A STUDY OF EMG PEAK AMPLITUDE MEASURES.

P.Mansell[1]

1. Introduction

In Mansell (1970b) an explanation was sought for the phemonenon of non-systematic variation in EMG waveforms associated with speech articulatory gestures. The waveforms under consideration were derived from surface electrodes on the upper lip and had undergone amplification, rectification and integration. It was noted that this variation was to be found on all the parameters conventionally employed to quantify these waveforms.

It was argued that the view we take of the significance of these variations to models of speech production is dependent upon our view of the relationship between the articulatory movement and the derived EMG signal. If we see the EMG as being directly related to the movement in each utterence of a token of a linguistic type, then we must develop production models which allow for variable outputs in response to a constant input. It is to be noted that this argument holds whether we take the EMG waveform to be the result of a set of "context-free" commands, with "coarticulation" being a purely mechanical effect weighting the relationship between the EMG and movement, or whether the EMG is taken to be the result of sets of commands which are themselves contect-sensitive. If, on the other hand, we are willing to see the relationship between EMG and movement as more complex, then alternative explanations for the significance of the variation can be derived. We might suppose, for example,

---

1) University of Essex.

that not all the muscle activity represented by the EMG waveform is of relevance to the articulatory gesture of concern - only a certain level of activity, say, may be required for such a gesture, the higher levels of activity recorded being representative of overshoot phenomena. Continuing this hypothetical line of argument, we might further suppose that amplitude and timing measures within this overshoot region (and peak amplitude measures for a large number of tokens would necessarily fall in such a region) would be liable to variation and of little significance in investigation of the organization of speech gestures.

In Mansell (1970a, 1970b) some implications of the former approach were explored. Basic to both approaches, however, is a more satisfactory description of the nature of the variation. In Mansell (1970b) the correlation between a limited number of the conventionally used parameters, including EMG peak amplitude, was investigated, and the resulting very low correlations regarded as providing some evidence of the inappropriateness of the parameters concerned to the description of relevant aspects of the EMG signal. The nature of the variation is not directly approached in this paper (but see section 5.2. below): rather, I am concerned to ask whether the events occurring at any electrode site are typical of lip activity as a whole, as measured simultaneously at other sites. This measure I have chosen to call the reliability of EMG peak amplitude measures. To demonstrate the reliability at a suitable level of significance will not decide anything about the causes of variation at particular sites. But if demonstrated this would resolve any doubts about the contribution of purely local factors to the variation.

2. Experimental method

The utterances for this experiment, all occurring monosyllables of English, and listed in Table 1, were

Table 1: Utterances used.

$C_n V_r C_n$	$C_l V_n C_n$	$C_n V_n C_l$
Cool	pack	cap
	back	cab
	mac	cam

where n = neutral with regard to lip activity
l = labial
r = rounded

chosen to provide a number of situations typical of those studied in the EMG speech literature, and at the same time provide reasonably simple derived waveforms for amplitude measurement. The articulator of interest is the upper lip, and the utterances allowed for the study of the contrast between voiced, voiceless and nasal labial stops, in syllable-initial and syllable-final positions, as well as the contrast between activity for labial consonants and a rounded vowel.

Four channels of EMG were utilized. The subject's upper lip was cleaned with detergent and the electrodes (sanded silver cup, approx. 5 mm. in diameter) were deployed in a Common Reference Derivation (see Cooper, Ossleton & Shaw 1970). That is to say, four active electrodes were taped to the upper lip as symmetrically as possible with respect to the midline, and all connected to a single reference electrode taped to the nose. Connecting all the active electrodes to a common reference electrode enables us to derive signals which are not only synchronous in time but which also have strictly comparable amplitudes in that one of the inputs to each of the four differential amplifiers is a constant. It was felt that such an electrode system was particularly suited to this study, since it enabled comparisons to be made among a number of amplitude measures during the same utterance.[2]

The centres of the electrodes labelled Right Centre and Left Centre in the presentation of results were sited approximately 1 cm. from the midline of the lips. The electrodes labelled Right Corner and Left Corner were placed with centres approx. 2 cm. lateral to the centre electrodes (but see section 5.1.2.). Thus they were not placed sufficiently laterally to pick up interference from other muscle groups which might have been active for spreading gestures.

