Speech Synthesis and Models of Speech Production - II

Mark Tatham
Kate Morton
Phil Mansell

Reproduced from Interim Report (to January 1971): Science Research Council Award B/SR/6733 (July 1969-March 1971).

The experimental work and its relationship to our model of speech production documented in this interim report is not intended to be exhaustive in any way. The processing and writing up of experimental data is extremely time consuming and we have sought to present here notes on a few major experiments. Our final report will contain the appropriate documentation of all the Project's experimental activity.

INTRODUCTION

The experiments reported in the following pages are concerned with the development and refinement of experimental techniques for deriving information which is felt to essential for any theory of speech production, and consequently for a physiologically based scheme of speech synthesis.

Only a résumé of the theoretical reasoning which has lent these particular experimental concerns their significance will be given. We assume a phonology, and reasons have been adduced elsewhere for supposing that the output from this phonology represents a statement of linguistic intention, framed in perceptual terms (Mansell [1971]). The task of the phonetic component is seen in general terms to be the generation and control of a motor output which reflects the intersection of the linguistic intention, and the many other factors relating to the precise nature of the articulation, such as speech rate, conditions of external noise, familiarity of the speech material, and so on. The changes associated with the operation of a number of such factors are documented in a general way in the literature; for changes associated with speech rate, see Chistovich *et al*. [1967], with external noise conditions Lane *et al*. [1970], with familiarity of the speech material, Lieberman [1964]

Three levels of processing are distinguished:

1. motor planning,
2. the execution of the motor plan,
3. the control over that execution.

The experimental question to be decided is: what aspects of the speech output can be attributed to the operation of which of these levels of processing? A prerequisite of such an investigation is an adequate analysis of the output stage of the speech production model, and it is on experiments intended to increase the feasibility of such an analysis that this report is in the main concentrated.

This is inevitable, since the analysis of the relationship between the types of movements and the type of control is at present largely a non-experimental investigation (but see **Section 2**); the feasibility of adequate control in experiments directly limiting available feedback information (such as those proposed by Hardcastle [1970]) has yet to be demonstrated. Considerable progress is being made in this non-experimental investigation however, and will be briefly described. Russian and East European work on the degree of central involvement necessary in even reflexive movement (Anokhin [1960]; and especially Konorski [1967]), together with recent American work on types of feedback models and their relationships to particular types of motor tasks (Greenwald [1970]; Keele [1968]; Adams [1968]), and work in engineering psychology on the effect of feedback variables on the performance of complex tasks (see, e.g. Bahrick [1957]) have all had an influence on our thought in this direction; it is suggested that the account of feedback in speech being developed will more adequately reflect the variety of motor activity in speech than has been the case with previous accounts. Non-experimental work has also been continuing on the level of motor planning; it has been possible to show that a system of ordered rules expressing unitary processes but with variable outputs depending on the simultaneous presence or absence of variables representing the effect of the external factors referred to above (see Bierwisch [1967] for phrasing rules which produce different results depending on the speech rate externally specified).

Reverting to the experimental investigations; what precise aims do these investigations have which do not directly duplicate the conventional experimental alms of articulatory phonetics? In the first place, preliminary consideration of the phenomena of speech shows us that the *articulator* as such is an inconvenient unit of analysis. We prefer to adopt as the unit of analysis the effect which a particular group of muscles acting in some balance can have on a particular body of tissue. This approach commits one to investigating electromyographic activity during speech; the following reasons may be adduced:

The adoption of the muscle group as the unit of analysis multiplies the parameters of the analysis of speech movements, but considerably simplifies the specification along each parameter.

Without recourse to the muscular level of analysis it is impossible to distinguish between active and passive movements.

As a defence for the retention of the mention of the body of tissue involved in the definition of the unit of analysis, it can be noted that this reference considerably simplifies the writing of rules to account for coarticulatory phenomena between the activity of muscle groups associated with the same body of tissue

The ultimate aim is to dispense with the arbitrariness associated even with articulatory synthesis by rule; to replace the mechanical assignment of movement durations and velocities to ‘articulators’ with empirically derived equations specifying the time constants of the articulatory unit (muscles + tissue), the nature of the movement, and the significance attached to that movement by the higher processing, and the relationship between the minimum time necessary for the accomplishment of that movement and the maximum time implicit in the choice of speech rate. It is significant that Ohman [1967a] in perhaps the most complete theory of articulation yet proposed declared himself dissatisfied with the explanatory power of his model as long as it remained on the level of ‘articulators’.

The experimental methods reported here include not only electromyography but also aerometry and photo-electric glottography, as well as experimental methods of our own C devising

1. ELECTROMYOGRAPHY

A survey of the literature revealed that electromyography had proved a satisfactory experimental technique only in two general classes of experiment:

- Where the purpose of the experiment was exploratory in nature and the only information required was the presence or absence and timing of electromyographic activity;
- Where the specific hypothesis of the experiment could be again unambiguously investigated by demonstrating simply the presence or absence of muscle activity (see, e.g., Ohala [1970], who falsified a major premise of Lieberman [1967] by demonstrating laryngeal activity during stressed syllables).

In all other cases electromyographic investigations did not seem to have provided unambiguous demonstrations of hypotheses for two main reasons:

- because of variability leading to widely overlapping distributions in sets of signals to be compared,
- because of the unreliability of illicit comparisons of parameters of the EMG with an articulatory classification of speech movements.

These considerations suggested that EMG was an appropriate technique to use for discovering what the grouping of muscles in our 'articulatory units' were; but that investigations of the functional activity of these articulatory units in electromyographic terms would either have to be constrained such that the hypothesis could be adequately investigated by' means of data on a restricted number of parameters, or that basic investigation was needed as a preliminary into the nature of variations in EMG signals and into the relationship between these signals and articulator movement.

