(ARIPUC 13, 1979)

ACOUSTIC-STATISTICAL ANALYSIS OF VOICE QUALITY

Børge Frøkjær-Jensen

Abstract: The investigation deals with long-time analyses of:
(1) fundamental frequency,
(2) intensity above 1000 Hz relative to the total intensity, and
(3) total intensity.
Data are sampled 20 times per second. For each parameter, the mean and the coefficient of variance are calculated. The method seems useful for comparisons of changes in prosody taking place e.g. during voice training. Especially the coefficient of variance seems to be a good measure for the dispersion of pitch, "voice quality"[1], and intensity.

1. Purpose

The primary purpose of this investigation is to get an idea about which significant changes in fundamental frequency range, "voice quality", and intensity we can extract from conventional registrations of the fundamental frequency and intensity carried out as long-time analyses. From a phonological point of view the changes are not important, but for the description of speech characteristica for individuals and groups of individuals they are extremely important.

2. Method

2.1 Material

In order to get a normal material, it was decided to use a class of 16 students with normal, healthy voices at a teachers'

---

1) Parameter 2 is in the following text referred to as "voice quality" in quotation marks in order to indicate that there is no unambiguous correspondence between the subjective term voice quality and parameter 2.

college (Poul Hansen, 1979). In connection with their instruction the students were given 15 45-minute lessons of voice and speech training. Tape recordings of the students were made before and after voice training. The tape recordings consisted of easy texts, each of approximately 1 minute's duration. The same text, recording room, etc. was used for both recordings. One student was left out of the investigation because of some noise which made the recordings incomparable.

2.2 Acoustic analysis

When listening to the tape recordings before and after training, some changes are clearly audible in intonation and "voice quality". These changes must be caused by corresponding changes in the acoustic spectrum, primarily in the fundamental frequency, the spectral energy distribution (i.e. the relation between the energy above 1000 Hz and the total energy[1]), and the total intensity[2]. Computing their mean values and dispersions, we may get some statistical measures which will correlate with the audible changes. Only in cases with a breathy and aspirated voice quality parameter 2 is unapplicable because the value of the parameter would be too high. For a good voice the "voice quality" parameter will shift between -10 and -15 dB.

The whole text has been analyzed by a fundamental frequency meter and an intensity meter (F-J Electronics), extracting the mentioned parameters, and the resulting curves have been recorded on a mingograph (Elema 800). The measurements have been carried out on two randomly chosen sentences in the middle of the text (duration of the sentences: approximately 6 seconds).

In these sentences all voiced passages have been measured every 1/20 second (approximately 100 measures per parameter), and transferred to punched cards.

---

1) In other papers (Frøkjær-Jensen and Prytz 1976, Wedin et al. 1977) "voice quality" has been defined as the relation between the energy above and below 1000 Hz, but this was not possible in this investigation because of lack of instrumentation.

2) For practical reasons all intensities have been measured as sound pressures, i.e. a doubling of the intensity does not correspond to 3dB but to 6dB. The term intensity has been preserved because the signals are rectified and integrated over 20 ms.

2.3 Statistical analysis

The data have been processed by an EDP-program which gave the following statistical information:

1. A histogram showing the number of observations per 10 Hz interval for the fundamental frequency, and per 1 dB interval for the remaining two parameters: the intensity relation and the total intensity.
2. Arithmetic mean.
3. Standard deviation.
4. Coefficient of variance ($SD/\overline{X}$).
5. Test of normality based upon a χ^2-test.
6. Skewness with calculated divergence from the normal distribution.
7. Kurtosis with calculated divergence from the normal distribution.
8. Geometric mean, and 20, 25, 50, 75, and 80 percentile points.

The test of normality must be considered a necessary prerequisite for the application of statistical methods based on the normal distribution. The data samples were tested as to normality by means of χ^2-tests. Skewness and kurtosis were also calculated. Table 1 shows the results of the normality test of all data samples. Out of 81 data samples which could be χ^2-tested, only six samples diverged so much that they could not be accepted as belonging to normally distributed data populations.

The coefficient of variance is a measure of the modulation span irrespective of the average being high or low. It appears to be well suited as a measure for the modulation, both for the fundamental frequency, the "voice quality", and the total intensity. The higher the coefficient of variance, the greater the modulatory span of the three parameters.