The items analyzed in this paper were elicited as part of a larger corpus involving, in addition to the utterances reported here, more complex interactions between

---

2) It will be seen that the techniques of handling the data used in this paper are of such a nature as to cancel out the effect of constant variations between electrode sites. Such constant differences are not likely to affect correlation measures, and there is the check of bilateral symmetry to guard against such effects on amplitude measures. The possibility remains, of course, that uncorrelated variations may occur at each electrode site as a result of peripheral processes, which phenomena would be confounded with data relating to lip activity. Such a possible source of error is allowed to stand here, since the examination is essentially of the feasibility of current practice of using amplitude measures as discriminants of EMG activity: and current practice does not include any means of controlling for peripheral variations.

the labial stops and the rounded vowel, together with items involving a labial fricative. The items, in orthographic form, were arranged in lo randomized lists, which were read by the subject at a steady pace. The frame for each monosyllable was:

I'd like to say a ________ again.

As noted, the signals from the electrodes were led to four differential amplifiers (Fonema), and were recorded together with the audio signal on an Ampex SP 300 FM recorder running at 15 ips. The signals were then HP filtered to eliminate low-frequency movements artefacts due to changes in skin-electrode contact. They were finally rectified and integrated (5o msec. time constant) before being displayed on a 4-channel Mingograf ink-jet recorder. Two runs of the data were made, each run containing two channels of EMG, audio oscillogram and a rectified integrated version of the audio.

The nature of the design required that amplification be identical on all EMG channels. The system was calibrated and adjusted before the experiment and checked again after the experiment was over. The amplification and time relationships of the Mingograf channel were checked repeatedly throughout run-out of the data.

Measurements were carried out manually from the Mingograf traces. Peak amplitudes were measured in arbitrary linear units from an arbitrary baseline, set slightly higher than the amplitude of the inter-utterance noise. It was observed that the baseline remained stable on all channels throughout the experiment. lo tokens were measured for each type. In the case of the rounded vowel in cool, however, it was observed that on a number of occasions the derived amplitude curve contained two peaks. These were treated as separate items of data for the purposes of this experiment. The number of measurements taken for cool, 15, is therefore higher than in the case of stops. First and second peaks for any token of cool are distinguished in Tables 2-7 as (a) and (b) after the number of the token.

3. Objectives and hypotheses

The specific objectives and hypotheses of this experiment can be summed up under the following heads:

3.1. Dependence of Peak Amplitude on electrode placement

The basis for deriving hypotheses here should be an account of how response can be graded along the length of the upper lip. Such an account appears impossible to give. Certainly, the variety of labial articulations— "a speech articulation system in miniature" (Öhman, Leanderson & Persson 1965)— would suggest that the sphincteric function commonly ascribed to the orbicularis oris muscle cannot be the only mode of operation of this muscle in speech. From the anatomical point of view, the picture seems only generally established. According to Hollinshead (1968, p. 336) it is not known to what extent true sphincteric fibres exist. Sicker, DuBrul & Lloyd (1970 p. 143-144) note that the fibres of the muscle can be divided into an upper and a lower group which cross each other lateral to the corner of the mouth. This division appears to correspond to the observation in speech research that EMG signals from upper and lower lips in bilabial stops reveal a different mode of operation, and in labial fricatives a different degree of involvement. In addition, however, it is noted by the above authors that the majority of upper and lower fibres are confined to one side only,

TABLE 2

Raw amplitude scores (arbitrary linear units)

2.1 PACK

No.	R. Centre	R. Corner	L. Centre	L. Corner
1.	34.5	27	37.5	23
2.	50	34	60	40
3.	38	25	33	28
4.	42	32	40	31
5.	45.5	29	49	32
6.	39	25.5	51.5	28
7.	40	28.5	45	36.5
8.	32.5	23	34.5	26.5
9.	35	21	37	21.5
10.	41	20	46	33

2.2 CAP

1.	34	27	33.5	28
2.	43	32	40	29
3.	37.5	31.5	42	32
4.	36.5	23	32	21
5.	38	37	42.5	32.5
6.	39	38	44	25
7.	50	21	39	33
8.	31	19.5	40.5	31
9.	40.5	32	40	27
10.	25	14	27.5	14.5

TABLE 2 - continued

2.3 BACK

No.	R. Centre	R. Corner	L. Centre	L. Corner
1.	43.5	22	43.5	28
2.	46	31	5o.5	35
3.	42.5	22	37.5	28.5
4.	52.5	32	38.5	37.5
5.	37.5	19.5	45	3o
6.	43	37.5	44	28
7.	41.5	29	39	36
8.	31.5	22.5	4o.5	32
9.	42	32	49.5	36
1o.	29	21	3o	16

