Working Papers 25 1983
Linguistics-Phonetics
Lund University, 195-211

ACOUSTIC ANALYSIS OF CHINESE FRICATIVES AND AFFRICATES

Jan-Olof Svantesson

The purpose of this investigation is to find a method of analyzing fricatives (and the fricative phase of affricates) which makes it possible to describe the available "fricative space" of a language, and to apply this method to Standard Chinese.

For this purpose parameters for characterizing fricative spectra are suggested and applied to the Chinese fricatives. Using these parameters the Chinese fricatives (and fricative components of affricates) as said by different speakers can be compared, and also the totality of the Chinese fricatives can be compared to those of other languages, such as Arabic, which has been investigated with the same methods (Norlin 1983). A comparison with the Swedish fricatives is also made, although this is difficult due to the lack of comparable data. Nartey 1982 uses a similar method of analyzing fricative spectra but the Chinese fricatives are not analyzed by him.

The fricatives and affricates in Standard Chinese

Standard Chinese has six fricatives and six affricates:

f	s	sh [ʂ]	x [ɕ]	h [x]
		r [ʐ]		
	z[ts]	zh [tʂ]	j [tɕ]	
	c[tsh]	ch [tʂh]	q [tɕh]	

The official pīnyīn transcription is used (with IPA symbols in brackets if they differ).

There are restrictions on the possible combinations of consonants and a following vowel, so that only these combinations occur:

	___ i [i]	___ i[ɿ] ___ i[ʅ]	___ ü[y]	___u	___uV	___V (other vowel)
s sh r z c zh ch	–	+	–	+	+	+
x j q	+	–	+	–	–	–
h	–	–	–	+	+	+
f	–	–	–	+	–	+

The palatals (x, j and q) are thus in complementary distribution with h and f, and in nearly complementary distribution with s, sh, etc., the only contrasts in the phonemization implied by the pīnyīn transcription being before i. The allophones of i after the fricatives and affricates are:

xi [çi]   si [sɿ]   shi [ʂʅ]   ri [ʐʅ]
ji [tçi]   zi [tsɿ]   zhi [tʂʅ]
qi [tçhi]   ci [tshɿ]   chi [tʂhʅ]

Thus it is possible to regard [ɿ] and [ʅ] as allophones of a phoneme which is distinct from i, and to let the palatals be allophones of one of the other series.

On the other hand, there are pairs of syllables such as:

sa [sa]   xia [çia]~[ça]
sao [sao]   xiao [çiao]~[çao]

where the i written in the transcription can be regarded as a vowel onglide conditioned by the palatal, so that these pairs may be seen as minimal pairs for [s] and [ç] . Zhāng et al. 1982 have shown that there is less perceptual confusion between s and x than between s and sh although, as shown below, the acoustic difference between s and x is smaller than that between s and sh. Obviously the different environments of s and x make perception easier.

Historically the palatals have developed from alveolars and velars in palatalizing environments:

[*s], [*x]> x [ɕ]
[*ts],[*k]> j [tɕ]
[*tsh], [*kh]> q [tɕh]

Procedure

The acoustic properties of the fricatives and affricates were investigated using four informants, who said each sound twice. The fricatives and affricates were said in word-initial position in a sentence frame, so that they were preceded and followed by the vowel [a] . The palatals (x, j and q), which do not occur before a were said before ia (see above).

Four male native speakers of Standard Chinese were used. Two of them (C and D) were born and grew up in Běijīng, one (B) was born in Jiāngsū province and moved to Běijīng in his sixth year and the fourth (A) was born in Liáoníng in northeast China and moved to Běijīng in his teens. All speak Standard Chinese with Běijīng pronunciation.

The recordings were made in sound-treated rooms in Lund and in Stockholm.

Analysis

Fourier transform (FFT) spectra were made in the middle of each fricative and in the fricative components of the affricates. The sampling frequency was 20 kHz, and the sampling time 26.5 ms. By running the tape at half speed when sampling, spectra in the frequency range up to 10 kHz could be made.

The FFT spectra were converted into critical band spectra according to Schroeder et al. 1979. The critical bands were determined by Schroeder's formula

$$f_n = 650 \sinh(n/7)$$

where f_n is the upper boundary of band n.