1.1 The Nature of EMG Variations

The variations observed in derived EMG traces have been described in Mansell (to appear), and are evident from the large ranges to be observed in published findings. Variation is inherent in biological systems (Rosenblith [1965]) and is evident in the performance of all skilled motor tasks. We have been influenced in our approach to this variation by the formulations of Welford [1958] who sees the central mechanisms responsible for skilled performance as

> ...capable of producing a response which is formed *ad hoc* by a kind of calculation based on many influences derived from the present aims and past experience of the subject and the sensory data of various kinds available at the time. ... Such a system results in a response which is unique on each occasions, although it is determinate and based on constants which are, at least in principle, observable. (p.27)

A phonetic component utilising computations of this nature has been described in Mansell [1971]; the implication is that the variation is central in origin, of course. This conclusion was reached by various experimental procedures, described in more detail in Mansell (to appear).

Three hypotheses were investigated:

1. That the variation was a low-level phenomenon,
2. That the variation was a function of the types of data analysis used — i.e. that the categories applied to the data were inappropriate,
3. That the variation was a function of higher level processing and represented to an extent the way in which the speaker organised his task at this high level.

Three areas were investigated under hypothesis 1. Findings are summarised below:

The variations could be due to time varying factors affecting the transmission of signals from muscle source to electrode pick-up. The major inhomogeneity in this transmission path is the resistive layer generally thought to be localised in the *stratum corneum* (Tregear [1966]). Mathematical methods exist for the analysis of the effects of inhomogeneities on the volume conduction of electric potentials (Rush [1967]; Geselowitz [1967]), and an attempt was made to derive empirical data on the time-course of variations in the tissue resistance in order that a model could be set up. Using methods advocated by Lykken [1959], however, no variations could be detected during speech or during the time-course of a typical experiment. It is clear that the variations of skin resistance employed in psychological experiments are extremely slow, having a time-course of hours. This negative result, together with the fact that extremely low values of skin-electrode impedance could be achieved by adequate preparation of the skin (Almasi and Schmitt [1968]), suggesting that the resistive layer had been bypassed, lead to the abandonment of this direction of research, and the rejection of this sub-hypothesis.

Evidence exists on the fluctuation of excitability of populations of neurons (see references in Mansell [to appear]); it was concluded that the variations encountered were likely to be of too high a magnitude to be accounted for by this effect.

EMG experiments typically involve large numbers of repetitions; a possible hypothesis therefore is that the observed variations could be a functions of neuromuscular fatigue. It was evident, however, that the EMG did not show any consistent trends of variation over time, and showed none of the characteristics associated with surface electrode study of fatigue (Cobb and Forbes [1923]).

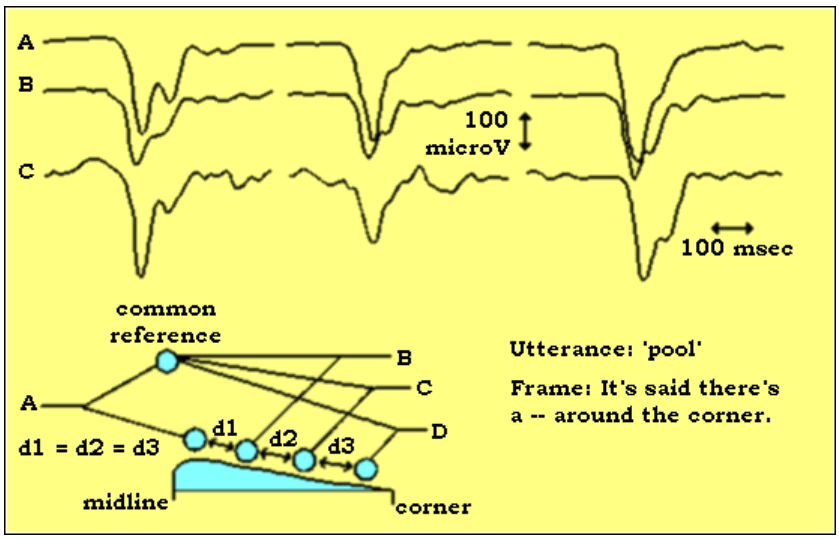

Fig. 1 Specimen traces from Common Reference Derivation. NB – the trace from electrode D was of very low amplitude and is not shown.

Hypothesis ii. was investigated mainly by statistical means. The principal effort was towards limiting the extent of the variation by attempting to find correlations among the varying parameters. In all cases investigated the correlation coefficients were very low. An interesting result was that while great variation was found on the EMG parameters the audio parameters of amplitude and duration were nearly constant throughout the material investigated. This result is typical in our experience.

Hypothesis iii. can be adopted in a model where, first, the activity along certain articulatory parameters is rated low in terms of the functional load of that parameter (for this concept see Wang [1967]; Mohr and Wang [1968]), and which incorporates a model of articulation where certain articulators are held to operate according to intrinsic time constants rather than extrinsic spatio-temporal demands (see **Section 2**).

If the variation is a function of central processing, then we should expect the movement resulting from the muscle activity also to vary. The expectation is fulfilled in preliminary results presented in Mansell and Allen (to appear).

That the variation is not a function of strictly local events has been exceptionally clearly shown by the results derived from the development in this laboratory of electrode placements corresponding to the system known as *Common Reference Derivation* in EEG work (Cooper, Ossleton and Shaw [1969]). Here, equally spaced electrodes are placed on the lip, each electrode part of a bipolar pair; in each case, however, the other member of the bipolar pair is the common reference electrode, attached in our case to the nose. This system permits the unambiguous comparison of amplitude and waveshape from each of the four active electrode sites. Typical results are presented in Fig. l. It will be seen that, whatever our interpretation of the derived signals from the lips, it is the case that this waveshape is in some sense typical of lip activity as a *whole*. This point has been frequently assumed in EMG work, but we believe that it has not been possible to show this unequivocally before.