Concerning the fundamental frequency I have observed (based on more than 100 subjects) that the coefficient of variance varies from below 0.1 for voices with a poor modulation to more than 0.2 for voices with a lively fundamental frequency modulation.

Since the samples in the majority of cases did not deviate significantly from the normal distribution, the comparison of the recordings before and after training were carried out using a series of t-tests.

Table 1

Normality of data populations
+ indicates P>0.05 (normal distribution)
0 indicates 0.01<P<0.05 (borderline case)
- indicates P<0.01 (data not normally distributed)

Parameter No. 1 (fundamental frequency):															
Subject	1	2	3	4	5	6	7	14	8	9	10	11	12	15	16
before	-	0	+	+	0	-	+	+	+	0	0	+	0	+	+
after	+	+	+	0	+	+	+	+	-	0	+	0	+	0	+
Parameter No. 2 ("voice quality"):															
Subject	1	2	3	4	5	6	7	14	8	9	10	11	12	15	16
before	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
after	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Parameter No. 3 (total intensity):															
Subject	1	2	3	4	5	6	7	14	8	9	10	11	12	15	16
before	0	+	+		0		0		0	+	+	+	+		
after	-	-	-		0	+	0		+	0	+	+	+		

Table 2

Calculations for subject No. 1
Fo: Fundamental frequency
Alfa: Intensity above 1000 Hz relative to total intensity
Int.: Total intensity

Calculations for	Before voice training			after voice training		
subject No. 1	Fo	Alfa	Int.	Fo	Alfa	Int.
Arithmetic mean	248.1	-15.5	30.3	300.5	-13.5	32.7
Standard deviation	29.1	6.59	6.0	43.6	6.17	5.3
Coefficient of var.	0.117	-0.426	0.199	0.145	-0.456	0.161
χ^2	38.5	9.80	19.2	9.22	13.20	35.8
Probability	0.0001	0.2000	0.0140	0.3241	0.0674	0.0001
Degrees of freedom	8	7	8	8	7	9
Skewness	1.26	-0.38	-1.14	0.27	-0.17	-1.35
Kurtosis	1.20	-0.49	0.96	-0.86	-0.83	1.58
Geometric mean	246.8	-16.1	30.1	297.9	-14.1	32.5

The results of the t-tests were checked, however, by two different non-parametric tests (the Walsh test and the Wilcoxon test). Apart from one instance the three test methods gave identical results regarding significance levels achieved.

3. Results

3.1 Results for an individual subject

3.1.1 Fundamental frequency:

Table 2 shows the results for an individual subject before and after voice training (the corresponding frequency distributions are shown in Figure 1A and Figure 1B). The most important calculations are the arithmetic mean and the coefficient of variance. Comparing the fundamental frequency before and after voice training it is found that the mean frequency has been raised from 248 to 300 Hz. Furthermore, the coefficient of variance has been increased from 0.117 to 0.145[1]. This means that not only has the fundamental frequency been raised, but the dispersion of the fundamental frequency (the intonation range) has also been considerably increased.

3.1.2 "Voice quality"

Table 2 shows the calculations for the "voice quality" (the corresponding histograms are shown in Figure 2A and Figure 2B). The χ^2-tests show that the populations for "voice quality" cannot be considered significantly different from the normal distribution. We may therefore compare the means which show that the "voice quality" is increased +2 dB (from -15.5 to -13.5 dB) during training.

Furthermore, the coefficient of variance has increased by 7% (from 0.426 to 0.456) indicating a tendency to make better use of variations in "voice quality" after the voice training.

---

1) This comparison is not statistically justified because the data before voice training are not normally distributed (see table 1). Instead, we may base the comparison on the geometric mean. However, if we calculate the corresponding coefficient: $(P_{75}-P_{25})$/geometric mean, we get an even better result: 0.121 before compared with 0.268 after voice training.

