XIII. SPEECH ANALYSIS

Prof. M. Halle
A. Cohen
G. W. Hughes
J.-P. A. Radley

A. INVESTIGATIONS OF FRICATIVE CONSONANTS

1. Spectral Measurements

The aim of the work reported here was to establish criteria by means of which it would be possible to decide whether a fricative consonant belonged to one of the following classes: /f/ or /v/; /s/ or /z/; /ʃ/ or /ʒ/. (When talking of these classes in what follows, we shall designate them by the voiceless member; thus /f/ will mean both /f/ and /v/.)

Available evidence indicates that the desired information is contained in the spectrum of the sound and that relatively little is contributed by the behavior of the formants in the adjacent vowels. (See the experiment reported below.) The measurements were made on 50-msec gated central portions of fricative consonants in about 200 English words spoken by four subjects (2 females and 2 males).

For purposes of filtering we used the set of low- and highpass filters with the very steep skirts which was described in the Quarterly Progress Report of April 15, 1955. The particular bands to be measured were selected after an examination of a large number of detailed energy spectra of fricative consonants. To eliminate the voicing component, all measurements were made with a 720-cps highpass filter. The following measurements were actually made.

1. The difference (in db) between the total energy in the sound (720 cps to 10,000 cps) and that in the band from 4200 cps to 10,000 cps.

2. The difference (in db) between the energy in the band from 720 cps to 6500 cps and the energy in the band from 720 cps to 2170 cps.

3. The frequency location of the maximum in the spectra between 1500 cps and 4000 cps.

4. The difference (in db) between the energy in a 500-cps band centered about the maximum (located in measurement 3) and that in the band from 720 cps to 1370 cps.

Measurement 1 indicates the presence or absence of a peak above 4 kc/sec. With the exception of samples from male speaker E, all /s/ had peaks above 4 kc/sec. Several /s/ of speaker E had peaks in the region between 3 and 4 kc/sec. No /ʃ/ had peaks in that region. Some /f/ also had peaks above 4 kc/sec. Thus, if measurement 1 was small, this could be taken as an indication that the sound was either /s/ or /f/; if it was large, this would indicate that the sound was either /f/ or /ʃ/.

Sounds giving small values for measurement 1 could be uniquely determined as being either /s/ or /f/. Small values for measurement 2, reflecting a concentration of energy

in the frequency region below 2 kc/sec, were always /f/; large values, /s/.

Sounds giving a large value for measurement 1 could be uniquely determined as being either /f/ or /ʃ/. Large positive values for measurement 4, reflecting the presence of a strong maximum in the region between 1500 cps and 4000 cps, were always produced by /ʃ/; small positive or negative values were produced by /f/.

The difficulty with speaker E resulted from the fact that most of his spectra for /s/ were shifted downward by about 1 kc/sec (if compared with /s/ of other speakers), often placing them in the region between 3 kc/sec and 4 kc/sec. If this fact is taken into consideration and measurement 1 is adjusted to compare the total energy in the sound with that in the region above 3 kc/sec (instead of 4 kc/sec as for the remaining speakers), then the data can be made to fit all the other criteria. That his speech is somewhat at variance with that of the other speakers can be seen in the results of the perceptual tests discussed in the following section.

M. Halle, G. W. Hughes

B. PERCEPTUAL TEST

1. Purposes and Procedure

An experiment was set up to determine the identifiability of gated portions of voiceless fricatives and to establish correlations between certain spectral properties of the sound presented and its identification by a group of listeners. Gated portions, each 50 msec long, of 46 /s/, 32 /f/, and 22 /ʃ/, taken from isolated words spoken by 1 female and 3 male speakers were recorded twice in random order on a test tape. This tape, containing 200 samples in all, was played through a sound system having a reasonably flat response up to 10 kc/sec and presented to a group of 10 listeners under conditions of high S/N ratio, with instructions to identify every sample as one of the three fricatives /f/, /s/, or /ʃ/. The samples were spaced at 6-sec intervals with a 45-sec pause after each group of 25.

2. Learning

Since each stimulus was presented twice, the change in agreement among listeners from the first presentation to the second presentation was taken as a rough indication of the learning that had gone on in the interim. In 51 out of 100 cases there was an increase in agreement; in 23 cases there was a decrease in agreement, while in 26 cases there was no change; thus we conclude that a certain amount of learning took place.