1.2 The Relationship between Parameters of the Derived EMG Signal and Movement

The general position with regard to the relationship between EMG movement is expressed by Partridge and Huber [1967], when they state (p.1277) that

> '... it is not yet possible to examine an FMG record and interpret any detail of the movement pattern resulting from that neural-muscle activity'.

A linear relationship between 'muscle tension' shown by the physical expansion of the muscle mass itself and the envelope of the EMG signal has been shown many times. It should be noted, however, that this relationship holds only in isometric conditions, and breaks down under more realistic physiological conditions (Inman *et al*. [1952])

The one researcher to show a correlation between movement curves and EMG envelope was Fritzell [1969]. He showed that the envelope of muscle activity from *m. levator palatini* was of the same essential shape as the movement curve of the velum along its own axis derived from cineradiography, although the movement curve, as expected, lagged in time

In order to investigate this correlation for the lips a capacitance movement transducer was developed, and is described in Mansell and Allen (to appear) . In this device a high frequency carrier wave (1Mhz) is amplitude modulated by the change in capacitance between two small metal plates, one (the receiver) attached to tile lips, the other (the transmitter) held at a distance from and parallel to the first. The output after demodulation and amplification is shown to be proportional to 1/d, where d is the distance between the plates.

Work has continued in the following ways:

The design and development of a linearising circuit for the device.

The design of an electrode and electrode placement techniques that will satisfy tile demands of the movement transducer in terms of area shown to the transmitter, and also the demands of EMG recording in terms of satisfactory skin-electrode contact.

Designs were sought for new general purpose EMG amplifiers with limited bandwidth and high gain. These are now in the process of being built and will prove ideal for the picking off of the lower frequency EMG signals from the dual-purpose electrode.

1.3 The Experimental Delineation of Synergic Muscle Groups

In the introductory section of this paper mention was made that various theoretical reasons had led us to make the grouping of muscles round tissue bodies the unit of analysis in our investigation of speech movements. The task of discovering what these groups are therefore becomes important.

The lips have been our main concern. Something is known of the mechanical factors involved in lip movement (Lindblom [1967]), but there exists as yet no sizeable body of data on the functional deployment of the muscle groups around the lips. Three factors make the task difficult:

The small size, proximity, and complexity of insertion of the muscles make their isolation difficult.

The volume conduction of muscle potentials, a problem with needle electrodes as well as surface electrodes (Dedo and Dunker [1966]), becomes a large problem here.

Certain misconceptions as to the function and innervation of especially *m. orbicularis oris* are perpetuated in the speech literature. Thus many writers have referred to the sphincteric or 'draw-string' action of this muscle, whereas the anatomical evidence for sphincteric fibres is largely lacking (Sicker, Dubrul and Lloyd [1970]), and EMG records show clearly that the superior and inferior sections of this muscle show different patterns of activity (see, e.g. Ohman [1967]).

The nature of the electrodes to be used for this kind of work has been the subject of research. Our work has largely been with surface electrodes, and Fromkin [1965] used monopolar surface electrodes to map muscle involvement in certain facial gestures and speech sounds. In one of the very few experimental papers on the effects of different kinds of deployment of surface electrodes, Velé and Janda [1965] showed that bipolar deployment gave more effective discrimination between the action of closely adjacent muscles. We decided to adopt a bipolar system.

Another consideration was ease of application of electrodes. Fromkin used suction electrodes [1965], with the negative pressure supplied through a tube from the electrode to a vacuum pump. Following Novikova [1961] we have developed a suction electrode consisting essentially of a fairly stiff rubber pipette bulb stretched over a silver ring to which is attached, in our case, a conventional cup silver electrode. Negative pressure is supplied by squeezing the bulb before application of the electrode to the skin. Excellent results have been obtained with this method. We are at present investigating the involvement and sequential recruitment of muscle groups in the sequences [fu] and [uf] embedded in appropriate frames

2.0 THE NATURE OF SPEECH MOVEMENTS

Among the hypotheses derived from a consideration of the literature on speech movements was the possibility that there exist a restricted number of articulatory units which in certain linguistic contexts may be said to operate in a maximally non-determined manner, i.e. according to their natural time constants and according to simple laws relating extent of movement to velocity of movement. We considered that the agonist and antagonist groups around the lips, the velum and the jaw might operate in this way in contexts where there existed no absolute necessity for inhibition of the movement of this articulator by virtue of the segmental context of the utterance either to the left or right of the segment supposed to require their action. Confirmation of this view has come in a general way from our own work on the lips (Mansell and Allen, [to appear]), and from that of Ohala on the jaw [1968, 1970], and that of Björk on the velum [1961].

2.1 Experiments on Lip Excursion

Further experimental evidence on the nature of speech movements was sought. With an appropriate analog means of recording continuous movements it was possible to take advantage of the findings of Woodworth [1899] and Welford [1968] that the curves of movements where on the one hand the time taken to accomplish the movements was of paramount importance, and where on the other the precise extent of the movement was important, show strikingly different characters. The latter curves show a much more gradual initial phase and a flattening of the curve as the target area is approached. In this case the movements towards the target could take up as much as two thirds of the total time taken for the movement. Time dominated curves, on the other hand, show a much more symmetrical appearance between initial and terminal phases and show little evidence of corrective influence on the shape of the curve.

It was decided to construct an experimental situation in which whatever the true nature of the lip excursion should be, the subject would be required to treat the task as if it were the

first type described above, i.e. dominated by spatial co-ordinates. The movement curves derived from this situation would then be compared with those derived under normal speaking conditions, where the nature of the task would be assumed to be as the subject habitually construed it.