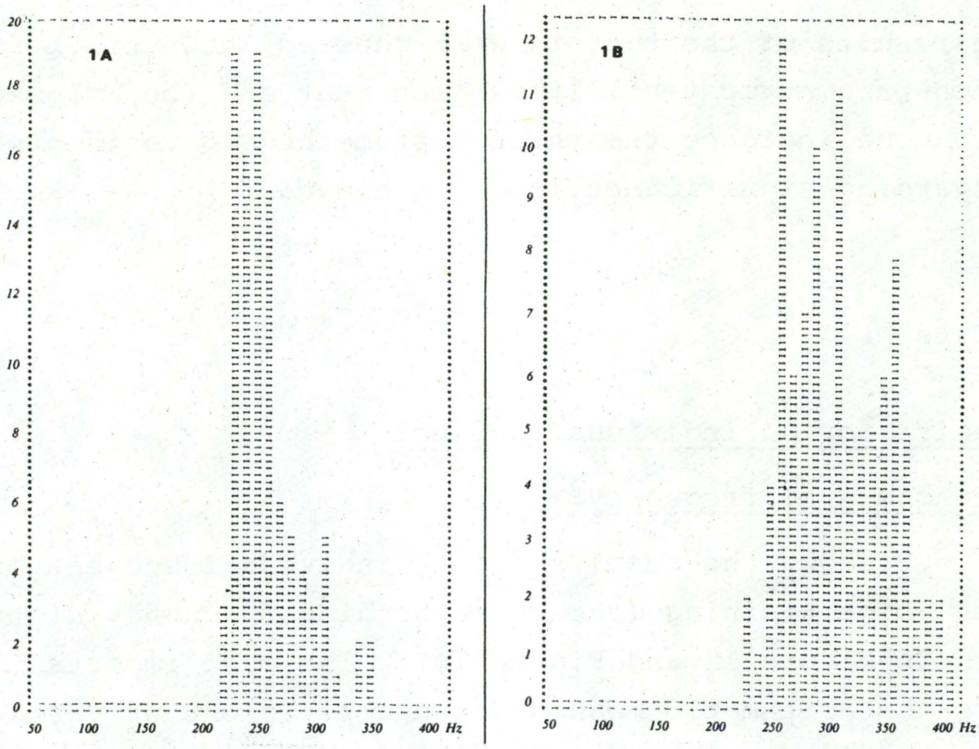

Figure 1A

Figure 1B

Histograms showing the distribution of fundamental frequency measurements before and after voice training.

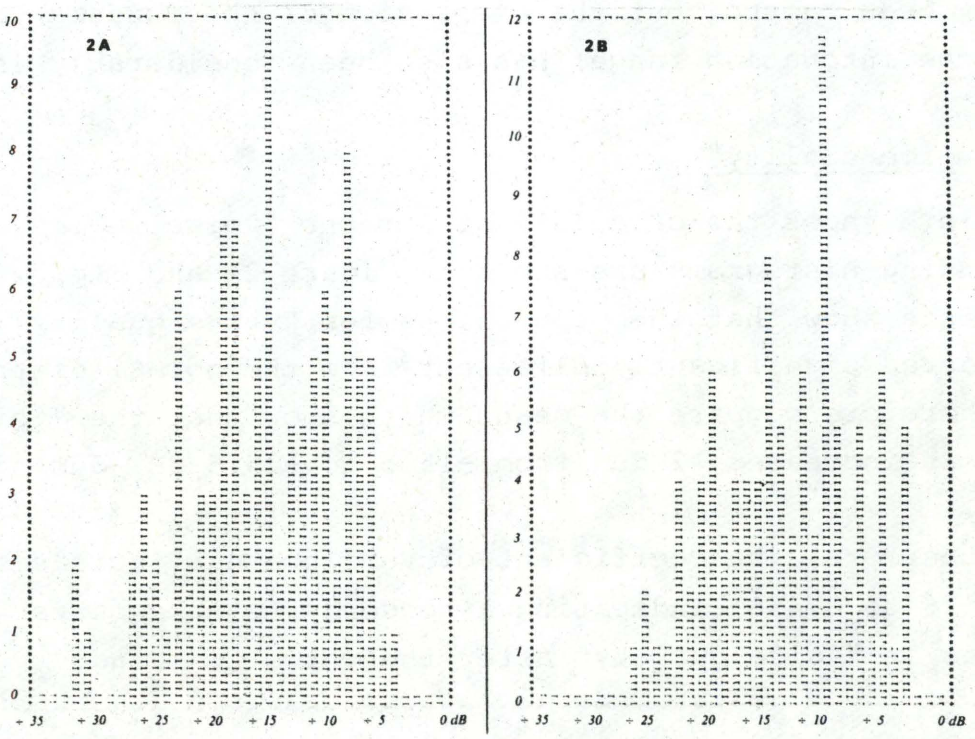

Figure 2A

Figure 2B

Histograms showing the distribution of "voice quality" measurements before and after voice training.

