Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

On the synthesis and perception of voiceless fricatives

Mártony, J.

journal: STL-QPSR
volume: 3
number: 1
year: 1962
pages: 017-022

KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

II. SPEECH PERCEPTION

A. ON THE SYNTHESIS AND PERCEPTION OF VOICELESS FRICATIVES

In the following an investigation of the Swedish voiceless fricatives /f, s, ʂ, ç/ is reported. These sounds are generally classified as labiodental, dental, prepalatal [x], and palatal respectively.

Spectra of these fricative sounds were analyzed and described mathematically in terms of poles and zeros. An attempt was made to simplify and approximate this mathematical description with a fewer number of poles and zeros. Moreover, a perceptual study was carried out intended to reveal something of the relative influence of spectral shape, intensity, and adjacent vowel transitions on identification. The results of this experiment are relevant to the choice of quantization steps in the vocoder synthesis of fricative spectra and intensities and is also indicative of the perceptual implications of errors in formant tracking, intensity, and spectrum measurements.

VCV-words were used for analysis V standing for [a]. All the utterances were produced by the same speaker. Spectral sections were made on the spectrum analyzer RASSLAN using 125 c/s analyzing filters and a time constant of 80 msec.

This choice of analyzing and smoothing filters gives rise to an uncertainty in the ordinate values of $\sigma_y = 2.5$ dB [1]. Fig. II-1 shows the measured spectra which were matched according to the technique described in an earlier QPSR [2]. A maximally accurate matching calls for a fairly great number of poles and zeros. Table II-1 indicates frequencies and bandwidths of the poles and zeros. These spectra deviate from the natural samples by ± 5 dB in the frequency region where the matching was done (600 c/s → ~ 7 kc/s).

An approximate description of the spectrum can be performed with 2 poles and 1 zero. No attention is here paid to closelying poles and zeros which usually have the effect of neutralizing each other and appear in the spectrum as a small dip followed by a peak. Such a zero-pole pair raises the level at higher frequencies but little.

A series of expert judgments indicated that such approximations are permissible from a perceptual point of view. Similar results have been obtained by Heinz and Stevens [3].

---

x) Generally retroflex. Spectral energy of lower frequency than the palatal.