The spectrum is thus described as 24 bands, with bandwidths of approximately 100 Hz below 500 Hz, and of about 1/6 of the center frequency above 1000 Hz. The critical bands computed by this formula are:

n	f_n	n	f_n
1	93	13	2031
2	188	14	2357
3	287	15	2732
4	392	16	3136
5	505	17	3658
6	628	18	4228
7	764	19	4884
8	915	20	5640
9	1086	21	6512
10	1278	22	7516
11	1497	23	8674
12	1746	24	10010

The mean level (in dB) within each critical band was estimated, and the spectra were redrawn as histograms with each critical band represented as a bar with constant breadth, and with the baseline at -30 dB. This amounts to a rescaling of the spectra to an auditively more correct form, since each critical band corresponds to an equal distance (1.5 mm) on the basilar membrane, or to 1200 primary auditory nerv fibres (Schroeder et al. 1979)

For practical reasons, only bands 2-24 were used.

For the characterization of the spectra of the different fricatives, we propose to use the center of gravity and the dispersion of the critical band spectra, and also the mean intensity level (in dB), as computed by the following formulas:

center of gravity $$m = \sum_{n=2}^{24} n \cdot 10^{(x_n/10)}/F$$

dispersion $$s = \left(\sum_{n=2}^{24} (n-m)^2 \cdot 10^{(x_n/10)}/F\right)^{1/2}$$

mean intensity level $\bar{x} = 10 \log (F/23)$

where $$F = \sum_{n=2}^{24} 10^{(x_n/10)}$$

and x_n is the mean level (in dB) in band n, as estimated from the FFT spectra.

The center of gravity is a measure of the overall pitch level of the spectrum, and the dispersion can be regarded as a measure of its flatness.

Results

Figure 1 shows oscillograms of the sound waves, Fourier transform spectra in linear scale, the same spectra in logarithmic units (dB), and critical band spectra of each fricative.

Table 1 gives the centers of gravity of the critical band spectra (measured in critical band units, and also given in Hz), the dispersion, and the mean intensity level (in dB). The mean intensities are given as deviation from the average for each series of fricatives (read at the same occasion), in order to make them roughly comparable also between different speakers.

Figure 1 shows that s and x in Chinese are characterized by a high peak in the upper part of the spectra (bands 20-23 = 5-9 kHz). They differ by s having a more abrupt fall than x down to the level in bands 17-19 (3-5 kHz). This is reflected in the lower centers of gravity and greater dispersions of x than of s (Table 1). The s of speaker D differs from the other s by having a lower center of gravity and much greater dispersion.

In the spectra of sh, the higher levels of the spectra go further down (to bands 14-15 = 2-3 kHz), and the high level area is relatively large and flat. Thus the center of gravity falls further down, and the dispersion increases compared to s and x.

The intensity of s is somewhat lower than that of sh and x.

The spectrum of f is flatter and has a lower level than these of the sibilants. This is reflected in a greater dispersion, and in a lower mean intensity. Also h usually has greater dispersion, and lower intensity than the sibilants, and lies between them and f in these respects.

The only voiced fricative r differs markedly from the others by having an energy concentration in the low bands (below band 9 = 1 kHz) and also a peak in the area around band 15.

The fricative components of the affricates are generally similar to the corresponding fricatives (i.e. the fricative components of z and c are similar to s, those of j and q are similar to x, and those of zh and ch are similar to sh). Also here speaker D differs: his zh and ch have a lower center of gravity and higher dispersion than his sh.

In Figure 2 the dispersion is plotted against the center of gravity for each fricative. This is one way of representing the fricative space of Chinese in a way which makes comparison with other languages possible.

In Figure 3 an alternative representation of the fricative space is given. Here the mean intensity level over the critical bands is plotted against the center of gravity.

On both diagrams the individual fricatives are fairly well separated, but there is some overlapping, especially between different speakers, which may indicate that the perception of fricatives involves normalizing between different speakers, as is the case for vowels.

Discussion

The method of making critical band spectra and computing the proposed parameters from them makes it possible to characterize the place of the fricatives of a language within the available fricative space, and to compare the fricatives of different languages.

The fricatives space of Chinese, as defined by the center of gravity and dispersion (Fig 2) is rather crowded in the voiceless sibilants area (low dispersion and relatively high center of gravity), and there is less variation between different productions of the same sibilant than with h and especially r, which is the only voiced fricative.

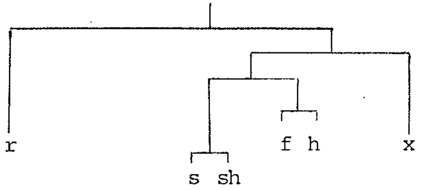
Arabic has also been analyzed with the same methods (Norlin 1983). Comparison of the three voiceless sibilants in Arabic (s, ṣ, and š) with the Chinese ones show that, at least for some speakers, Chinese s is more high-frequent than Arabic s, and that Arabic š and Chinese sh lie in approximately the same area. The high dispersion and low center of gravity area is more utilized by Arabic, which has 7 voiceless and 4 voiced fricatives, than by Chinese.