3.1.3 Total intensity

Again, refer to table 2 for the calculations for parameter 3. (The corresponding histograms are shown in Figure 3A and Figure 3B.) As shown by the tests, the data populations for the total intensity are not normally distributed. Therefore, we will have to compare the geometric means. Both the geometric mean and the percentiles show that the intensity level has been increased approximately 2 dB during training. These results are only to be considered as a tendency because the two samples cannot be regarded as significantly different. The coefficient of variance is reduced by 19% for this single subject. This is opposite to the normal findings which show an increase of approximately 10% on the average.

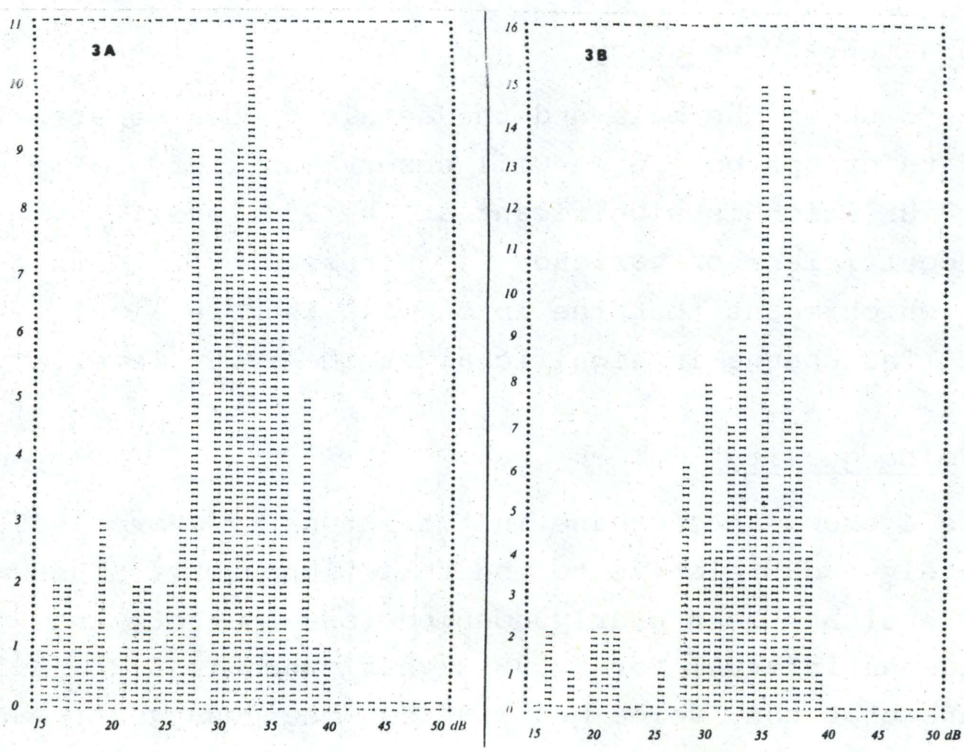

Figure 3A Figure 3B

Histograms showing the distribution of total intensity measurements before and after voice training.

3.2 Comparisons of the results before and after voice training for all 15 subjects

Tables 3, 4, and 5 show the changes in each of the three parameters during voice training. Most important are the changes in the mean and in the coefficient of variance. There is a clear tendency to change the mean value and to increase the coefficient of variance.

In order to test whether this change is significant, a Student's t-test has been applied to both statistic parameters for all subjects pooled (the two extreme right columns in each table). Roughly, a good agreement has been found between non-parametric tests such as the Walsh and the Wilcoxon test and the Student's t-test, though a few of the data populations differ from the normal distribution.

3.2.1 Fundamental frequency

Table 3 shows the male and the female voices separately. For the two data groups pooled we will observe an increase in F_o of 12.2 ± 10.7 Hz which is significant at the 95% level.

The coefficient of variance is increased by 24% (from 0.122 to 0.151) which means that the intonation is more lively after the training. The change is significant at the 99.5% level.

3.2.2 "Voice quality"

Table 4 shows an increase in the intensity above 1000 Hz of approximately 5 dB relative to the total intensity. This means that the level has been nearly doubled (see footnote 2 on page 190). The confidence interval for 99.9% significance is 5.2 ± 3.2 dB, which means that the "voice quality" has been changed towards a brighter timbre and consequently a better intelligibility.