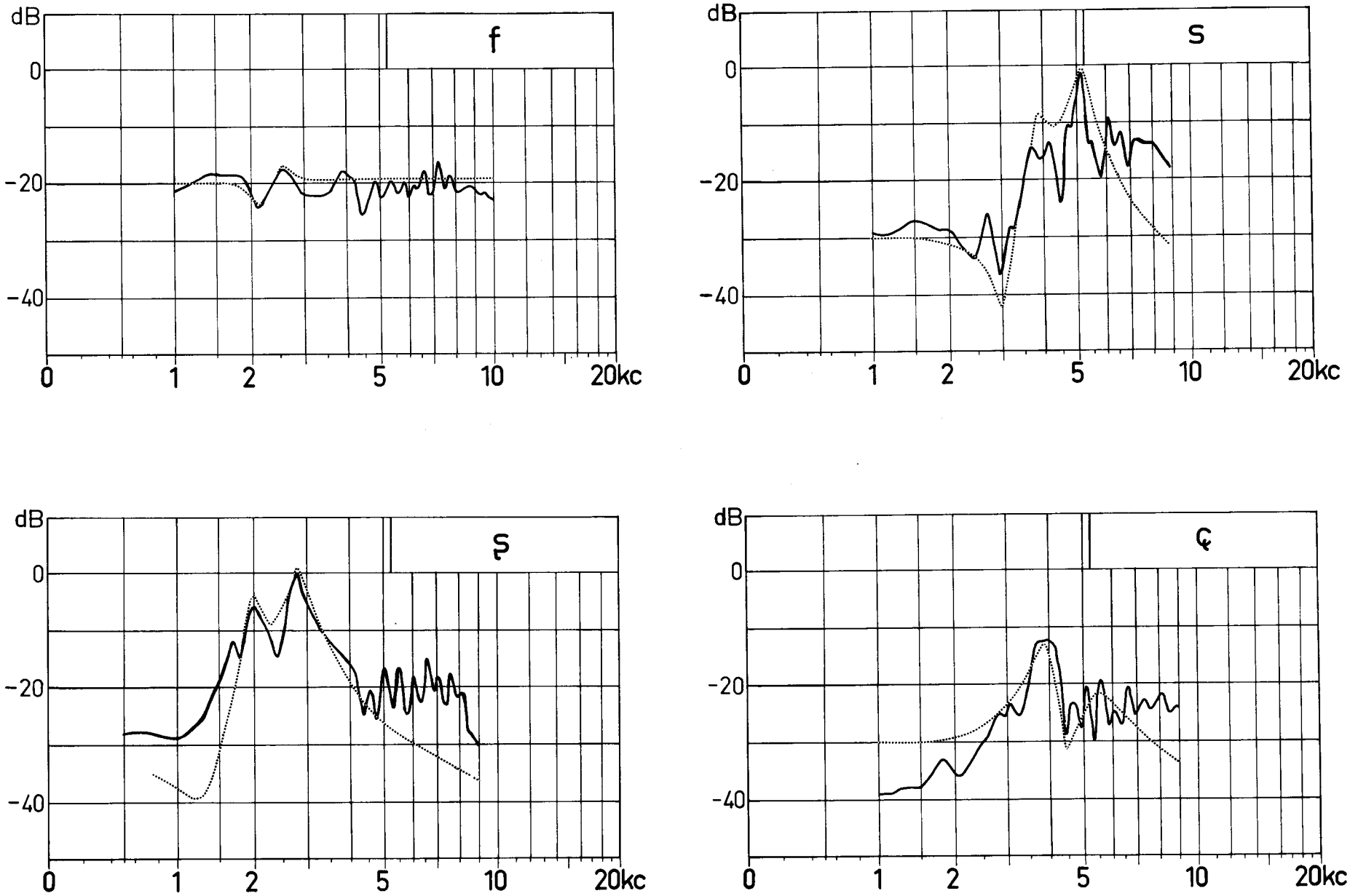

Fig. II-1. Spectrum of [f], [s], [ʂ], [ɕ]. Measured spectra are represented by solid lines, approximations in terms of two poles and one zero by dotted lines.

The approximation utilizes the poles and zeros marked with x in Table II-1. The associated spectra are compared with the natural samples in Fig. II-1.

TABLE II-1

	[f]			[s]	
	f c/s	B c/s		f c/s	B c/s
zero	2150	250 x	z	2400	250
pole	2450	250 x	p	2800	250
zero	3400	300	z	3100	250 x
pole	3700	300	p	3750	250
zero	4750	450	z	3950	250 x
pole	5400	450	p	4200	250
			z	4350	250
			p	4700	250
			p	5200	250 x

	[ʂ]			[ç]	
	f c/s	B c/s		f c/s	B c/s
z	1250	250 x	z	1450	250
p	1650	250	p	1800	250
z	1750	250	z	2100	400
p	2000	250 x	p	2800	400
z	2300	250	z	3300	300
p	2500	250	p	3500	300
p	2900	250 x	p	4000	300 x
			z	4500	600 x
			p	5600	700 x

TABLE II-1. Frequency and bandwidth of poles and zeros for description of the fricatives [f, s, ʂ, ç]. Poles and zeros marked with crosses enter the approximation used for synthesis. It is to be noted that the second pole of /f/ is above the region in which the matching was done.

To evaluate the relative importance of cues for the identification of fricatives the above syllables were synthetized with systematic variation of vowel transitions, spectral properties, and intensity. A first series involved no transitions but changes in intensity and the spectral pattern as determined by 2 poles and 1 zero.

The intensity variations were determined by the level differences between vowel and consonant in the four words i.e.,

in /afa/ the consonant was 30 dB below the vowel
" /asa/ " " " 10 dB " " "
" /aşa/ " " " 7 dB " " "
" /aça/ " " " 17 dB " " "

f, s, ş, ç for intensity indicate that the noise level is 30 dB, 10 dB, 7 dB, 17 dB below the vowel.

This series was thus made up of 16 stimuli. These stimuli are called AB. The first letter indicates the type of fricative spectrum the second letter the intensity of the noise. The vowels contain no transitional segments.

In a second series also the vowel transitions were varied. 4 new patterns to combine with the 16 already made would in all have given 64 stimuli. To reduce this figure two intensities were chosen, one value belonging to the fricative intended by the specific spectral pattern and the other value associated with the fricative intended by the transitions. In other words there were 28 different stimuli in the second series. The stimuli of this series are denoted by ABA and ABB. The first letter refers to the vowel transitions the second to the noise spectrum and the third to the intensity of the noise.

The remaining parametric values were coded in accordance with the human samples.