Swedish also has three sibilants which are roughly comparable to the Chinese ones. Available data on Swedish fricative spectra

(Lindblad 1980) are not directly comparable to our data, but judging from the spectra published by Lindblad it appears that the Swedish sibilants are related to each other in a way similar to Chinese ones, i.e. that the center of gravity decreases and the dispersion increases in the series [s] - [ɕ] - [ʃ] (or [ʂ]; other variants of Swedish /ʃ/ are not comparable). It also seems that the Chinese sibilants have the rise into the high-level area of the spectrum higher up than the Swedish counterparts (which) probably implies that they have higher centers of gravity).

Zhang et al. 1982: 201 give a hierarchical clustering diagram, which shows the perceptual similarity (as measured by the tendency to be confused in a perception test) between Chinese initial consonants.

If the fricatives are isolated from this diagram it becomes:



This can be compared to the dispersion/center of gravity plot (Fig 2). In both diagrams the voiced r is clearly set apart from the others (by its low center of gravity in Fig 2). The pairs s, sh and f, are also well separated in both diagrams (by the dispersion parameter in Fig 2).

The position of x [ɕ] differs markedly in the two diagrams, however. Acoustically it is similar to s and sh , and comes between them in Figure 2 but perceptually x is clearly separated from s and sh. As noted above, this discrepancy can be explained by the nearly complementary distribution of x and s or sh.

The features compact/diffuse and grave/acute are defined by Jakobson, Fant and Halle 1952 in terms of spectral properties, and measures similar to those proposed here are discussed by them.

They define compact phonemes in the following way:

> Compact phonemes are characterized by the relative predominance of one centrally located formant region (or formant).

> They are opposed to diffuse phonemes in which one or more non-cėntral formants or formant regions predominate."

They also suggest that the second moment about the mean (i.e. a measure similar to the dispersion measure used here) should be used as a measure of compactness. From the definition and the examples given by them (/ʃ/ compact, /s/ diffuse for instance), it appears that compactness is related to a non-extreme value of the center of gravity (and perhaps a not too high dispersion).

The feature grave/acute is defined as:

> "... the predominance of one side of the significant part of the spectrum over the other. When the lower side of the spectrum predominates, the phoneme is labeled grave; when the upper side predominates, we term the phoneme acute."

As a measure of this, the third moment about the center of area is suggested (and also the center of area, a measure similar to the center of gravity as used here). It seems that this feature cannot be directly related to any of the parameters used here. The term "significant part of the spectrum" used in the definition seems to imply that a local measure of the center of gravity could be used to distinguish pairs of phonemes which differ in this feature, rather than the "global" center of gravity used here.

Acknowledgements

The work reported here is supported by the Swedish Council for Research in the Humanities and Social Sciences.

References

Jakobson, R., Fant, G., and Halle, M. 1952. Preliminaries to speech analysis. 2nd pr. Cambridge, Mass.: MIT Press

Lindblad, P. 1980. Svenskans sje- och tje-ljud i ett allmänfonetiskt perspektiv. Lund: Gleerup

Nartey, J. 1982. On fricative phones and phonemes. UCLA Working Papers in Phonetics 55, Los Angeles

Norlin, K. 1983. Acoustic analysis of the fricatives in Cairo Arabic. To be published in Working Papers, Phonetics Laboratory, Lund

Schroeder, M.R., Atal, B.S., and Hall, J.L 1979. Objective measure of certain speech signal degredations based on masking properties of human audititory perception. In Björn Lindblom and Sven Öhman, (Ed.) Frontiers of speech communication research, London: Academic Press 1979, pp 217-229

Zhāng, J., Lǚ, S. and Qí, S. 1982. A cluster analysis of the perceptual features of Chinese speech sounds. Journal of Chinese Linguistics, Berkeley, 10, pp 189-206

Table 1. Center of gravity, dispersion and mean intensity level of the critical band spectra of fricatives and fricative components of affricates in Standard Chinese.