2.4 CAB

1.	32.5	19.5	39.5	24.5
2.	39.5	3o	42	35.5
3.	45.5	34.5	48.5	39.5
4.	36	17	46.5	22.5
5.	38	33.5	32	18.5
6.	42.5	33.5	39.5	22.5
7.	5o	37.5	54.5	33
8.	2o.5	15	35	18.5
9.	4o	2o.5	47.5	3o
1o.	29	15	31	16.5

TABLE 2 - continued

2.5 MACK

No.	R. Centre	R. Corner	L. Centre	L. Corner
1.	38	21	45.5	25.5
2.	38.5	31	49	31
3.	37.5	24.5	48.5	33
4.	35.5	22	37	25
5.	37.5	27.5	39.5	25
6.	31.5	28.5	37.5	23
7.	51.5	3o	41	41
8.	35	22.5	4o.5	29.5
9.	37.5	24.5	33.5	23
1o.	38.5	25	31.5	26.5

2.6 CAM

No.	R. Centre	R. Corner	L. Centre	L. Corner
1.	43.5	3o	4o	34
2.	42.5	17	47	28.5
3.	3o.5	22	38.5	24
4.	33	22.5	39	2o.5
5.	58	35.5	43	36
6.	47	32.5	5o	33
7.	43	38	41.5	31.5
8.	41.5	29.5	5o	3o
9.	35	23	42	29
1o.	53	29	48	31.5

TABLE 2 - continued

2.7 COOL

No.	R. Centre	R. Corner	L. Centre	L. Corner
1.	2o	18	24	23
2a.	15.5	14	16	17
2b.	13	13	16	19
3.	32	2o.5	36	28.5
4.	25.5	2o	3o.5	29.5
5a.	38.5	17	45	25.5
5b.	27	24	32.5	2o.5
6a.	37	23.5	39.5	31.5
6b.	19	14	18.5	24
7a.	31.5	13.5	36.5	26
7b.	18	12	16	25
8a.	28	17	33.5	32.5
8b.	26.5	15	24	2o.5
9.	17	11	15.5	14.5
1o.	27	18	22	24.5

TABLE 3

Tests of difference among parameters
of same type: Mann-Whitney U.

Type	A-B		B-D		C-D		A-C	
	U	Sign	U	Sign	U	Sign	U	Sign
Pack	3	.001	33	NS	7	.001	42	NS
Cap	16.5	.01	50	NS	8.5	.001	42	NS
Back	8	.001	34.5	NS	6	.001	46	NS
Cab	15.5	.01	46.5	NS	9	.001	38	NS
Mac	0	Abs.	33.5	NS	7.5	.001	36	NS
Cam	7	.001	41	NS	0	Abs.	46.5	NS
Cool	40	.001	27.5	.001	98.5	NS	102.5	NS

where: A = Right Centre   C = Left Centre
B = Right Corner   D = Left Corner

All significance tests one-tailed.

TABLE 4

Kendall Coefficient of Concordance: W.

Type	N	k	$s(=\sum(R_j - \frac{\sum R_j}{N})^2)$	$W(=\frac{s}{\frac{1}{12}k^2(N^3 - N)})$	$X^2(=k(N-1)W)$	Sign. (df=N-1)
Pack	10	4	973.5	.738	26.568	.01
Cap	10	4	756	.572	20.592	.02
Back	10	4	689	.522	18.792	.05
Cab	10	4	1019	.772	27.936	.001
Mac	10	4	703.5	.533	19.188	.05
Cam	10	4	910	.689	24.8	.01
Cool	15	4	3408	.761	42.616	.001

TABLE 5

Spearman Rank Correlation Coefficient: r_s.

Type	A-B r_s	A-B sign.	B-D r_s	B-D sign.	C-D r_s	C-D sign.	A-C r_s	A-C sign.
Pack	.646	.01	.51	NS	.676	.05	.758	.01
Cap	.549	NS	.118	NS	.445	NS	.33	NS
Back	.564	.05	.491	NS	.273	NS	.273	NS
Cab	.92	.0005	.639	.05	.855	.005	.743	.05
Mac	.388	NS	.316	NS	.655	.05	.288	NS
Cam	.661	.05	.828	.0005	.388	NS	.541	NS
Cool	.617	.01	.46	.05	.737	.005	.945	.0005

where: A = Right Centre, B = Right Corner, C = Left Centre, D = Left Corner

All significance tests 1-tailed.