The two series were presented in random order to a group of 12 phonetically untrained listeners. In each series every stimulus was repeated 4 times. Thus 48 responses per item were obtained. The task of the subjects was to identify the fricative and to make a judgment as to the quality of the sample. They were instructed to indicate whether the quality was good (quality 3), not good but identifiable (quality 2),

or bad and hard to identify (quality 1). The quality judgments are of course to some extent also dependent on the whole synthetic stimulus and not only on those parts contributing to the identification of the fricatives. Histograms in Fig. II-2 and Fig. II-3 show the result of the listening test. Each phonemic response was multiplied by the associated quality judgment.

Series 1. No transitions. The identification is mainly based on the spectrum of the fricative. Intensity is of minor importance except in cases where the intensity deviates more than ca. 15 dB from that of the original sample. The responses favor /s/ when fs, fʂ are presented whereas the number of /s/-responses decrease for ff or fç.

The number of /ç/-responses are low. The majority of responses to spectra intended as /ç/ were /ʂ/. The reason for this may be attributed to the absence of appropriate transitions and the close similarity of /ʂ/-spectra to /ç/-spectra.

Series 2. For stimuli of ABB-character the main cue for /s/ and /ʂ/ seems to be the spectrum and intensity. This is partially true for /f/ as well. The only case when the spectral properties and the intensity do not suffice to yield good identification is in combinations with /ç/-transitions. Particularly in the case of friction of /f/-type (stimulus çff) this effect is pronounced. The greatest number of responses favor not /f/ but /s/ and /ç/ in about equal proportions.

For stimuli in which the intensity and the transitions refer to the same phoneme (ABA) we get /s/- and /ʂ/-responses provided that the spectra pertain to these sounds and we obtain /f/-responses if transitions and intensity refer to /f/. Stimulus fsf is identified as /s/ whereas fʂf gives /f/. (The quality of these sounds is generally considered to be poor. This shows that /s/ responses are more resistant to variations in intensity than /ʂ/-responses.)

A comparison of the responses to ABB and ABA for B=/f/ shows low intensity to be a crucial cue for the identification of this sound.

The same effect appears in series I i.e., an intensity deviation of more than 15 dB from the original intensity of the noise impairs identification.

The effect of including transitions in the stimuli appears from a comparison of the results of series I and series II. Only when transitions and friction refer to the same phoneme the results improve especially quality judgments. Since these two series were presented separately comparison is not quite permissible.

The /ç/-stimuli seem to have caused our listeners a great deal of trouble. In Swedish /ç/ does not appear often in intervocalic position. Thus the low conditional probabilities for /ç/ in our test may account for the fact that this response was avoided. But the test was also taken by a phonetically trained subject who responded by /ç/ and 3 for good quality to all the synthetic /aça/ (ççç) except in one case which was rated as not good but intelligible (quality 2) (92 %).

Essential deviations between the responses of the group and the phonetician occurred only regarding /ç/ which supports the above hypothesis that linguistic structure affects the results.

The validity of possible conclusions being directly dependent on the success of the synthesis in copying the original sample care was taken to ensure as good agreement between these as possible.

The results can be summed up as follows. A good /f/ presupposes appropriate vowel transitions and low intensity. An exception is fsf which, in spite of low intensity, is identified as /s/ on basis of the spectrum. If the transitional part is erroneous or missing the spectrum becomes an important cue in addition to the low intensity. But the identification is difficult and the quality is rated as poor. For large deviations from the original transitional pattern as in çff there are practically no /f/-responses. Similar results were obtained by K. Harris for American English fricatives (4)(5)(6).

/s/-responses seem to depend chiefly on the spectral properties. Great deviations in intensity and the transitional pattern are tolerable. They affect the quality more than the intelligibility of the sound.

In the study reported by K. Harris the /s/-responses seem to be more dependent on intensity than in our test. Possibly this might

be due to the grosser approximation of /s/-spectra used by Harris (HP-filtered white noise).

The same remarks can be made for /ʂ/ i.e., the identification depends primarily on the fricative spectrum.

In the case of /ç/ the formant transitions serves as primary cues. This statement is based on the responses obtained from the phonetician.

J. Mártony

(1) Holmes, J.N. and Liljencrants, J.: "Analysis of random signals", STL-QPSR 2/1960, pp. 3-4.

(2) Fant, G. and Mártony, J.: "Pole-zero matching techniques", STL-QPSR 1/1960, pp. 14-16.