	Speaker	Center of gravity		Dispersion	Mean intensity
		crit. band units	Hz	crit. band units	dB
s	A	22.31	7314	1.24	3.74
		22.12	7118	1.06	5.84
	B	20.66	5771	1.26	3.83
		20.84	5923	1.44	1.57
	C	21.48	6493	2.34	-2.79
	D	19.90	5174	2.82	4.41
		18.53	4246	4.57	-0.89
x	A	19.84	5129	1.95	6.28
		21.38	6400	1.69	4.22
	B	19.33	4766	1.61	7.22
		19.73	5049	1.65	9.10
	C	19.82	5114	2.18	3.10
		20.93	6000	1.74	3.04
	D	20.29	5472	1.97	4.41
		19.82	5114	2.70	0.16
sh	A	17.11	3456	2.02	8.76
		16.54	3181	2.35	5.59
	B	17.96	3910	2.69	6.81
		19.52	4898	2.32	6.08
	C	17.41	3610	2.71	4.97
		16.17	3014	1.98	7.35
	D	16.05	2961	2.95	3.24
		17.49	3652	2.50	3.60
h	A	19.45	4849	2.98	-2.18
		17.67	3749	2.70	2.23
	B	14.43	2333	4.45	-2.63
		17.35	3579	3.50	-1.93
	C	19.79	5092	2.22	3.13
		15.64	2789	3.48	2.21
	D	12.78	1822	2.57	3.73
		13.13	1921	2.53	2.89
f	A	20.25	5441	3.57	-7.83
		20.10	5325	3.14	-10.23
	B	18.03	3950	3.98	-6.43
		19.51	4891	3.48	-6.85
	C	17.97	3916	3.38	-6.58
		19.25	4711	3.70	-9.18
	D	15.84	2872	3.77	-11.36
		15.40	2692	3.05	-5.87
r	A	11.45	1485	3.55	-8.80
		12.56	1762	2.36	-7.66
	B	13.86	2143	4.52	-8.79
		11.46	1488	5.39	-7.98
	C	4.83	428	1.56	-1.83
		6.35	609	3.94	2.02
	D	6.08	575	4.56	-4.40
		5.29	480	2.88	0.14

Table 1 (cont.)

	Speaker	Center of gravity		Dispersion
z	A	22.52	7538	0.84
	B	21.15	6192	1.88
	C	21.19	6228	2.46
	D	19.70	5027	3.69
c	A	21.80	6798	1.11
	B	20.55	5681	2.55
	C	21.44	6456	2.44
	D	20.32	5496	4.36
j	A	20.84	5923	1.63
	B	20.33	5504	1.34
	C	20.92	5991	1.41
	D	20.41	5568	2.29
q	A	20.18	5386	2.03
	B	19.84	5129	1.30
	C	20.43	5584	1.81
	D	20.27	5457	1.76
zh	A	15.54	2748	2.59
	B	20.41	5568	1.77
	C	15.45	2712	3.36
	D	13.47	2022	4.63
ch	A	17.96	3910	2.78
	B	19.55	4919	2.43
	C	16.48	3153	3.10
	D	15.69	2809	4.96

Figure 1a. Wave-forms of Standard Chinese fricatives.