TABLE 6

Tests of differences across types.
Mann-Whitney U.

Types	A-A		B-B		C-C		D-D	
	U	Sign.	U	Sign.	U	Sign.	U	Sign.
PACK/CAP	38.5	NS	45.5	NS	34	NS	41	NS
BACK/CAB	36	NS	44	NS	50	NS	33.5	NS
MAC/CAM	31.5	NS	38.5	NS	30	NS	36.5	NS
PACK/BACK	39.5	NS	50	NS	48	NS	42.5	NS
PACK/MAC	36	NS	43.5	NS	41.5	NS	38.5	NS
BACK/MAC	31	NS	46.5	NS	42	NS	31	NS
CAP/CAB	48	NS	44.5	NS	39	NS	44.5	NS
CAP/CAM	31.5	NS	49	NS	25	.05	37	NS
CAB/CAM	32.5	NS	42	NS	39.5	NS	35	NS

where: A = Right Centre C = Left Centre
B = Right Corner D = Left Corner

All significance tests one-tailed.

TABLE 7

(See text, section 5.2)

Type	Nx, where x > 32	% contrib. of error score to <u>s</u>	Nx, where y < 12	% contrib. of error score to <u>s</u>
Pack	2	34.7	1	33.3
Cap	1	42.9	1	16
Back	1	42	1	17.6
Cab	2	39.9	2	47.2
Mac	1	15.7	2	57.7
Cam	2	49.7	1	18.6

interlacing at the midline with the fibres of the other side. The deepest fibres of the muscle, however, cross from side to side without interruption. We should expect bilateral operation in speech, and hence can attach little functional significance to this arrangement. Some authors are content to observe that the bundles of the orbicularis oris corresponding to the red zone of the lips are finer and more densely arranged than are those in the periphery, while Hamilton (1958 p. 165) suggests a further functional division. He writes:

> "The upper and lower lips are provided with a series of small muscular fasciculi which act as labial tractors, i.e. they tighten the lips...."

Such an observation does not receive any support that I am aware of in the literature, and seems an uncertain basis on which to build an account of graded response.

In the absence of such an account, therefore, a specific hypothesis about the nature of the relationship between electrode siting and amplitude of derived waveform was not possible. The experience of many researchers that the highest amplitude traces were to be expected near to the mid-line of the lip led to the expectation that the more laterally placed electrodes would exhibit lower amplitude traces.

3.2. The reliability of peak amplitude measures

It was suggested that the following two conditions should be taken as measures of reliability:

(i) That even if the absolute amplitudes of the peaks from the various positions should vary, there should nonetheless be a high overall correlation among the amplitude scores paralleled by high correlations between pairs of electrodes. Failure of correlation measures to reach acceptable standards of significance would be interpreted as suggesting that events at the electrode sites were to a degree independent, and that hence the measure as a whole was unreliable.

(ii) That each output channel should show the same measure of discrimination on peak amplitude measures as between different linguistic types. Thus, if it is possible to discriminate between a syllable-initial and syllable-final stop by means of the signal from one electrode site, then, if the measures are to be held reliable, this discrimination should also be possible at all other electrode sites. It was predicted that a set of utterances which could be shown to be deviant on this measure would either exhibit such deviance through absolute unreliability, or because different parts of the lip were being used only partially differently in the two linguistic types. That is, the behaviour of the centre of the lips in two linguistic types may be different while the behaviour of the corners of the lips may be the same. In order that the latter explanation be entertained it is suggested that the deviance should be exhibited bilaterally in any given case.

4. Results

Two factors led to the choice of non-parametric tests for the statistical analysis of data. In the first place, experience with the use of parametric tests had led to the conclusion that the very large standard deviations encountered suggested that the assumptions about the distribution of the data required by these tests were not being met. Secondly, in the present design, the multiplication of electrode sites required that the number of utterances be kept fairly small. The choice does not appear to limit the validity of the results, however, since, according to Siegel (1956) the use of non-parametric tests increases the generality of the findings.