3. Identification

The following table shows the degree of agreement between the judgments of the listeners and the intention of the speaker.

In 26 cases	10 (all) listeners agreed with speaker
22	9
41	8
30	7
37	6
156	6 or more

In the remaining 44 instances there were 16 cases in which 6 or more listeners agreed on the identity of a sound, but disagreed with the speaker.

We conclude from this that it is possible to obtain correct identifications of fricative consonants in isolation.

4. Speakers

Table XIII-1 presents the responses of the listeners to the different phonemes uttered by the speakers.

Table XIII-1

Stimuli		Speaker E			Speaker T			Speaker R			Speaker H		
		s	f	ʃ	s	f	ʃ	s	f	ʃ	s	f	ʃ
Responses	s	107	23	13	184	52	22	112	26	22	175	27	13
	f	17	143	6	26	97	7	13	111	29	29	109	8
	ʃ	156	14	61	10	11	71	75	23	89	16	4	99
Total		280	180	80	220	160	100	200	160	140	220	140	120

Table XIII-1 shows that with the exception of /s/ uttered by speaker E the great majority of the judgments agreed with the speakers as to the identity of the phonemes. The reasons for the deviations in the case of speaker E are advanced in the first part of this report. Identical reasons hold for the relatively high number of /ʃ/ judgments for /s/ uttered by speaker R (who was not among the speakers studied in the first part of this report). The high number of /s/ judgments for /f/ uttered by speaker T (female) is due to the presence of peaks in the region above 4 kc/sec in many of T's /f/.

5. Physical Properties of Stimuli

In this section we shall discuss the listeners' responses in terms of the physical properties of the stimuli as represented by the four measurements described in the first part of this report.

a. Measurement 1

Value of measurement 1	Responses			
	/f/	/s/	/ʃ/	Total
0	102	216	22	340
0.1-1.9	169	343	88	600
2.0-3.9	195	85	100	380
4.0-5.9	67	52	161	280
6.0-7.9	40	45	115	200
⩾ 8.0	28	42	130	200

It appears from this table that the percentage of /ʃ/ responses increases with an increase in measurement 1, whereas the percentage of /s/ responses decreases. The percentage of /f/ responses increases initially, and after reaching a maximum value between 2.0 and 3.9, decreases. The interpretation of the data is quite simple if we recall that low values of measurement 1 show a concentration of energy in frequencies above 4 kc/sec, while high values show a concentration of energy below 4 kc/sec.

b. Measurement 2

Value of measurement 2	Responses			
	/f/	/s/	/ʃ/	Total
⩽ 5	401	134	185	720
> 5 ⩽ 10	102	59	79	240
> 10 ⩽ 15	17	87	176	280
> 15 ⩽ 20	26	195	139	360
> 20	46	314	40	400

Predominance of energy in the low-frequency region is indicated by small values of measurement 2; /f/ was the most frequent response for stimuli where the energy was concentrated in the area below 2000 cps. For high values of measurement 2, where, in other words, there was a great amount of energy in the high-frequency region, the most frequent response was /s/. For intermediate values of measurement 2, /ʃ/ was the most frequent response.

c. Measurement 3

Peak frequency	Responses			
	/f/	/s/	/ʃ/	Total
> 1000 < 1500	180	44	56	280
> 1500 < 2000	35	19	46	100
> 2000 < 3000	19	30	151	200
> 3000 < 4000	38	117	225	380

As was to be expected from measurement 2, peak frequencies in the low region would lead to frequent /f/ responses and peaks in intermediate regions would lead to /ʃ/ responses. The absence of a large number of /s/ responses results from the fact that peaks above 4 kc/sec were not considered in this tabulation (see discussion of measurement 1).

d. Measurement 4

Value of measurement 4	Responses			
	/f/	/s/	/ʃ/	Total
< 0	178	39	23	240
> 0 ⩽ 10	53	32	95	180
> 10 ⩽ 20	15	46	159	220
> 20	26	93	201	320

Again we see that the predominance of energy in the low-frequency region favors /f/ judgment. A strong concentration of energy round the peak frequency region, indicated by high values for this measurement, gives rise to a majority of /ʃ/ judgments.

A. Cohen