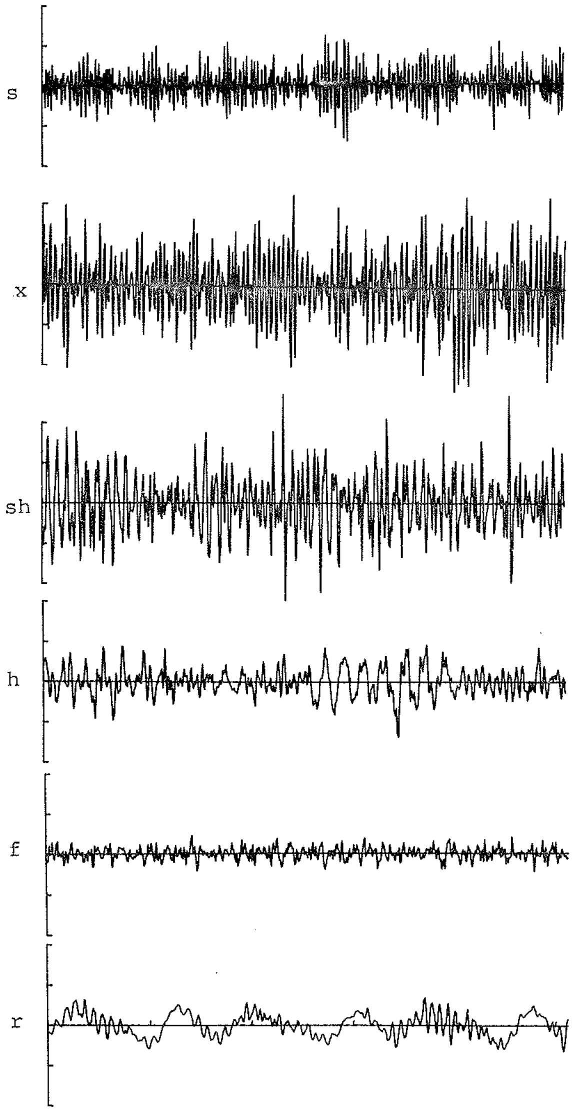


Figure 1b. FFT spectra in linear scale

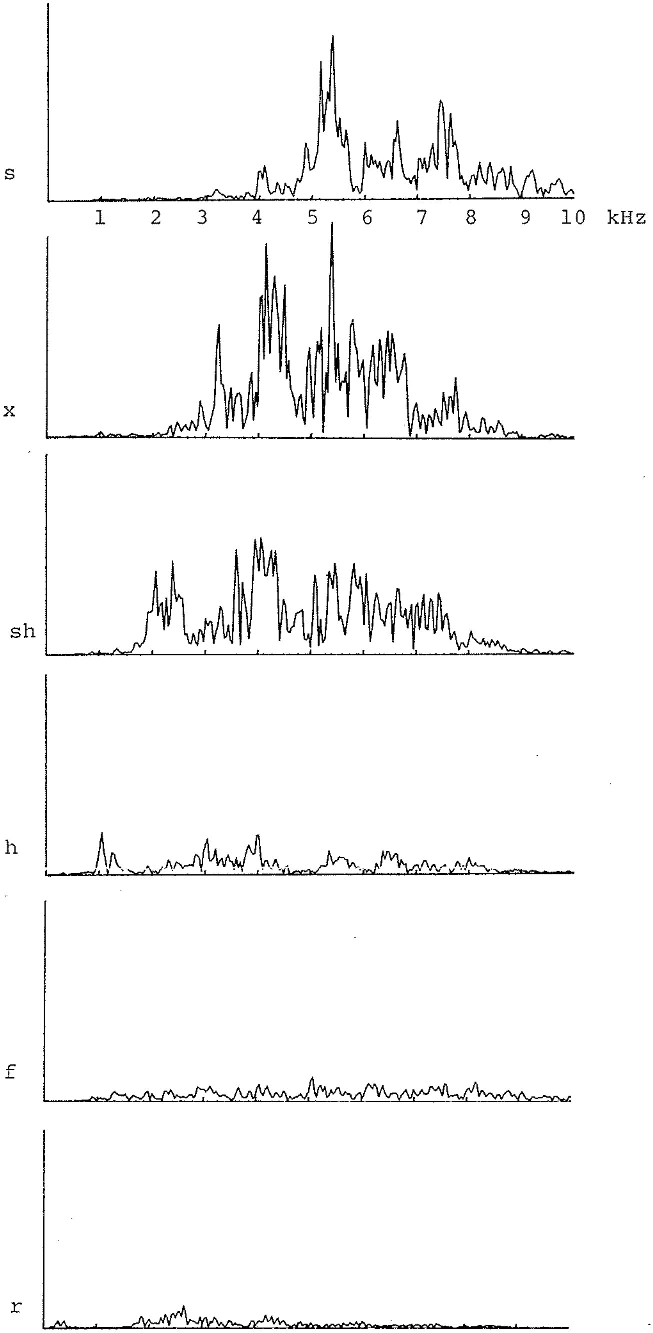


Figure 1c. FFT spectra in logarithmic scale (dB)