The tests employed here are the Mann-Whitney test for differences between two independent samples, the Kendall Coefficient of Concordance, a measure of agreement between results where the parameters are greater in number than two, and the Spearman Rank Correlation Coefficient. For the case of r_s, significance was determined by using a table of critical values, with extreme and doubtful cases calculated according to

$$t = r_s \sqrt{\frac{N-2}{1-r_s^2}}$$

The raw amplitude scores, expressed in arbitrary linear units, appear in Tables 2.1 - 2.7 The results obtained from statistical tests on these scores are given in Tables 3 - 7.

4.1. Dependence of amplitude scores on electrode placement

The first set of results relate to the dependence of the amplitude scores on electrode placement. This was investigated by looking for significant differences between the scores from amplitude sites within types. The results are presented in Table 3. H_o in this case was that the scores from different sites did not differ. Taking the stop series first, it will be seen that in every case H_o could not be rejected at a satisfactory level of significance for the comparisons between amplitude scores for centre sites, and between amplitude scores for corner sites. On the other hand, it will be seen that for each stop also, H_o can be rejected at a high level of confidence (above .01 in every case) for the comparisons between right center and right corner, and between left centre and left corner. Indeed, it will be seen that U, a measure which can vary from 0 to

50 in this case, achieves absolute significance in one case for each of these sets of comparisons.

In the case of the rounded vowel, the case is rather different. Here there are no grounds for rejecting H_o in the comparison between left centre and left corner, and between right centre and left centre, whereas we can reject H_o at a high level of significance for the comparisons between right centre and right corner, and right corner and left corner.

4.2. Results relating to first criterion of reliability

Tables 4 and 5 give results relating to the first criterion of reliability, given in 3.2. (i) above. The results for the Kendall Coefficient of Concordance (Table 4) show significant correlations (above .05 level) for all electrode positions for all types. Table 5, results of the Spearman Rank Correlation test, shows the nature of these correlations in greater detail.

There is a distinct lack of agreement between these two measures in the case of cap, which according to the Kendall test has a W significant at the .02 level, whereas it demonstrates no significant correlations among its electrode positions on the Spearman tests. This discrepancy appears to be explained by suggesting that for the Kendall test the value of s was inflated by the fact that the amplitude scores on all electrode positions for cap 10 (see Table 2.2.) were uniformly ranked lowest at each site. The conjunction of lowest rankings gave rise to a large difference score.

From the stop sequence there appear to be few regularities to be extracted on the Spearman tests. The exception is the case of cap vs. cab: while the cab scores provide grounds for rejecting H_o at a significant level for all four sets of comparisons, in the case of cap there are no grounds for rejecting H_o for any of the sets of comparisons.

Pack and back share a lack of correlation between contralateral corner electrode sites, and a significant correlation between right centre and right corner sites. They are distinguished by the fact that back, unlike pack, lacks a significant correlation between the two left sites, and between the two centre sites.

Cam, which might be expected to follow cab, is distinguished from it by having an extremely significant correlation between the scores on corner sites, and not showing a significant correlation between centre sites, or between sites on the left side. By the same token, mac is distinguished from back by showing a correlation between left-hand sites whereas it does not show a correlation between right-hand sites where back does.

Cap and pack share a lack of correlation on the comparison between corner sites. On the other comparisons cap shows no significant correlation while pack shows varying degrees of significance.

Cab and back share a correlation between right-hand sites, although the significance is higher in the case of cab. On other comparisons, however, they are distinguished by degrees of significance on the part of cab and lack of significance on the part of back.

Cam and mac share a lack of significant correlation between centre sites. They are distinguished, however, on the comparison between left sites by the fact that cam shows no significant correlation while mac exhibits a correlation significant at the .05 level. On other comparisons, mac shows no significance, while cam does.

In the case of the rounded vowel in cool, on the other hand, there is consistently significant correlation among the sites, with the highest being between the centre sites and the lowest being the corner sites.

4.3. Results relating to second criterion of reliability

The second criterion of reliability, given in 3.2. (ii), was investigated by comparing scores from identical electrode sites across linguistic types, utilizing once again Mann-Whitney U tests. In all cases except one no significant differences could be found across types. The single exception occurs at the left centre site for the comparison between cap and cam. Here H_o could be rejected at the .05 level of confidence.

5. Discussion

The results will be discussed in the order in which they were presented.