Also the coefficient of variance has been increased (39% from 0.297 to 0.414). This means that the "voice quality" after the training disperses over a greater voice quality range, - the subjects are using the fluctuations towards the hypo- and the hyperfunctions as part of their natural modulatory pattern. The observed changes are significant at the 99.0% level.

Table 3

Fundamental frequency: comparisons before and after voice training.

$\overline{X}$: average of N_x measurements

SD: standard deviation

SD_x^2: variance of $\overline{X}$

VC_x: coefficient of variance for $\overline{X}$

Df: degrees of freedom

SD_d: standard deviation of the difference

SE_d: standard error of the difference

t: average/standard error (Student's t)

P: probability

K: female subject

M: male subject

	Before training				After training				Difference	
subject	$\overline{x}$	N_x	SD^2_x	VC_x	$\overline{y}$	N_y	SD^2_y	VC_y	$\overline{y}$-$\overline{x}$	VC_y-VC_x
1 K	248,2	100	849	0,117	300,5	99	1897	0,145	52,3	0,028
2 K	239,3	81	942	0,128	276,0	85	941	0,111	36,7	−0,017
3 K	230,8	78	845	0,126	214,6	66	736	0,126	−16,2	0,000
4 K	237,1	92	524	0,097	240,8	72	1053	0,135	3,7	0,038
5 K	230,0	96	465	0,094	239,6	78	874	0,123	9,6	0,029
6 K	245,5	89	720	0,109	227,5	71	1233	0,154	−18,0	0,045
7 K	221,4	86	225	0,068	232,6	84	1178	0,148	11,2	0,080
14 K	201,3	33	1501	0,192	242,3	33	1734	0,172	41,0	−0,020
mean	231,7			0,116	246,7			0,139	15,0	0,023
8 M	110,4	92	99	0,090	130,2	87	500	0,172	19,8	0,082
9 M	111,3	75	163	0,115	112,9	91	274	0,147	1,6	0,032
10 M	150,5	106	513	0,151	155,0	99	1003	0,204	4,5	0,053
11 M	135,5	106	427	0,153	137,3	97	516	0,165	1,8	0,012
12 M	127,1	62	457	0,168	133,0	66	616	0,186	5,9	0,018
15 M	125,5	50	131	0,091	139,6	45	254	0,114	14,1	0,023
16 M	158,6	39	456	0,135	173,6	42	858	0,169	15,0	0,034
mean	131,3			0,129	140,2			0,165	9,0	0,036
mean M+K	184,8			0,122	197,0			0,151	12,2	0,029
								Df	14	14
								SD_d	19,32	0,2947
								SE_d	4,988	0,00761
								t	2,446	3,828
								P $<$	0,05	0,005
								significance level % $>$	95,0	99,5
								confidence interval	12,2±10,7	0,029±0,025

Table 4

"Voice quality": comparisons before and after voice training.

subject	Before training				After training				Difference	
	$\bar{x}$	N_x	SD^2_x	VC_x	$\bar{y}$	N_y	SD^2_y	VC_y	$\bar{y}-\bar{x}$	VC_y-VC_x
1 K	−15,45	100	43,4	0,426	−13,54	99	38,1	0,456	1,91	0,030
2 K	−20,56	81	14,1	0,183	−14,33	85	15,6	0,275	6,23	0,092
3 K	−18,61	76	24,2	0,264	−13,73	66	24,7	0,362	4,88	0,098
4 K										
5 K	−24,71	96	51,4	0,290	−16,08	78	43,2	0,409	8,63	0,119
6 K	−18,26	89	38,9	0,341	−15,20	71	37,6	0,403	3,06	0,062
7 K	−19,64	86	36,4	0,307	−16,45	84	30,6	0,336	3,19	0,029
14 K	−17,91	33	42,8	0,365	−11,00	33	27,5	0,476	6,91	0,111
8 M	−16,48	92	27,1	0,316	−11,91	87	24,7	0,418	4,57	0,102
9 M	−18,95	75	12,5	0,186	−13,70	91	18,8	0,317	5,25	0,131
10 M	−14,05	106	29,6	0,387	−11,56	99	20,3	0,390	2,49	0,003
11 M	−16,35	106	18,5	0,263	−13,88	97	17,1	0,298	2,47	0,035
12 M	−18,69	62	26,9	0,277	−15,98	66	13,8	0,232	2,71	−0,045
15 M	−21,30	50	23,0	0,226	−12,42	45	71,7	0,683	8,88	0,457
16 M	−20,33	39	43,8	0,326	−9,43	42	49,1	0,743	10,90	0,417
mean	−18,66			0,297	−13,52			0,414	5,15	0,117