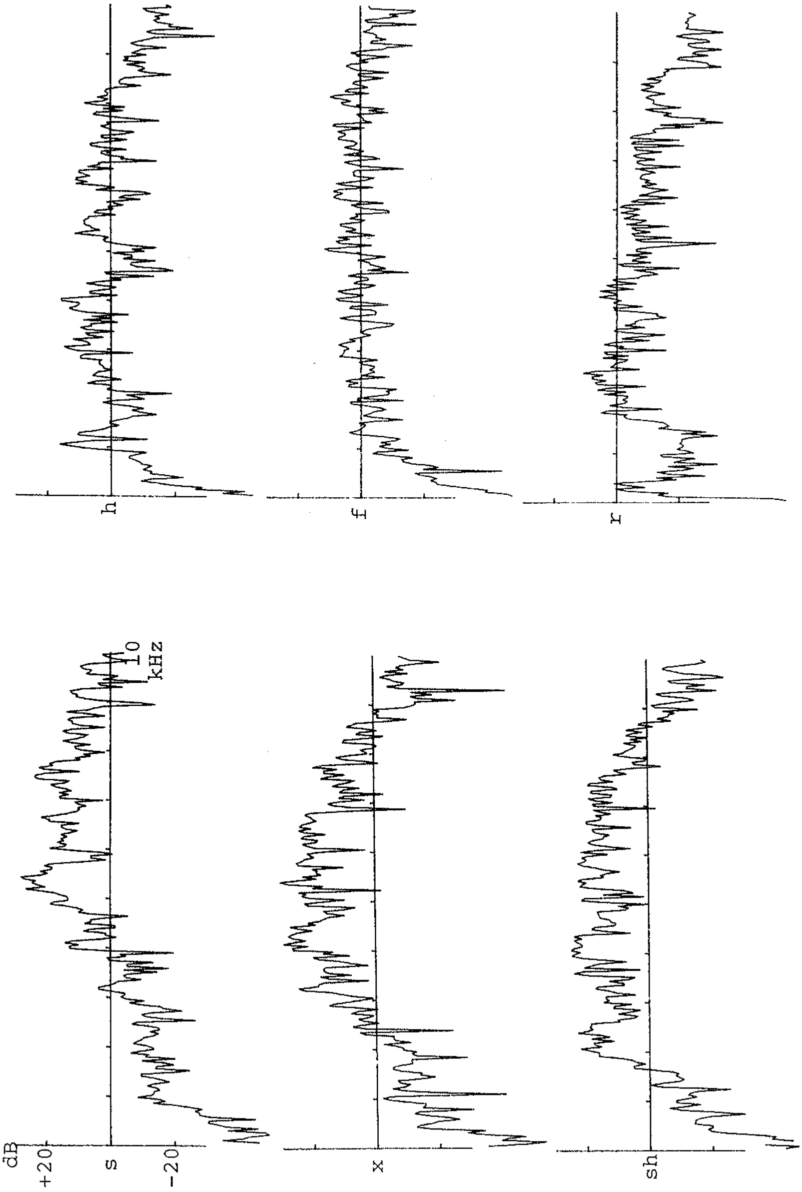


Figure 1d. Critical band spectra

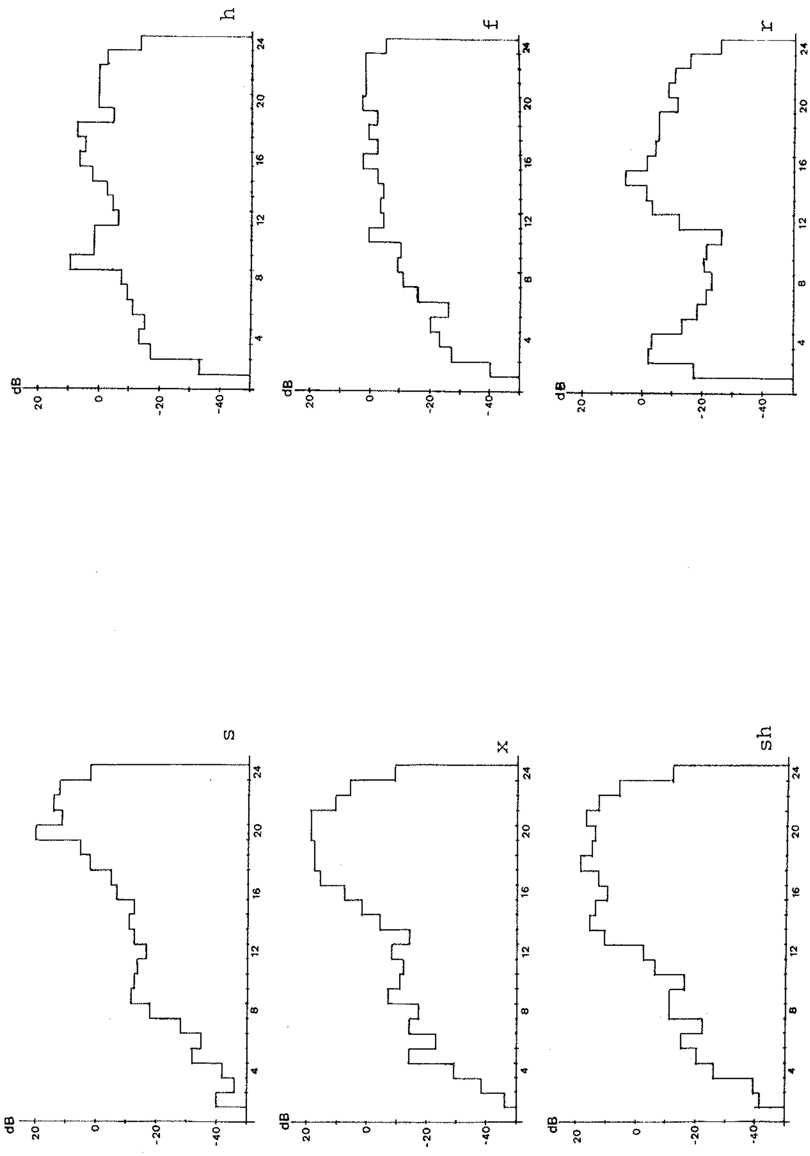


Figure 2

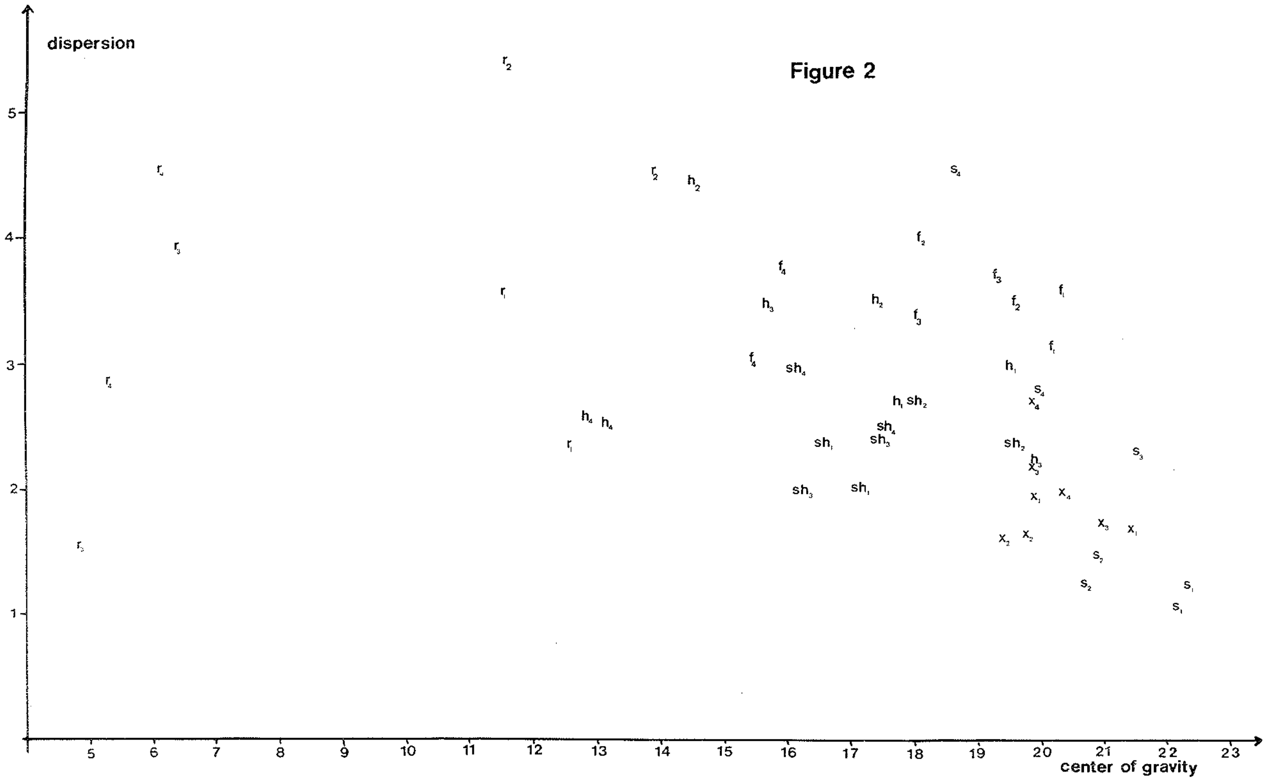
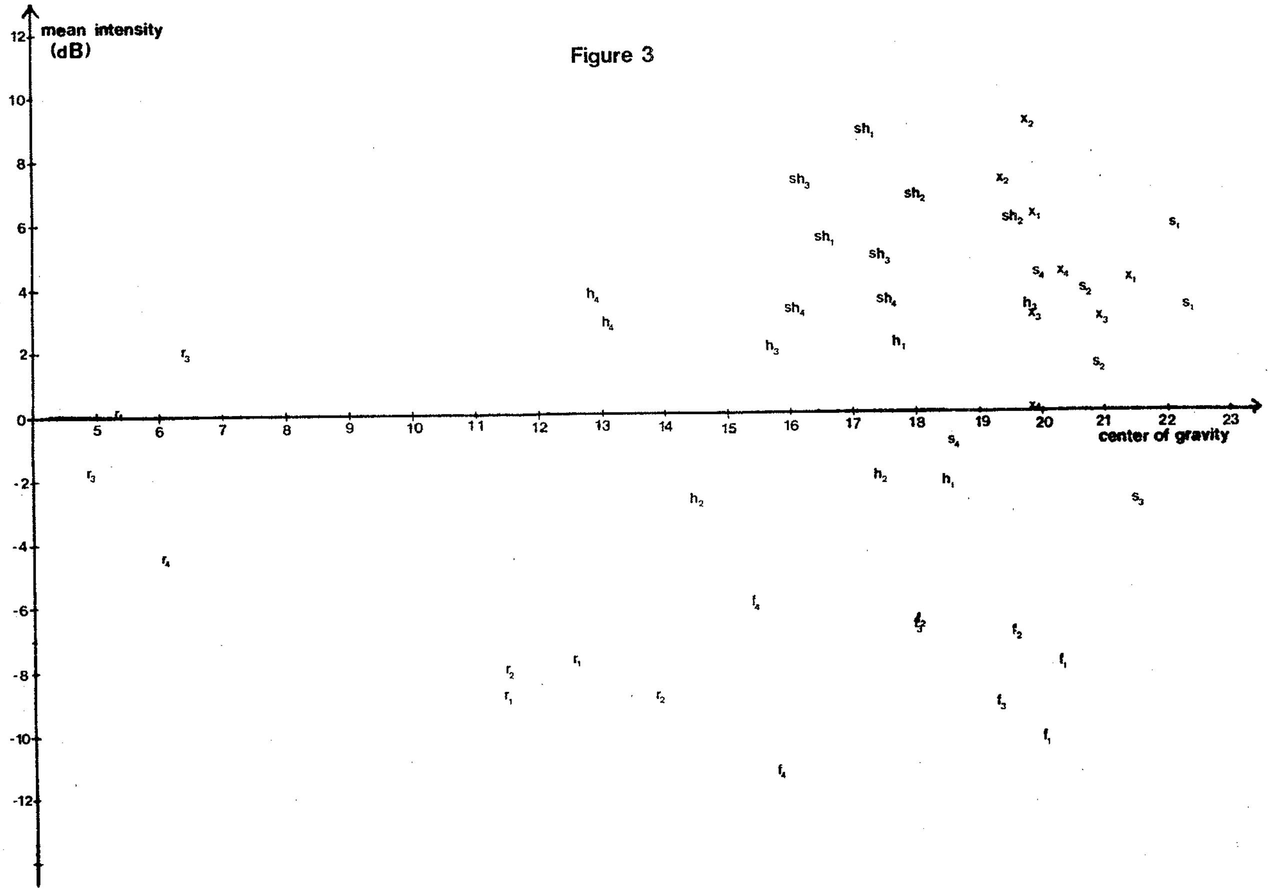

dispersion
center of gravity
5
4
3
2
1
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
r_2
r_4
r_3
r_1
r_4
r_5
r_2
h_2
s_4
f_2
f_4
f_3
f_1
h_3
h_2
f_2
f_3
f_1
f_4
sh_4
h_1
s_4
h_1 sh_2
x_4
h_4 h_4
sh_4
r_1
sh_1
sh_3
sh_2
s_3
h_3
x_3
sh_3
sh_1
x_1
x_4
x_3
x_1
x_2
x_2
s_2
s_2
s_1
s_1

Figure 3


mean intensity (dB)
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
center of gravity
sh₁
x₂
sh₃
sh₂
x₂
sh₂
x₁
s₁
sh₁
sh₃
s₄
x₄
x₁
s₂
h₄
sh₄
sh₄
h₃
x₃
x₃
s₁
h₄
r₃
h₃
h₁
s₂
r₄
x₄
s₄
r₃
h₂
h₁
h₂
s₃
r₄
f₄
f₂
f₃
f₂
r₁
f₁
r₂
r₁
r₂
f₃
f₁
f₄