5.1. Dependence on electrode position

The results relating to the dependence of amplitude scores on electrode position are initially very clear. During each of the stops the lips appear to operate symmetrically, since there is no means of distinguishing between sets of signals from right centre and left centre, and between right corner and left corner. On the other hand, in every case for the stop sequence the amplitudes from the centre sites were found to be significantly greater than those from the corner sites. There are two possible explanations for this latter finding:

(i) That the signals on the corner sites are related to those on the centre sites by volume conduction of potentials from the centre sites to the corner sites. (See Dedo & Dunker, 1966).

(ii) That evidence has indeed been given for the operation of a mechanism which grades response along the length of the upper lip, with maximum amplitude concentrated at the centre of the lips for labial stop consonants. It is maintained that there is sufficient evidence contained in the present experimental design to enable us to choose the latter explanation with some degree of confidence. In the first place it should be noted that the volume conduction hypothesis includes a more extreme form of the grading hypothesis - namely that there is little or no activity at the corner sites, the derived signal having been transmitted from other sites. Further, in the case of the volume conduction hypothesis, we should expect there to be a consistently high degree of correlation between centre and corner sites. It is evident from Table 5 that this is not in fact the case.

Acceptance of explanation (ii), however, must involve leaving open the question of the nature of the mechanism involved. The functional basis for the mechanism seems clear - the centre of the lips has further to travel to achieve closure than do the corners. The physiological basis, though, as implied by Section 3.1., is extremely problematical. I am not aware of any work relating to the existence and nature of an "innervation zone" in the lips such as is found in skeletal muscles and has been reported by Sonesson to be the case in the <u>vocalis</u> muscle (Sonesson 1960 p. 24). Hence the possibility of selective involvement of motor units in different types of operation cannot be discussed.

5.1.1. The possibility of artefacts

It appears possible to explain the asymmetries in the results in Table 3 for cool relatively simply as an artefact of electrode placement. This explanation further has the merit of pointing to a possible difference in active lip region between stop consonants and rounded vowels.

It will be remembered that while on the right side the significant difference between centre and corner sites familiar from the stop cases was also found for cool, this difference was not observed on the left side. Further, while a lack of difference was found between the centre sites, again following the stop case, a difference was found at a high level of significance between the two corner sites. Now, suppose that the left corner electrode had been placed further towards the midline than the right corner electrode. Let us further suppose, following Section 5.1.1., that the amplitude of the signal is graded in some way from highest at the centre to lowest at the corner, then the asymmetries in the data would be explained, in that the amplitudes of left corner and left centre sites would become more similar, while at the same time the amplitudes from the left corner site would become dissimilar to the amplitude scores recorded contralaterally.

This explanation, however, begs the question of why such an artefact of electrode placement does not show up in the stop data. To resolve this question, it is possible to make the speculation, given that there are, crudely, "regions of high activity" and "regions of low activity" along the length of the upper lip, that the region of high activity is wider for the rounded vowel than for the stop consonants, and hence included the asymmetrical left corner electrode. This electrode, in common with the right corner electrode would be supposed under this hypothesis to be outside the high activity region in the case of stops. The present study shows no direct evidence for or against such a hypothesis. As a first approximation to distinguishing between the form of the labial articulations in the two cases of closure and rounding, however, the hypothesis has a certain plausibility.

With the possible exception, then, of the rounded vowel, the results have shown that the peak amplitude of derived EMG waveforms is dependent upon electrode position. The nearer the midline the electrode is placed, the larger the signal can be expected to be. This relationship is found on both sides of the lips, and confirms the observations of researchers seeking optimum positions for recording lip EMG. Further, as expected, the lips act in a symmetrical fashion, there being for the stop group no significant differences between the sets of amplitude scores from centre sites, or from corner sites.

5.2. Correlation measures and the overshoot hypothesis

As will be apparent from the presentation of results in Section 4.2., the Spearman Rank Tests appear to give a more detailed account of the manner in which the reliability criterion of 3.2.1. is met than do the Kendall tests. The experience with the discrepancy between cap on the two tests reported in 4.2., however, suggests a means by which the Kendall tests in this case can be applied directly to the problem of the explanation of EMG variation as stated in Section 1 of this paper.

Computation of W in such tests involves ranking the scores for each electrode site separately within a linguistic type, and then, for each token, computing the sum of the ranks given the scores on each of the sites. The mean of these sums of ranks for all the tokens of a type is computed, and the difference between the total of ranks for each token and the mean is found. These quantities are squared and the squares summed to give s. Generally speaking, the magnitude of W, and hence the significance of W, grows as s increases in size for a constant number of parameters and number of tokens within the parameters.