(3) Heinz, J.M. and Stevens, K.N.: "On the properties of voiceless fricative consonants", J.Acoust.Soc.Am. 33 (1961) pp. 589-596.

(4) Harris, K.S.: "Cues for the identification of the fricative of American English" (Abstract), J.Acoust.Soc.Am. 26 (1954) p. 952.

(5) Harris, K.S.: "Some acoustic cues for the fricative consonants" (Abstract), J.Acoust.Soc.Am. 28 (1956) p. 160.

(6) Harris, K.S.: "Cues for the discrimination of American English fricatives in spoken syllables", Language and Speech 1 (1958) pp. 1-7.

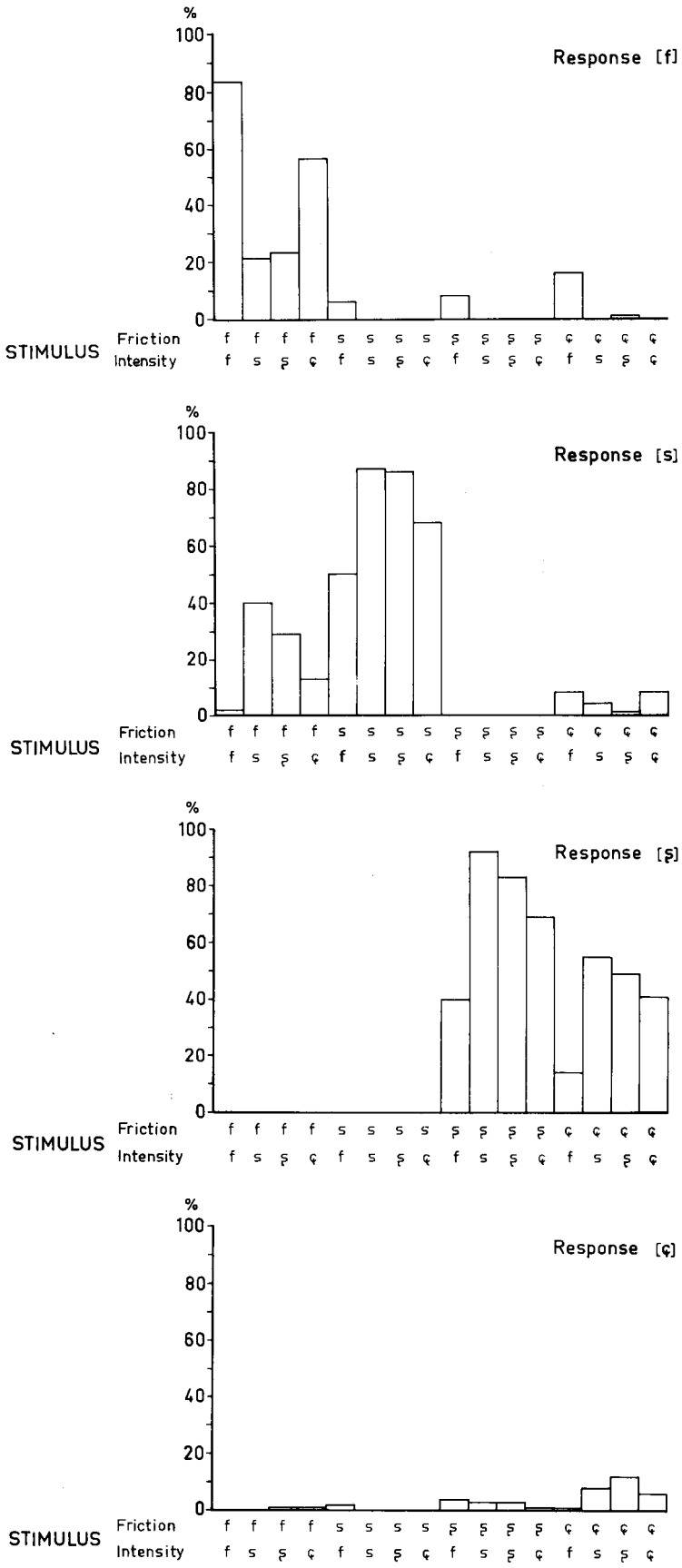

Fig. II-2. **Histograms of responses to stimuli in which the parameters varied were the spectral shape of the noise, the intensity of the noise segment relative to that of the vowel. There were no vowel transitions. The percentage on the ordinate was calculated from the responses weighted by a quality factor i.e., 100 % corresponds to the case when all presentation of a stimulus were consistently identified and designated as quality 3 (good).**

