Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 35, No. 4, 461–463, April, 1963

Electromyographic and Acoustic Study of the Production of Certain Final Clusters*

PETER F. MACNEILAGE

Haskins Laboratories, New York 17, New York
(Received 21 November 1962)

The possibility that speech may be perceived with reference to articulation prompted this preliminary study of muscle action during speech production. The question investigated was: what is the difference, if any, in the acoustic and articulatory characteristics of /f/ in utterance final position and /f/ embedded in final consonant clusters. The acoustic duration of /f/ friction, measured spectographically, was about twice as great in final position as in any other position. On the other hand, electromyograms were similar for all positions of /f/. The change in /f/ friction duration from final to prefinal position is therefore a contextual effect on acoustical properties of production alone. The motor gesture is more independent of context.

MUCH recent work at Haskins Laboratories has suggested that a knowledge of articulation processes may be important in understanding the perception of speech.[1–3] The specific suggestion is that influences attributable to a listener's articulatory habits may mediate between his reception of acoustic speech stimuli and his perceptual judgment of these stimuli. For instance, there is evidence that when the acoustic cues for certain speech sounds are synthetically made to vary continuously, subjects' judgments of the stimuli are sometimes discontinuous and categorical, and at other times continuous. They are discontinuous when "the acoustic cues are normally produced by discontinuous articulations"[3] (e.g., /b/ /d/ /g/) and continuous when the cues relate to continuous articulations (e.g., vowels). On the other hand, when the acoustic cues for a consonant such as /g/ (when followed by various vowels) form a discontinuous acoustical series, while the related articulation remains essentially the same, the perceptual judgment of /g/ also remains the same. These results suggest that judgments are being made in terms of the articulatory gestures which would have produced the acoustic stimuli, rather than in terms of the stimuli themselves.

There are other more general grounds for expecting an association between auditory speech stimuli and stimuli related to articulation. For instance, there is a good argument that speech develops through such an association.[4] Also, it is interesting to note that Hebb has presented evidence for a similar conclusion in the area of visual perception.[5] He states that "although the motor activations (related to eye movements) may be subliminal and do not always produce overt response, their role is essential in perception." In speech perception, as in vision, it seems likely that the motor action derived effects are, at least in the adult, not usually observable at the periphery when the perceptual events are occurring. However, this unobservability is not a critical disadvantage to an articulatory reference theory because, if perception is more intimately tied to articulation than to acoustic variables, the knowledge of what the relation is could well be found by study of articulation during production of speech.

The aspect of articulation most worth studying in this respect is, probably, not that of changes in vocal tract shape, but that of the muscle potential changes that control them. It is argued that the vocal tract changes

* This investigation was supported in part by P. H. S. Research Grant DE 01774–01 from the National Institute of Dental Research, Public Health Service.

[1] A. M. Liberman, J. Acoust. Soc. Am. 29, 117 (1957).
[2] F. S. Cooper, A. M. Liberman, K. S. Harris, and P. M. Grubb, 2nd International Congress on Cybernetics, 930 (1958).
[3] A. M. Liberman, F. S. Cooper, K. S. Harris, and P. F. MacNeilage, paper presented at the Stockholm Speech Communication Seminar (1962).
[4] F. H. Allport, *Social Psychology* (The Riverside Press, Cambridge, Massachusetts, 1924).
[5] D. O. Hebb, *The Organization of Behavior* (John Wiley & Sons, Inc., New York, 1949).

are complex resultants of what could be a rather simple series of motor commands, and electromyographic studies of motor action should allow direct inference of the nature of these commands.

Electromyographic work done so far encourages the hope of a simple motor description of articulation.[6,7] It has been found that labial stops /p/, /b/, and /m/ can be described in terms of the all-or-none action of three independent articulatory gestures localized at the lips, the velum, and the glottis. These gestures do not vary, depending on the context provided by spatially adjacent but simultaneous components of the utterance. The present study seeks to answer a related question. Does a speech gesture show a comparable independance of temporally adjacent gestures associated with other phonemes in successive speech output? In an attempt to answer this question for one specific instance, a comparison was made, in acoustic and electromyographic terms, between the phoneme /f/ when in utterance final position, and when embedded in various final clusters.