	$\bar{y}-\bar{x}$	VC_y-VC_x
SE_d	0,751	0,0386
t	6,85	3,036
P <	0,001	0,010
significance level % >	99,9	99,0
confidence interval	5,2 ± 3,2	0,117 ± 0,116

Table 5

Total intensity: comparisons before and after voice training.

subject	Before training				After training				Difference	
	$\bar{x}$	N_x	SD^2_x	VC_x	$\bar{y}$	N_y	SD^2_y	VC_y	$\bar{y}-\bar{x}$	VC_y-VC_x
1 K	30,3	100	36,2	0,199	32,7	99	27,8	0,161	2,4	−0,038
2 K	34,5	81	17,9	0,122	33,2	85	23,6	0,146	−1,3	0,024
3 K	33,6	76	34,2	0,173	28,2	66	29,3	0,192	−5,4	0,019
4 K										
5 K	34,7	96	39,8	0,182	28,4	78	29,2	0,190	−6,3	0,008
6 K					(28,9)	71	23,9	(0,169)		
7 K	31,0	78	20,0	0,144	28,9	84	26,1	0,176	−2,1	0,032
14 K										
8 M	30,1	92	16,5	0,135	32,4	87	25,2	0,155	2,3	0,020
9 M	32,8	75	42,9	0,200	31,2	91	38,6	0,199	−1,6	−0,001
10 M	34,2	106	40,6	0,186	33,5	99	41,2	0,191	−0,7	0,005
11 M	32,7	106	31,7	0,172	28,6	97	43,3	0,230	−4,1	0,058
12 M	33,3	62	31,6	0,169	33,5	66	37,6	0,183	0,2	0,014
15 M										
16 M										
mean	32,7			0,168	31,1			0,182	−1,7	0,014

	$\bar{y}-\bar{x}$	VC_y-VC_x
SE_d	0,9930	0,00779
t	1,779	1,809
P <	0,15	0,15
significance level % >	85	85
confidence interval	1,7 ± 1,5	0,014 ± 0,013

3.2.3 Total intensity

Table 5 shows a tendency to a minor reduction of the intensity level (2 dB), but the changes are not statistically significant. The reduced intensity may be explained psychologically as a result of the quite audible improvement of the "voice quality" where formant 2 and 3 become stronger. Consequently a significantly clearer voice with a better intelligibility and a greater ability to penetrate the background noise is obtained. The result could be that the subject does not need to talk as loud as before.

No significant change in the coefficient of variance is observed, though there is a tendency to greater variations (8%) in the intensity after voice training.

4. Summary

The results reported show that the dispersion of the single parameters expressed in terms of the coefficient of variance contains so clear information about the changes in the modulation that the three described coefficient of variances could be used as a standardized description of:

for parameter (1): the modulatory span of the fundamental frequency,

for parameter (2): the modulatory span of the "voice quality", and (to a lesser degree, perhaps)

for parameter (3): the span of the intensity modulation.

References

Davis, Steven B. 1958: "Acoustic characteristics of normal and pathological voices", Haskins SR 54

Davis, Steven B. 1976: "Computer evaluation of laryngeal pathology based on inverse filtering of speech", Monograph No. 13, Speech Communications Research Laboratory, Inc., Santa Barbara, California

Frøkjær-Jensen, B. and S. Prytz 1976: "Registration of voice quality" Technical Review 3, Brüel & Kjær.

Hansen, Poul 1979: "Åndedrætsfunktion og stemmedannelse", Dansk Audiologopædi, May.

Kitzing, Peter 1979: Glottografisk frekvensindikering (doctoral dissertation) Malmö, Sweden.

Smith, Svend and Kirsten Thyme 1976: "Statistic research on changes in speech due to pedagogic treatment", Folia Phon. 28, p. 98-103

Snedecor, Georg W. and W.G. Cochran 1976: Statistical methods, 6th edition/ 8th printing (The Iowa State University Press)

Wedin, S., R. Leanderson and L. Wedin 1977: "Evaluation of voice training. Spectral analysis compared with listeners' judgement", 17th International Congress of Logopedics and Phoniatrics (Copenhagen)