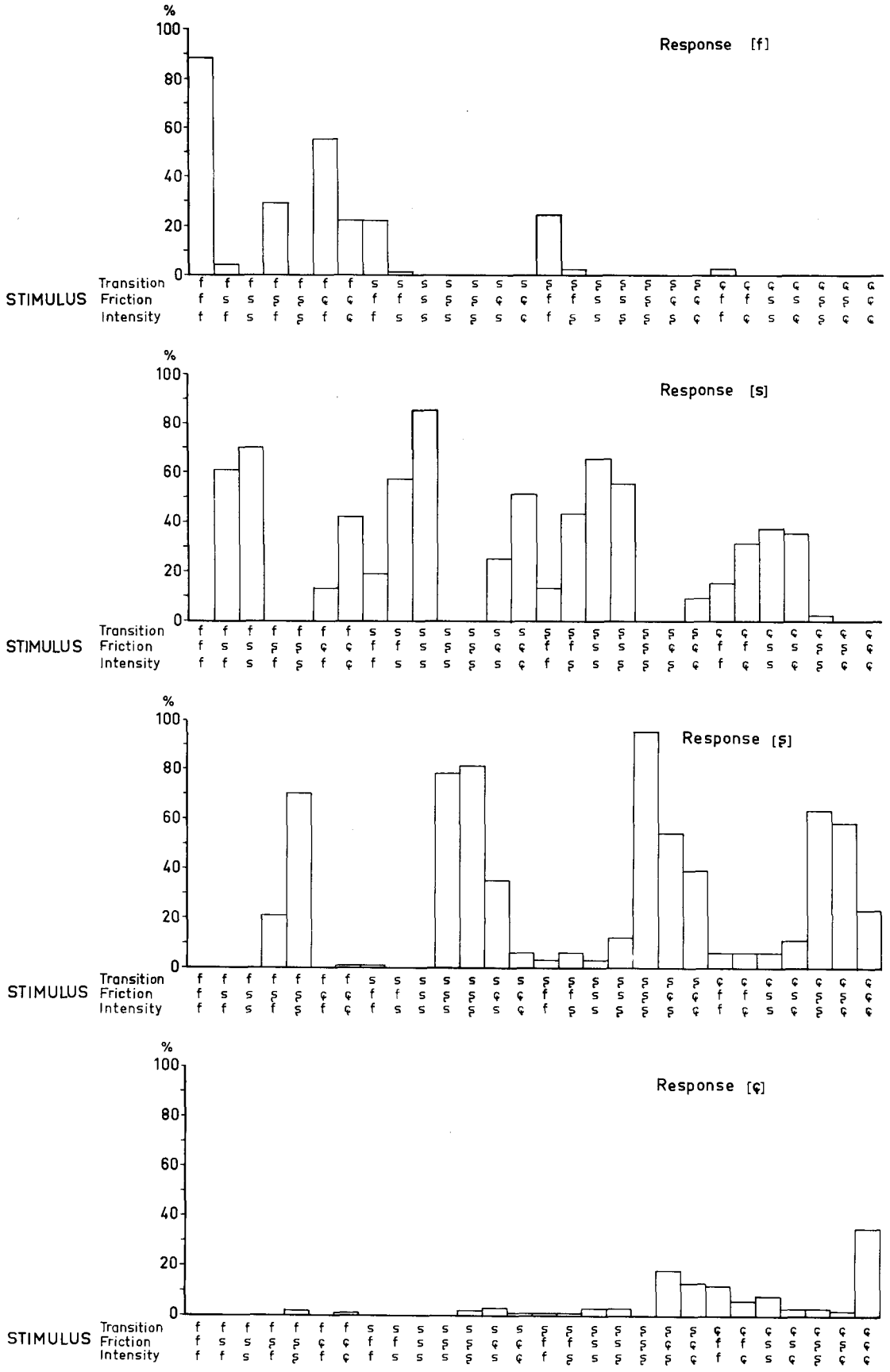

Fig. II-3. Histogram of responses to stimuli in which the parameters varied were the formant transitions of the vowel, the spectral shape of the friction and the intensity difference between the noise and the vowel. The percentage on the ordinate was calculated from the responses weighted by a quality i.e., 100 % corresponds to the case when all presentations of a stimulus were consistently identified and designated as quality 3 (good).