The largest contributions to s come from cases where, to illustrate from the case in hand, the four electrode sites are agreed in ranking scores for a particular token either high or low. In both these cases the difference between the sum of ranks and the mean of the sums of ranks will be great, and the contribution to s large.

A theory of EMG variation which supposed that there was a region of overshoot, within which amplitude scores were variable, would presumably predict that individual sites would more often agree in ranking amplitude scores for a particular token low than high, since the highest scores must be within the region of overshoot, whereas the lowest scores almost by definition will not. Therefore the "overshoot" hypothesis predicts greater variability at the high end of the sets of amplitude scores, and greater homogeneity at the lower end.

In order to test whether this prediction is borne out by the facts, the following crude test was devised. Conventionally, ranks are assigned with the highest number for the lowest rank. In the case that all four electrode sites were agreed in assigning the amplitude scores for a particular token the lowest rank, the sum of ranks will be 40. If, on the other hand, they assign them the highest rank the sum of ranks will be 4. The difference between these scores and the mean (22, i.e.

$$\frac{4\sum_{i=1}^{10} i}{10}$$

in our case) will be +18 and -18 respectively. The sign difference will disappear when the difference scores are squared and the same contribution will be made to s in each case. If there are differences between the tendency for high amplitudes and low amplitudes to exhibit greater and less variance, then this should be revealed by:

(i) the number of sums of ranks which are of small magnitude compared with the number which are of large magnitude.

(ii) The contribution which the squared difference scores related to these sums of ranks make to the magnitude of s expressed as percentages.

Such measures were derived from the stop set and are presented in Table 7. Arbitrary limits were set for the upper and lower ends of the scale of magnitude for sums of ranks, such that at the lower end the sum of ranks should be greater than 32, and at the upper end should be less than 12. In the case of perfect correlation this would limit the cases considered to those where all scores were ranked 1 or 2, or 9 or 10. Columns 1 and 3 of Table 7 show the number of times in each type that sum of ranks scores falling within the lower and upper limits respectively occurred. Columns 2 and 4 show the percentage contribution made to s by the squared difference scores associated with these sums of ranks scores.

There does not appear to be much information to be gleaned from the statement of numbers of sums of ranks scores falling in the two limits. In two cases the number in the lower limit is greater than that in the higher limit, the two contain equal numbers on three occasions, and the higher limit contains more occurrences than the lower in one case. A Mann-Whitney U test was applied to the percentage scores, however. A value of 15 was computed for U, which value has an associated significance of .35. Hence the test fails to find a significant difference between the two sets of percentages.

We may conclude, therefore, that insofaras the test described captures predictions under the "overshoot" hypothesis of Section 1, it does not provide any evidence in favour of it.

5.3. The first criterion of reliability

For the stop sequence, except for the single case of cap vs. cab, the linguistic categories with which the stimulus set was constructed, fail to provide any key to the irregularities revealed in Table 5, presenting results of Spearman Rank Correlation tests on the amplitude data. One is prevented from an automatic adoption of the view that EMG amplitude measures are unreliable by the fact that high and consistent correlations are observed for the case of cool. A saving hypothesis could be constructed by supposing that the signals were more variable in the case of the stop sequence. Now, amplitude measures for the stop sequence are of generally higher amplitude than for the rounded vowel (for example, the range for right centre electrode for cool is 13 - 38.5, while for the signals from the right centre electrode for the stop sequence taken together the range is 2o.5 - 53). Experience suggests further that the durations of the EMG signals will be shorter in the case of stop consonant gestures. It might be suggested that EMG measures for fast, high-amplitude gestures might be more variable than for slow gestures with lower amplitudes, such differences perhaps reflecting a difference in control.

It must be remembered, however, that Mansell (1970b) showed both amplitude and duration measurements for [u] in the frame [hək _ tə] to be highly variable, and to show only a small positive correlation with each other.

In the absence of a satisfactory criterion for explaining the asymmetries of Table 5, and lacking satisfactory evidence for a saving hypothesis, it must be provisionally concluded that the reliability criterion of 3.2. (i) is not met. The reservation must be included, however, that this conclusion applies only to the stop series.