The receiving electrode of the lip movement transducer was affixed to the midline of the lower lip of the subject with spirit gum. (The lower lip was chosen in this experiment since greater amplitude of movement is expected from the lower lip — Fromkin [1964].) The distance between the transmitting and receiving electrode was adjusted to 9mm. and the offset of the receiver adjusted so that a reading of zero volts was obtained on a digital volt meter (see Mansell and Allen [to appear] for justification).

The speech material had to meet three requirements:

1. It should show only one occasion of lip protrusion;
2. The frame for this example of lip protrusion should consist of oral consonants known to permit lip movement co-articulation, and vowels not involving extensive jaw movement;
3. The material should involve only semantically meaningful utterances.

The utterance chosen was:

'It's in the cooler again today.'

Fifty tokens of this type were elicited first in the control 'normal' conditions. The subject received no feedback on his performance other than that which lie could normally be assumed to be using for the performance of the protrusion gesture. lie was given no specific instructions other than to keep his utterances as similar as possible, and to adopt a resting position between utterances involving open rather than closed lips. Counting and timing were performed for the subject by displaying to him the illuminated panel of a digital timer/counter fed with a pulse at the rate of 0.25 Hz. Initiation of successive utterances was thus separated by approx. 4 sec.

The output from the lip movement transducer was fed to one channel of an Ampex SP300 FM tape recorder after having been passed through a low-pass filter set at 40Hz to eliminate high frequency noise. The audio trace was fed to a synchronous channel of the recorder with the input amplifier operating in direct mode and incorporating audio equalisation. Tape speed was 15 ips.

During the control phase of the experiment the peak amplitudes of the movement trace were monitored from a cathode ray oscilloscope not visible to the subject. During the rest period between the control and the experimental condition, the median value for the distribution of peak values was calculated. (Borda and Frost [1968] showed that for small distributions of unknown variance the median value was less sensitive to extreme fluctuations.)

A sinusoidal signal of peak-to-peak voltage identical with that of the median value of the control experiment was fed to the Y-axis of the CRO and the vertical deflection adjusted to cover 4 cm. on the display. The time base was adjusted to show a stationary spot. Markers were affixed to the graticule to show more clearly the rest and target levels in the vertical dimension.

The experiment was then repeated with the CRO in easy view of the subject. His attention was drawn to the fact that he should expect to see a vertical reading in the neighbourhood of the [u] in /cooler/ corresponding to the protrusion of the lips for that vowel. He was instructed that his prime task was to produce the criterion measure of protrusion. He was allowed ten repetitions of the experimental phrase as practice. This practice phase was recorded. The experimental session then began. As before, fifty tokens of the type *'It's in the cooler again'* were elicited.

Five types of information were sought from the comparison of the control and experimental phases:

1. Primary effects of the change in experimental condition as reflected in the shape of the movement;
2. Secondary effects as revealed in the timing principally in the audio;
3. Comparison of the homogeneity of the peak amplitude of lip excursions in control and experimental conditions, both overall and as a function of time ('learning curves', but see below);
4. Relationship of peak amplitude to audio onset in both conditions, as well as distributions of the absolute amplitudes of excursion achieved by audio onset;
5. A measure of co-articulation of lip rounding to the left and right of the [u] segment.

The general experimental technique described above has been now sufficiently practised. Research continues on the experimental design, however, since there are a number of factors which have been shown to have an influence on similar motor performance in psychological experiments which have not yet been controlled for. Of principal interest are the following two variables, the first of which has a possible influence on the difficulty of the task for the subject, the second of which concerns the processing of the data under 3. above:

Fitts [1954] suggested a formula for expressing the relationship between the amplitude of movements, the time taken to achieve them, and the accuracy required. the formula was as follows

$$\text{Movement time} = a + b \log(2A/W)$$

where W is the width of the target within which the movement is required to end, measured parallel to the direction of the movement; A is the amplitude of the movement from its starting point to the centre of the target; and a and b are constants. As Welford [1968] notes, the essential point of the formula and of all suggested emendations, is that it makes movement time constant for any given ratio between amplitude and target width.

The variable that is not controlled in the design above is that of the width of the target area; tests are now being made to gauge the effect of widening the target limits for the lip movement excursion upon the performance of the subject on both primary and secondary measures.

The second point concerns precisely the same area of the experiment but relates to the scoring of performance. It was realised early in the development of the experiment that an arbitrary criterion of success would need to be developed for the purpose of deriving information under 3. above. In an important paper Bahrick, Fitts and Briggs [1957] discussed 'time-on-target-scores', which give the total time that the absolute magnitude of the error voltage was smaller than a given magnitude and a possible choice for scoring 3. They showed that when different values were taken for the given magnitude of error in time on target scores, and the resulting curves were compared with the RMS error score for the same data:

> 'It can be seen that each of the curves . . shows a maximal slope at a different range of variation of the RMS value, and becomes insensitive to variations outside that range.' (p. 261)

Procedures for deriving the RMS error score these researchers recommend are at present being investigated.

3.0 OTHER EXPERIMENTAL METHODS OF DERIVING ARTICULATORY DATA

3.1 Photo-electric Glottography

It was noted by Mansell [1971] that a study of whispered speech could help to clarify the nature of the interaction between linguistic intentions, phonetic processes, and communicative pressures. In particular it was argued that more detailed knowledge of articulation in whisper could lead to a characterisation of the notion *simplicity* in this part of the phonetic component. Thus, if it was shown that for example the gross adduction — abduction manoevres of the larynx for segmental gestures were identical in whispered and voiced speech, then we would infer that simplicity was being maintained at a high level in the processing, at the expense of

redundancy on the articulatory level. If, on the other hand, the larynx during whispered speech showed no signs of gross abduction during, say, phonologically voiceless stops, then we could conclude that simplicity at the articulatory level is being aimed for, at the expense of complication in the higher processing.

