. Recent electromyographic data is referred to and the problem of the relative roles of active and passive (voluntary and involuntary) programming of the muscles associated with articulation is re-examined.

11. 'Model Building in Phonetic Theory'. M. Tatham, in *Occasional Papers* 8. Language Centre, Essex University, November 1970 (invited Forum Lecture given to the Linguistics Institute of the Linguistics Society of America, Columbus, Ohio, July 1970)

This paper examines criteria for the construction of a linguistically oriented model of speech production which is at the same time a working model capable of implementation in speech synthesis. The current model is discussed in some detail with emphasis on the temporal aspects.

12. 'Computer Processing of Electromyographic and Similar Data'. Katherine Morton, in *Occasional Papers* 5. Language Centre, Essex University, October 1969

The technique at present employed at Essex for the computer handling of EMG data is reviewed. Software integration and averaging of many tokens is dealt with and block diagrams of the current interface are presented.

13. 'The Phonetic Component'. Katherine Morton and M. Tatham, in *Occasional Papers* 8. Language Centre, Essex University, November 1970 — paper read to the Linguistics Association of Great Britain, Manchester, April 1970

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.

14. 'Defining the Bases of Phonetic Theory'. M. Tatham, in *Occasional Papers* 8. Language Centre, Essex University, November 1970 — paper read to the Linguistics Society of America, July 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.

15. 'A Linguistically Oriented Approach to Speech Synthesis by Rule'. M. Tatham, in *Occasional Papers* (forthcoming). Language Centre, Essex University — paper to be read to the Linguistics Society of America, Washington, December 1970

The majority of research projects in synthetic speech centre around the development of a set of rules to provide correct time varying control signals for driving a speech synthesiser. 'Correct' usually means capable of enabling the device to generate speech-like sounds where quality is determined by spectral analysis or subjective listening tests. This approach is criticised, and the implementation of a linguistically oriented speech production model is discussed. The model, it is argued, is inadequate if it seeks to generate speech from a simple quasi-phonemic input alone. The consequences of adding a time dimension by rule to a segmental input are reviewed, but it is argued that there may be better ways of incorporating such hitherto problematical features by denying the single channel input.

(iii) Publications by Project Consultants Resulting Directly from Their Consultancies

16. H. A. Whitaker (Department of Linguistics and Department of Psychology, University of Rochester, New York) 'Some Constraints on Speech Production Models', forthcoming in University of Essex, Language Centre, *Occasional Papers*; and read at the Essex Symposium on Models of Speech Production, September, 1970

> A model of speech production must be consistent with and constrained by data and evidence from a wide variety of sources: acoustics, neurology, psychology and linguistics. Three related constraints are examined : 1. whether the output side of the model can be described with an associative chain hypothesis, 2. some evidence pertaining to the nature of the units in the model and 3. a proposed tracking mechanism that may account for certain errors in speech production. Evidence against an associative chain hypothesis is given; disconfirmation of such models may await neuromuscular/acoustic data that is not currently available but which could be obtained. Slips of the tongue (spoonerisms) and aphasic speech seem to provide little direct evidence that the units of the model are discrete segments (phonemes) although it is possible that analysis into motor command groups which control specific parts of the vocal tract may do so. One source of these errors is seen as a result of either mis-timing or erroneously scanning the output of the grammar by the mechanism which produces and transmits the motor program.

17. P. Mansell (Language Centre, Essex University — PhD Student named as research officer on the current application for an extension to the present SRC award) and R. Allen (Dept. of Electrical Engineering Science, Essex University — named as research officer on the current application for an extension) 'A First Report on the Development of a Capacitance Transducer for the Measurement of Lip Excursion', forthcoming in University of Essex, Language Centre, *Occasional Papers*; and read at the Essex Symposium on Models of Speech Production, September 1970

> It is shown how a convenient transducer for lip movements is a pressing necessity in some current lines of experimentation on the electromyographic activity of the lip muscle, in particular in the investigation of the variations of EMG gestures often encountered. A number of design principles for such a transducer are given and transducers in current use are evaluated in terms of those principles. The development of a capacitance device capable of measuring the horizontal excursion of the upper lip in the posterior/anterior direction of a point on the upper lip is described. The limitations of the device are discussed, and results from a preliminary experiment using the device presented. A brief outline of projected future work is included.

18. P. Mansell (Language Centre, Essex University. This paper is included because Mansell has been associated with the Project's work during his PhD period and is named on the proposal for an extension) 'The Nature of Variation in EMG Signals', forthcoming in University of Essex, Language Centre, *Occasional Papers*; and read at the Essex Symposium on Models of Production, September 1970

> Stress is laid on the necessity of an investigation of the variations which occur in EMG gestures in speech research in response to invariant stimuli. It is suggested that even a partial explanation may point to more theoretically based procedures for comparing sets of gestures given in response to different stimuli. Examples are given of the parameters of variation. The views of experimenters both in speech EMG research and in other fields are examined for their contribution to a study of the variation. The unusual nature of the signals derived in speech EMG research is noted, together with the perhaps unusual expectations of researchers in this field. It is first enquired whether there arc transformations which can be performed upon the derived data, either to eliminate what might be only the appearance of variation, or to suggest the causes of real variation. Secondly, a schematic model of a typical experimentation is set up, and the possible sites where variation might enter the system are listed.

Where these hypotheses are amenable to experimental test within the limited scope of phonetic research, suggestions are made as to appropriate procedures.

APPENDIX B

General Bibliography

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

G. Fant (1960) *Acoustic Theory of Speech Production.* Mouton, The Hague

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

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

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

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

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

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

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

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

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

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

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

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

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

S. E. G. Ohman (19675) Numerical Model of Co-articulation. *JASA* 41.2

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

L. D. Partridge and P.C. Huber (1967) Factors in the Interpretation of the Electromyogram based on Muscle Response to Dynamic Nerve Signals. *American Journal of Physical Medicine* 46.3

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

I. M. Slis (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

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

M. Tatham (1970a) Articulatory Speech Synthesis by Rule: Implementation of a Theory of Speech Production. *Working Papers in Linguistics* 6 Computer and Information Science Research Center, Ohio State University 4

M. Tatham (1970b) Speech Synthesis: A Critical Review of the State of the Art. *Int. J. Man-Machine Studies* 2

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

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

E. Werner and M. Haggard (1969) Articulatory Synthesis by Rule. *Speech Synthesis and Perception Progress Report* 1

H. A. Whitaker (1969) On the Representation of Language in the Human Brain. *Working Papers in Phonetics* 12 UCLA Los Angeles

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