5.4. The 2nd criterion of reliability

Table 6 shows that the second criterion of reliability, given in 3.2. (ii) is met in all except one case. Further, since in this case, the comparison between amplitude scores at left centre electrode for cap and cam, the variance is not bilateral, we must conclude that this is a case of absolute unreliability. More evidence comes from a comparison between the patterns for cap and cam on Tables 3 and 5. In both cases the relationships for the left centre site and the right centre site and the left centre site and the left corner site show the same patterns of significance and lack of it. Although this result is at a relatively low level of significance and is found in the context of 35 satisfactory scores, I take it to demonstrate that there will be cases where unfortunate placement of an electrode may led to distictions between linguistic types on the basis of peak amplitude measures which will not reflect the activity of the lip as a whole. It must be noted that the rounded vowel in cool was not included in the tests reported in this section.

5.6. The cause of variation

Finally, since the reliability of the amplitude measures has not been shown, it must be noted that the possibility of a peripheral cause for signal variation is not ruled out. Further it is not possible on the basis of this experiment to distinguish between variation arising from factors at the electrode-tissue interface and variation in the events taking place beneath the electrode. If the former set of factors were responsible for variation, however, it is rather surprising that this study has shown that, with two exceptions, the range of variations from comparable parts of the lips is the same. In the case of the latter cause of variation, the explanatory problem that was the subject of Mansell (1970b) remains.

6. Conclusions

This study has investigated the dependence of EMG peak amplitude measures on electrode position, and the reliability of such measures as revealed by the degree of correlation among scores from electrode sites in a Common Reference derivation, and the degree of constancy across these sites in distinguishing between peak amplitude activity for different linguistic types. It was further possible to provide a crude measure of the plausibility of an "overshoot" hypothesis to account for EMG variation.

It was concluded that EMG peak amplitude scores are highly dependent on electrode positioning, with the highest amplitude being found closest to the midline of the lips for both stop consonants and spread vowel. The EMG peak amplitude measures for the stop consonants were judged unreliable, since the patterns of correlations among sites were complex, and not explicable in terms of the linguistic oppositions employed in deriving the stimulus array. Further it was possible in one case to make a presumably spurious distinction between linguistic types at one site, whereas the other sites in this case showed no difference in amplitude scores. It must be noted that the rounded vowel was not studied in the latter test, and in the former test of reliability proved to meet the criteria adequately.

The crude test of the overshoot hypothesis failed to find confirmatory evidence.

Finally, it is concluded that the degree to which lip activity as a whole can be studied in detail as a result of its use is sufficient proof of the usefulness of Common Reference deployment of electrodes in speech EMG research.

Acknowledgements:

The experimental work on which this paper is based was carried out in the Speech Research Laboratory of the Language Centre, University of Essex. I should like to thank Dr. M.A.A. Tatham for his help and cooperation in the running of the experiment. The processing of the data was carried out while the author was a visiting researcher at the Institute of Phonetics, University of Copenhagen, supported by a grant from the Rask-Ørsted Foundation. I should like to thank Professor Eli Fischer-Jørgensen and the staff of the Institute, especially Børge Frøkjær-Jensen, for placing their facilities at my disposal.

References

Cooper, R., Ossleton, J.W. and Shaw, J.C. 1969: EEG Technology. Butterworth.

Dedo, H. and Dunker E. 1966: "The Volume Conduction of Motor Unit Potentials", Electroenceph. Clin. Neurol. 2o, p. 6o8-613.

Hamilton, W.J. 1958: Textbook of Human Anatomy. Macmillan.

Hollinshead, W.A. 1968: Anatomy for Surgeons. Vol. I. The Head and Neck. 2nd. ed.

Mansell, P. 197oa: "Linguistic Parameters in Performance Models". University of Essex, Language Centre Occasional Papers, 8.

" 197ob: "On the Nature of EMG Variations". Proc. Essex Symposium on Speech Production Models, UELC Occasional Papers, 9.

Ohman, S., Leanderson, R. and Persson, A. 1965: "Electromyographic Studies of Facial Muscles During Speech", STL-QPSR 1/1966, p. 1-14.

Sicker, H., DuBrul and Lloyd, E. 197o: Oral Anatomy. 5th ed. C.V. Mosby.

Siegel, S 1956: *Nonparametric Statistics for the Behavioural Sciences.* Mcgraw Hill.

Sonesson, B. 1960: "On the Anatomy and Vibratory Pattern of the Human Vocal Folds", *Acta Oto-Laryng.* Suppl. 156.