The literature points in different directions on this general problem, depending on whether segmental or prosodic factors are being considered. In the latter case, there is now a considerable body of analysis showing the presence in whispered speech of features not present in the corresponding voiced speech (see Meyer-Eppler [1957] and the references in this paragraph). It appears clear that when asked to make intonational judgements on whispered speech, listeners must be responding in some way to this additional information (Trim [1970]) since judgements are better than chance (Kloster Jensen [1958]) and specific effects can be predicted (Fónagy [1970]). The status of this extra information from the point of view of the sender of the message is far from clear, however. Fónagy claims that the variations among his speakers means that (p.181):

'Il n'existe pas de procédé de transcodage etabli pour l'expression de 1'intonation dans la parole chuchotée.'

Since there have appeared no accounts of how the acoustic results of whisper may be accomplished articulatorily, and since we possess no model of the acoustic results of the different excitation source in whisper this question cannot be considered settled.

For segmental gestures, as already mentioned, the picture is entirely different, with experimenters having shown how whispered speech is in many ways identical with voiced speech. The researcher with the most quantitative data on this question is Slis [1969]. Following various findings in the speech literature on the difference between voiced and unvoiced plosives and fricatives, and on the different affects these segments had on their immediate context, he carried out a series of acoustic experiments on Dutch material. The tests were carried out on normal and whispered speech; those findings relating to amplitude and duration are relevant to both, and the findings were identical in each case. The findings are summarised as follows:

- Noise burst duration longer in voiceless than in voiced consonants;
- Amplitude of noise burst higher in voiceless than in voiced consonants;
- Preceding vowel longer for voiced than for voiceless consonants;
- Tendency for vowels adjoining voiced consonants to be slightly higher in amplitude than those in voiceless situations;
- When the mean difference in duration of the preceding vowels and that of the silent interval are compared they appear to balance one another. The results on this test are presumed similar for voiced and whispered speech since they are averaged together for presentation.

His conclusion with regards to whispered speech (p.473 is

> '... it seems justified to consider whispered speech as normal speech minus voice, leaving the time structure and probably transitional cues intact.'

Articulatory data are also available for the segmental behaviour of the glottis during whispered speech. As suggested above, they point towards an explanation of behaviour in whispered speech favouring redundancy at the periphery rather than complication in the central processing In a paper whose main purpose was the introduction of their photoelectric device for registering gross movements of the glottal aperture, Malécot and Peebles [1965] presented traces (their Fig. 3, p.549) of glottal abduction during [apa] in both voiced and whispered speech; no scales are shown for the two traces, but since the authors pronounce that traces for glottal movement during [aba] and [ama], also shown, are 'essentially the same' (p.550), we may assume the amplification to be constant. In this case, the amplitude of the abduction manoeuvre in whispered speech is greater than that observed during voiced speech, in the approximate ratio 1.15:1.00.

Further data comes from Slis and Damsté [1967]. Using a glass fibre optic passed into the pharynx *via* the nose combined with a photomultiplier they transduced the light transmitted through the glottal aperture from a light source beamed through the tracheal wall. They investigated principally glottal manoeuvres during phonologically voiced and voiceless plosives and fricatives in intervocalic position. They included data on whispered speech, but their main interest appears to have been in normal speech, since their results for whisper are largely uninterpreted. Their results for voiceless plosives and fricatives for both conditions are given in Table I.

	Average T[c/v]%	Average A[v1v2]mm	Absolute T[c]mm
Voiceless plosives (normal)	120	18.5	22.2
Voiceless plosives (whisper)	63	32	20.1
Voiceless plosives (normal)	207	13.2	27.32
Voiceless fricatives whisper)	96	46	44.04

Table I

Where:

T[c/v] = transillumination during the consonant relative to maximal transillumination during the adjoining vowels in %.

A[v1v2] = mean amplitude of transillumination during the adjoining vowels in mm on the recording.

Absolute T[c] = (T[c/v] x A[v1v2] / 100

We can extrapolate from this table the following two observations:

1. That the absolute measures of vocal fold separation for voiceless consonants in normal and whispered speech are comparable;
2. A[v1v2] represents the average of the transillumination during two vowels; it is safe to conclude that throughout whisper the glottis was open to a greater extent than in normal speech.

Further use of these figures is unsafe, however. It might be possible to conclude, for example, that during whispered speech in this experiment the vocal folds were *adducted* for voiceless consonants, given that in each case Average A[v1v2]. Absolute T[c] and T[c/v] < 100. Slis and Damsté, however, note (p.107) that their traces for whispered speech show:

> '… at the start of each word the glottis opened which opening movement continued during the word. The consonant movement seems to be superposed upon this opening movement.'

It would appear to be this factor which has artificially enlarged the average value for A[v1v2] in Table I.

Further information is added in Table II, where Absolute T[c] is compared for normal and whispered voiced and voiceless consonants:

	Abs. T[c] normal	Abs. T[c] whisper
Voiceless plosives	22.2	20.1
Voiced plosives	2.38 – 24.41*	7.59
Voiceless fricatives	27.32	44.04
Voiced fricatives	0.15 – 11.47*	22.2

Table II Assembled from data in Slis and Damsté [1967]. *extremes of range

Although the mode of presentation of the results for normal voiced plosives and fricatives precludes detailed conclusions, it is evident that:

In whispered speech and normal speech the same tendencies are observable for glottal opening in voiceless as against voiced consonants, with the opening area much larger in the former case.

Further confirmation of 2. above comes from consideration of the data on sub-glottal and intra-oral pressures which Slis and Damsté include in their paper. The values derived during voiced fricatives in normal and whispered speech are given in Table III.