METHOD

The subjects were four adult research workers (3 male and 1 female). Two of the males were trained linguists, and all were speakers of some dialect of General American English. Indices of /f/ articulation were obtained by placing a suction cup electrode on the lip of each speaker and passing the primary muscle potential through an amplifier to a penwriter. Signals were also integrated with a decay time constant of 20 msec and then fed to the penwriters. In addition, the signal was sent to low-level preamplifiers whose output was recorded on magnetic tape.[6]

Electrode placement was at the level of the junction of the upper and lower lip, and $\frac{1}{4}$ in. from the vermillion border. This position was chosen because the production of /f/ involves the drawing outward of the corner of the mouth, and a number of muscle groups which could subserve this function lie adjacent to this area. The exact position of the electrode chosen in each case was that which gave a relatively discrete burst of muscle potentials during the production of /f/, and it was decided upon by trial and error before the experiment. Vocal cord output was recorded by another penwriter connected to a small vibration pickup fastened lateral to the larynx. The subjects' utterances were recorded on tape. The stimuli were the list of clusters given below.

if	ilf	imf
ifs	ilfs	imfs
ift	ilft	imft
ifts	ilfts	
ifθ	ilfθ	
ifθs	ilfθs	

[6] G. F. Lysaught, R. J. Rosov, and K. S. Harris, J. Acoust. Soc. Am. 33, 862A (1961).
[7] K. S. Harris, M. M. Schvey, and G. F. Lysaught, J. Acoust. Soc. Am. 34, 743A (1962).

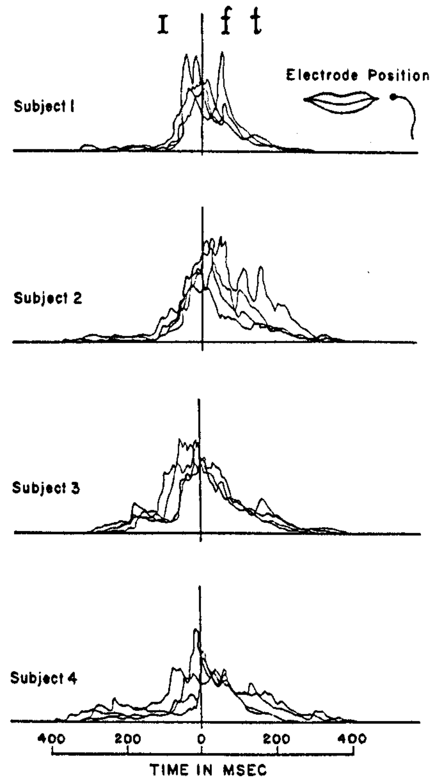

FIG. 1. Tracings of the integrated muscle potentials from each of the four subjects during four productions of the utterance /ift/. The vertical line indicates the offset of voicing.

/f/ is the only labial fricative among the speech sounds in these stimuli.

The subjects were asked to read, in a normal speaking voice, six lists of the above clusters, each list containing one of each cluster and each having a different random order of clusters. Subjects paused for about $1\frac{1}{2}$ sec between each utterance and for slightly longer between each list. Muscle potentials and vocal output were recorded throughout the test period.

Muscle potential patterns were measured by superimposing on tracing paper the outlines of a number of instances of the same utterance. The point of origin for these outlines was taken at the offset of voicing preceding /f/ friction in each response, a feature clearly visible in the voice trace and corresponding to the time of onset of /f/ friction visible in the spectrograms. The pattern of multiple traces obtained was then averaged by freehand drawing. The possibility of measurement error from this averaging technique was slight, as the traces showed much similarity from response to response. Figure 1 shows the consistency of the responses for each of the four subjects on a typical utterance. In this figure, and in the results presented later, the amplitude scale is different from subject to subject, as different gain settings were used. However, the voltage level reached was usually about 150–200 μV. Acoustical durations of the segments of the utterance were obtained by measuring spectrograms of the taped responses. The boundary between /f/ friction and /θ/ friction was often impossible to determine, so figures for the separate durations of these phonemes when they occurred together were not obtained.