	Sub-glottal (Pt) mm H2O	**Intra-oral (Pc) mm H2O**
Voiced fricatives (normal)	125	50
Voiced fricatives (whisper)	100	60

Table III

Given increased glottal aperture we would expect, other things being equal, decreased glottal resistance. Approximations to the glottal resistance can be derived by the following equation (Scully [1969]):

Rg = (Pt – Pc) / Ug ….. Eq. 1

where U represents air-flow. The area of a constriction can be approximated by the following empirically derived equation taken from Warren and Dubois [1964]:

Ag = Ug / (k sq-rt(2 delta P / D))

Where

U = volume flow rate of air

Delta P = pressure drop across the orifice

D = density of air

K is a correction factor

The value of k is assumed by Warren and Dubois to be constant at 0.65. Using this value of k,

A = (1.1 x U) / sq-rt delta P ….. Eq. 2

A in cm squared, U in litres/sec, delta P in cm H2O (taken from Scully [1969]

Assuming U is constant for both normal and whispered speech g (an error on the conservative side, as will be seen, since air-flow during whisper is normally assumed to be greater than in normal speech) we can solve equations (1) and (2) using the values given in Table III.

Normal speech	Whispered speech
Rg(n) = (12.5 – 5.0) / Ug(n)	Rg(w) = (10.0 – 6.0) / Ug(w)

Hence for Ug(n) = Ug(w). Rg(n) > Rg(w) ;

And for Ug(n) < Ug(w), Rg(n) >> Rg(w) .

Turning to Eq. 2,

Ag(n) = (1.1 x Ug) / sq-rt7.5 = (1.1 x Ug) / 2.74

Ag(w) = (1.1 x Ug) / sq-rt4 = (1.1 x Ug) / 2

Again, given equal flow, the expected result, Ag(w) > Ag(n), is obtained; given greater flow in the whispered case, then Ag(w) >> Ag(n).

Our own efforts to replicate and extend these findings with respect to the specific hypothesis of simplicity outlined above, have to a certain extent been hampered by the

limitations of the photo-electric method of recording, and by the variety of glottal configurations subsumed under 'whisper'

In the experiment reported here the instrumentation consisted of the Photo-electric Glottograph developed by Frøkjær-Jensen at Copenhagen (Frøkjær-Jensen [1969]. In this system, a DC light source is conducted to the external wall of the trachea *via* a clear plastic rod. The change in supraglottal illumination with glottal activity is picked up by a wide-angle photo-cell introduced into the pharynx through the nose. The amplified signal was recorded on an Ampex SP300 FM tape-recorder, synchronously with the audio signal, and subsequently displayed on an Elema-Schönander ink-jet recorder

The introduction of the catheter containing the photocell and its cable into the pharynx was monitored both visually *via* a laryngoscopic mirror and by means of monitoring output signals during sustained vowels on an oscilloscope. The materials for the experiment consisted of a set of eight sentences, which set was pronounced five times each in normal and whispered modes. The set of utterances was chosen to show contrast between voiced and voiceless cognates in intervocalic position; also it had to satisfy the requirement that only close and half-open vowels were to be used (this is a function of the instrument — Malécot and Peebles [1965]; Frøkjær-Jensen [1969] and that, apart from the sounds of interest, the other consonants should be consistently voiced. The set was as follows:

Are there any{tales, dales, bills, pills, goals, coals, seals, zeals} around?

The nature of the results achieved in this experiment and consistently with this subject in other, similar experiments is illustrated in Fig.2, showing the glottogram trances for normal and whispered pronunciation of 'Are there any pills around?' Here the curve for the whispered glottogram can be seen to follow closely the shape of the voiced glottogram; the amplitude of the movements in the whispered mode, is, however, x 1/10 of that for the voiced mode in all its aspects. The very large opening movements including the general level of illumination which in Slis's data characterised whispering, associated with the taking of breath between utterances in the whispered mode are nonetheless of strictly comparable amplitude to those observed between utterances in the voiced mode.

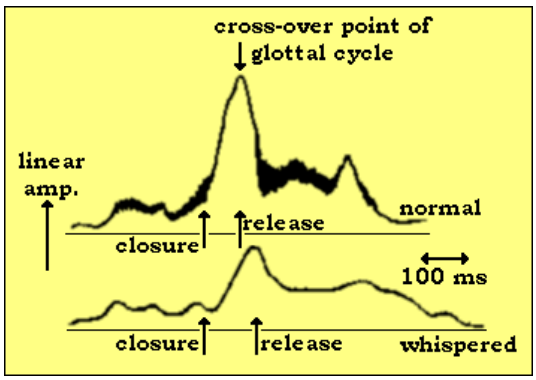

Fig. 2 Airflow traces in normal and whispered speech. Utterance: 'pills'; Frame: 'Are there any – around? The utterances are aligned on the moment of articulatory closure for [p].

In investigating this phenomenon, which takes precedence any more particular conclusions we might wish to draw the glottogram curves, irrespective of their amplitudes below), we have derived two hypotheses:

1. That the small amplitude of gross movements in this particular subject was a function of very high laryngeal tension consequent upon his using 'flute whisper' (van den Berg [1968], p.297);
2. That in the whisper type manifested by this subject the larynx tube above the glottis is very much constricted; this type of whisper was discussed and illustrated

in Ohala and Vanderslice [1965]. The result of constriction above the glottis would be acoustically, as the above authors note, that

> '… the entire length of this constricted tube contributes to the turbulence of the passing air which generates the characteristic noise of whisper.' (p.58)

The result from the point of view of electrical glottography would be a considerable diminution of the overall light level; a diminution in some ways comparable to the virtual extinction of the signal during the most open vowels (Frøkjær-Jensen [1969]).

It is evident that these experiments need to be replicated with a variety of subjects, and such replication is in preparation. Two general points can be made on the literature on the basis of the data collected so far:

The wide-angle lens of the photo-cell in use with the Frøkjær-Jensen glottograph does considerably increase the range of speech sounds that can be investigated. The inclusion of a nasal consonant in the test frame enabled us to gauge the effect of the photo-cell being displaced by movement of the velum. There was no cessation of the signal, although in some cases there was a slight diminution in the DC component of the signal level, as predicted by Frøkjær-Jensen [1969].

Malécot and Peebles [1965] point in their whispered traces to the apparent phenomenon of glottal *adduction* (as evidenced by diminished transmission of light) during intervocalic phonologically voiced consonants. heir account of this phenomenon would make it in our terms an example of peripheral redundancy. They do not however offer any hypothesis as to why this adduction movement should take place from the starting point of the characteristic vowel position for whisper. We believe that the explanation lies elsewhere. This phenomenon of a diminution in amplitude is visible in our records of normal) voiced speech. We would attribute such short-term effects on the general DC level to physical movements of the transducer as a result principally of movements of the tongue.

3.2 Another Aspect of Glottographic Traces

The glottographic traces obtained in the above experiment for normal speech are currently being analysed for the light they throw on the time relationship between articulatory parameters in running speech. It has been claimed (see, e.g. Lenneberg [l967]) that there exist timing relationships between events on different articulatory parameters which are too fine to be mediated *via* proprioceptors acting in a closed-loop manner. Such a claim leads to a position essentially that of Lashley [1951] who maintained that to account for the co-ordination of rapid movements an open-loop system was entailed. These claims have concerned the relationships between glottal and supraglottal articulations, and have been challenged by Ohala [1970] who makes the assertion in this that there is no controlled co-ordination between these two systems.

The opposite view has been put forward by Rothenburg [1968] with respect to the relationship between the release of an articulatory closure at the lips or in the mouth and the abduction-adduction manoeuvres of the larynx. This relationship has been shown to be an important one from a classificatory and perceptual point of view (Lisker and Abramson [1964]; and there exists a physiological model of this relationship (Frøkjær-Jensen and Rischel [to appear]). We are at present therefore investigating the variability of the time relationship between the release of the stop and the cross-over point of the glottal abduction-adduction cycle (point A in Fig. 2) . At present the moment of release is derived from the audio traces, but we hope to combine glottography with intra-oral pressure traces.

3.3 Air-flow Data

Much the same considerations as those outlined in the previous section have lead us to make use of the Electroaerometer, again developed by Frøkjær-Jensen. We are at present analysing air-flow and air pressure data with respect to the timing relationships of sequentially released oral stops, as in, e.g. 'catgut'.

However, the degree to which precise articulatory data can be derived from simultaneous measures of oral pressure and air-flow (see especially Scully [1969]) makes it extremely likely that great use will be made of this technique in the future. At present progress is being delayed by the manufacture of calibration equipment.

References

Adams, J. A. (1968) Response feedback and learning, Psych. Bull. 70, pp. 486-504

Almasi, J. J. and Schmitt, O. H. (1968) The dependence of skin-thru-electrode impedance on individual variations, skin preparation, and body location. Proc. Ann. Conf. Engineering in Medicine and Biology 10:BA2

Anokhin, D. K. (1961) A new conception of the physiological architecture of conditioned reflex, in Brain Mechanisms and Learning ed. Delafresnaye, Thomas

Bahrick, H. P. [1957] An analysis of stimulus variables influencing the proprioceptive control of movements. Psych. Rev. 64:5, pp. 324-328

Bahrick, H. P., Fitts, P. M. and Briggs, G. E. [1957] Learning curves — facts or artefacts? Psych. Bull. 54:3, pp. 256-268

Bierwisch, M. [1966] Regeln für die Intonation deutscher Sätze, Studia Grammatica VII, pp. 99-198

Björk, L. [1961] Velopharyngeal function in connected speech, Acta. Radiol. Supp. 202

Borda and Frost [1968] Error reduction in small sample averaging through the use of the median rather than the mean, Electroenceph. Clin. Neurophysiol. 25, pp. 391-392

Chistovich, L. *et al*. [1965] Speech: Articulation and Perception. Trans. US. Dept. Commerce

Cobb and Forbes [1923] EMG studies of muscular fatigue in man, Am. J. Physiol. 64:2, pp. 234-251

Cooper, R., Ossleton, J. W. and Shaw, J. C. [1969] EEC Technology. Butterworth

Dedo, H. and Dunker, E. [1966] The volute conduction of motor unit potentials, Electroenceph. Clin. Neurol. 20, pp.608-613

Fitts, P. M. [1954] The information capacity of the human motor system in controlling the amplitude of movement, J. Exp. Psychol. 47, pp. 381-391

F6nagy, J. [1969] Accent et intonation dans la parole chuchotée, Phonetica 20, pp. 177-192

Fritzell, B. [1969] The velo-pharyngeal muscles in speech, Acta Oto-laryng. Suppl. 250

Frøkjær-Jensen, B. [1969] Construction and comparative tests of two different types of glottographs, Paper delivered at the XVII Northern Congress of Otolaryngology

Frøkjær-Jensen, B. and Rischel, J. [to appear] Glottograph studies of the voice source, in Form and Substance ed. Hammerich, Jakobson and Zwirner, Akademisk Forlag, Copenhagen

Fromkin, V. [1964] Lip positions in American English vowels, Language and Speech 7, pp. 215-225

Fromkin, V. [1965] Some Phonetic Specifications of Linguistic Units: an EMG Investigation, UCLA WPP 3

Geselowitz, D. B. [1967] On bioelectric potentials in an inhomogeneous volume conductor, Biophys. J. 7, pp. 1-11

Greenwald, A. G. [1970] Sensory feedback mechanisms in performance control, Psych. Rev. 77:2, pp. 73-99

Hardcastle, W. [1970] The role of tactile and proprioceptive feedback in speech production, Department of Linguistics, Edinburgh University, Work in Progress 4, pp. 100-111

Inman, V. T. *et al*. [1952] Relation of human electromyogram to muscular tension, Electroenceph. Clin. Neurol. 4, pp. 187-194