RESULTS

It can be seen from Fig. 2 that the muscle potential patterns show a broad similarity for all utterances. Potentials begin from 200–250 msec before cessation of voicing, reach their peak near the point of cessation, and decline to zero over the next 300 msec. There are some slight systematic variations in the time courses of the potentials. The onset of muscle action is slightly earlier when /il/ and /im/ precede /f/ than when /i/ alone precedes /f/. Also the amount of muscle potential preceding /f/ friction is greater when /im/ precedes /f/ than in the other two cases, probably because of the labial action involved in production of /m/. However, the muscle pattern following the onset of /f/ friction is similar in all cases, regardless of what precedes or follows /f/ in the utterance.

In contrast to this similarity of muscle pattern, the main spectrographic finding is that the acoustic duration of /f/ friction in final position (about 275 msec) is approximately twice as great as its duration in any embedded position, although the different embedded positions show similar durations. This large difference is independent of the type of utterance segment which precedes /f/. No other contextual effects on acoustical duration occur. The utterance segments which precede /f/ and follow /f/ show consistent duration characteristics regardless of variations in their contexts. It is interesting to note that the duration of final /t/ closure is much less than that for final /f/, and similar to its duration when embedded.

DISCUSSION

From these results it can be concluded that the muscle action patterns associated with /f/ show considerable similarity whether /f/ occurred in final or any of the several embedded positions in clusters. The fact that this similarity existed when the reference point used for the comparison of muscle patterns was the cessation of voicing, means there was also an invariance in relative timing between the lip gesture for /f/ and the glottal gesture of devoicing regardless of whether the preceding segment was /i/, /l/, or /m/. These results are specially significant in contrast with the acoustical results in that the 2:1 difference in acoustical duration between final /f/ friction and embedded /f/ friction has no corelater

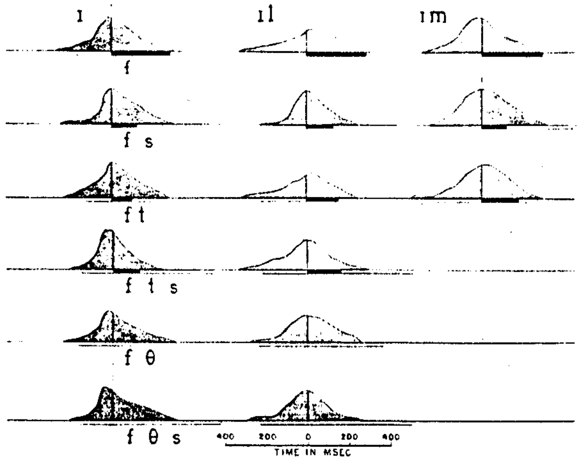

FIG. 2. Mean integrated muscle potentials (envelopes) and mean spectrographic durations (blocked portions beneath envelopes) of 4 subjects, associated with the production of various final clusters containing /f/. The vertical line indicates the offset of voicing.

in muscle pattern variation. The results suggest further that when technical problems of recording from more inaccessible muscles are overcome, similar results for other gestures showing prolonged acoustical duration in final position (e.g., /s/) may be found.

The present finding of relatively invariant muscle patterns is perhaps surprising in the light of the typical view of speech production obtained from visual observation of the process. A statement by Miller[8] seems representative of this view: "The vocal machinery does not produce phonemes the way a typewriter prints letters. The shapes of the resonating cavities change in a continuous sort of movement, with articulatory thrusts modified by the positions assumed for the preceding and following sounds." This view gives the impression of much more interdependance between articulatory events than the present findings suggest. However, this is no inconsistency but just what would be expected from the view mentioned earlier: that vocal tract shape changes are a complicated manifestation of what may be a relatively simple sequence of motor events. In conclusion, although many more studies with different speech gestures must precede any general statement, this study has shown that in one case at least, the motor dynamics of utterance in context show an encouraging amount of invariance, and may be more independent of context than their concomitants in acoustic output and vocal tract shape change. They may, therefore, provide a simpler description of speech production.

[8] G. A. Miller, *Language and Communication* (McGraw-Hill Book Company, Inc., New York, 1951).