Keele, S. W. [1968] Movement control in skilled motor performance, Psychol. Bull. 70, pp. 387-403

Kloster-Jensen, M. [1958] Recognition of word tones in whispered speech, Word 14, pp. 187-196

Konorski, J. [1967] Integrative Activity of the Brain. University of Chicago Press

Lane, H. *et al*. [1970] Regulation of voice communication by sensory dynamics, JASA 47:2(2), pp.618-624

Lashley, K. S. [1951] The problem of serial order in behavior, in Cerebral Mechanisms in Behavior, ed. Jeffress, Wiley

Lenneberg, E. H. [1967] Biological Foundations of Language. Wiley

Lieberman, P. [1963] Some effects of semantic and grammatical context on the production and percept ion of speech, Language and Speech 6, pp. 172-187

Lieberman, P. [1967] Intonation, Perception and Language. MIT Press

Lindblom, B. E. F. [1967] Vowel duration and a model of lip mandible co-ordination, STL QPSR 4/67, pp. 1-29

Lisker, L. and Abramson, A. S. [1964] A cross language study of voicing in initial stops: acoustical measurements, Word 20, pp. 384-422

Lykken, D. T. [1959] Properties of electrodes used in electrodermal measurement, J. Comp. Physiol Psychol. 52:6, pp. 629-634

MacKay, D. G. [1970] Spoonerisms: the structure of errors in the serial order of speech, Neuropsycholgia 8

MacNeilage P. F. [1970] Motor control of serial ordering of speech, Psych. Rev. 77:3

MacNeilage, P. F. and Declerk, J. L. [1968] On the motor control of co-articulation in CVC monosyllables, Haskins Labs: SR-12

Malécot, A. and Peebles, K. [1965] An optical device for recording glottal adduction-abduction during normal speech, Z. Phon. 18, pp. 545-550

Mansell P. [1971] Linguistic Parameters in Performance models, University of Essex, Language Centre, Occasional Papers 8

Mansell, P. [to appear] On the nature of EMG Variations, Proc. Essex Symposium on Models of speech production, University of Essex, Language Centre Occasional Papers

Mansell, P. and Allen, R. [to appear] A first report on the development of a capacitance transducer for the measurement of lip excursion, Proc. Essex Symposium on Models of Speech Production, University of Essex, Language Centre, Occasional Papers

Meyer-Eppler, W. [1957] Realisation of Prosodic features in whispered speech JASA 29 pp. 104-106

Mohr, B. and Wang, W. S-Y. [1968] Perceptual Distance and the specification of phonological features, Phonetica 18, pp. 31~45

Novikova, L.A. [1961] Electrophysiological investigation of speech, in Recent Soviet Psychology ed. O[1]Connor, Pergamon Press, pp. 210-226

Ohala, J. [1966] A new photo-electric glottograph, UCLA WPP 4, pp. 40-53

Ohala, J. [1970] Aspects of the Control and Production of Speech, UCLA WPP 15

Ohala, J. and Vanderslice, R. [1965] Photography of states of the glottis, UCLA WPP 2, pp. 58-59

Ohman, S. [1967a] Peripheral motor commands in labial articulation, STL QPSR 4, pp. 30-63

Ohman, S. [1967b] Numerical model of co-articulation, JASA 41, pp. 310-320

Partridge, L. D. and Huber, F. C. [1967] Factors in the interpretation of the electromyogram based on muscle response to dynamic nerve signals, Am. J. Phys. Med. 46:3, pp. 1276-1285

Rosenblith, W. A. [1965] The quantification of electrical activity in the nervous system, in Mathematics and Computer Science in Biology and Medicine. HMSO, pp. 131-138

Rothenberg, M. [1968] The breath-stream dynamics of simple released plosive production, Bibliotheca Phonetica 6

Rush, 5. [1967] A principle for solving a class of anisotropic current flow problems and applications to electrocardiography, IEEE Trans. on Bio-med. Eng. 14:1

Scully, C. [1969] Problems in the interpretation of pressure and air flow data in speech, University of Leeds, Phonetics Dept. Reports 2, pp. 53-92

Sicker, Dubrul, Lloyd [1970] Oral Anatomy 5th ed.

Slis, I. H. [1969] On the complex of the voiced-voiceless distinction: acoustic measurements, IPO MS 129

Slis, I. H. and Damsté, P. H. [1967] Transillumination of the glottis during voiced and voiceless consonants, IPO Annual Prog. Rep. 2/1967, pp. 103-109

Tatham, M. A. A. [1969] The control of muscles in speech, University of Essex, Language Centre, Occasional Papers 3

Tregear, R. T. [1966] Physical Functions of the Skin. Academic Press

Trim, J. [1970] Cues to the recognition of some linguistic features of whispered speech in English, Proc. 6th Int. Congr. Phon. Sci. Academia Publishing House of Czech Acad. Sci., pp. 919-923

van den Berg, J. W. [1968] Mechanism of the larynx and the laryngeal vibrations, in Manual of Phonetics, ed. Malmberg, North Holland, pp. 278-308

Warren, D. W. and Dubois, A. B. [1964] A pressure-flow technique for measuring velopharyngeal orifice area during continuous speech, Cleft Palate J. 1, pp .52-71

Véle and Janda [1965] Comparison of the monopolar and bipolar polyelectromyographic technique, Electromyography 5, pp. 97-98

Welford, A. T. [1968] Fundamentals of Skill. Methuen

Whitaker, H. A. [1971] Some constraints on speech production models, Proc. Essex Symposium on Models of Speech Production, University of Essex Occasional Papers, 9

Wickelgren, W. A. [1969] Context-sensitive coding, associative memory and serial order in (speech) behavior, Psych. Rev. 76:1

Woodworth, R. S. [1899] The accuracy of voluntary movement, Psych. Rev. Mon. Suppl. 3:2, pp. 54-59