# REGISTRATION OF VOICE QUALITY[1]

Børge Frøkjær-Jensen[2] and Svend Prytz[3]

Abstract:

Long-time-average-spectra recordings of normal voices as well as an average spectrum of such LTAS-registrations are shown and discussed.
For comparisons of voice qualities we have tried to set up a new parameter, $\alpha$, which is a measure of the intensity relations in the higher and the lower parts of the speech spectrum:

$$\alpha = \frac{\text{intensity above 1000 Hz}}{\text{intensity below 1000 Hz}}^{4}$$

Because the spectrum above 1000 Hz is normalized relative to the spectrum below 1000 Hz, $\alpha$ is independent of microphone distance, amplification level, etc. $\alpha$ seems to be a good acoustic correlate to the physiological term "medial compression", and preliminary research indicates that it is relevant to evaluations and comparisons of voice qualities.
The "quality parameter" is represented graphically by histograms showing the number of $\alpha$-values automatically sampled during a read text, and it is displayed on a storage oscilloscope along the vertical axis. The horizontal axis is used for displaying the total speech intensity.

**We define voice quality as an auditory property, i.e. an aspect of the perception of the human voice. A good voice quality depends on (1) certain typical formant patterns, (2) absence of noise in the acoustic spectrum, and (3) a high degree of absence of aperiodicity in the fundamental frequency.**

---

1) Paper read at the International Congress of Phonetic Sciences in Leeds, August 1975. In the present version, only minor modifications have been made.

2) Audiologopedic Research Group, University of Copenhagen, 96 Njalsgade, 2300 Copenhagen S, Denmark.

3) Svend Prytz, M.D., Phoniatric Laboratory, ENT-department, University Hospital, Blegdamsvej, 2100 Copenhagen Ø, Denmark.

4) Cf. the paper by Svend Smith and Kirsten Thyme in this issue of ARIPUC.

In 1963 Wendahl and Moore found a direct relation between the jarring, rough, and hoarse voice quality in voices suffering from unilateral recurrent paralysis and the variations in periodicity between adjacent pitch periods.

Lieberman has defined these pitch variations in terms of the so-called Lieberman pitch perturbation factor, and he has analyzed the magnitude of this factor in different larynx disorders by computer.

Smith and Lieberman found significant differences between normal subjects and patients suffering from cancer of the vocal folds, polyps on or adjacent to the vocal folds, and acute and chronic laryngitis.

Koike improved this method and got similar results, whereas Hecker and Kreul could not reproduce Lieberman's results, even though the methods were almost identical. They defined instead a "directional pitch perturbation factor", which depends on the direction of the perturbation change. This factor was a significant improvement in the discrimination of pathological from healthy voices. Furthermore, they found a more narrow distribution of the fundamental frequency in pathological voices than in normal voices, and they established that the averaged fundamental frequency and duration of phonations were reduced compared to the normal voices. However, the patients used in this investigation were all selected and matched, and it was found that they all suffered from laryngeal cancer.

Hans von Leden and Iwata have investigated pitch perturbations (among other diseases) in 10 patients of unilateral recurrent paralysis before and after teflon®-injection in the paralysed vocal fold. They found the method reliable and useful in the phoniatric clinic. The discrimination between different laryngeal diseases was poor, however.

Just as important as the cycle-to-cycle variations in pitch is the acoustic structure of the speech spectrum. According to the literature, this spectrum has its origin in the voice source and decreases by 12 dB per octave. However, during normal speech we find variations in the slope. Glottography and inverse filterings seem to show that the slope for voiced consonants is about -15 dB per octave, and thus steeper than the slope for vowels. On the other hand, we find changes in the opposite direction during high voice effort, such as shouting.

Changes in voice quality are used by singers and actors as an artistic way of expressing their emotions and moods, whereas vocal disability as well as voice disorders create unpleasant voice qualities, such as breathiness, hoarseness, and roughness.

Within the phoniatric and logopedic clinic there is a great demand for developing instrumentation and methods for registration of changes in voice quality.

The present paper is a preliminary report, dealing with three different methods for voice quality analyses:

(1) Long-time-average-spectral analysis based on a read text of a duration of 45 seconds.
(2) Histograms of the voiced part of speech showing the amplitude level above 1000 Hz, relative to the level below 1000 Hz.
(3) The relative amplitude parameter shown as a function of the total amplitude level on a storage oscilloscope screen.

The analyzer used for the long-time-average-spectral analyses is a Brüel & Kjær 400 channel measuring system, which consists of a spectrum analyzer, an averager, and a 12" display with a level recorder for paper curve recordings. For further details we refer to the B & K manual and to our previous paper (Frøkjær-Jensen and Prytz 1974). In that paper we pointed out that the dynamic range of the instrument was too restricted for speech analysis. However, this problem was overcome by introducing a 6 dB per octave high shaping in the analyzed frequency range up to 5000 Hz. Furthermore, we have introduced a gating system which cuts out all voiceless speech segments. In this way the unvoiced sounds do not contribute to the total energy of the LTAS-analyses.

The first illustration (fig. 1)[1] shows four LTAS-analyses taken from 22 normal voices. Along the X-axis we have the relative amplitude level in dB. The dotted line indicates the above-mentioned preemphasis of +6 dB per octave. We observe deviations among the four voices. Especially for voices Nos. 2 and 22 we observe a pronounced depiction of the lower harmonics, which we may interpret as restricted variations in the fundamental frequency or intonation for these two voices. It may be due to the subjects' behaviour during the recording procedure. We do not find this harmonic pattern in voices Nos. 10 and 20.

The next illustration (fig. 2) shows the spectral distribution of 10 normal voices set up in the same graph. Notice the relatively small dispersion among the curves, which indicates that the spectral distribution of normal and healthy voices is fairly constant, at least up to about 3000 Hz.

In fig. 3 the solid line shows the average curve based upon the just shown 10 normal and healthy voices. The dotted line indicates the commonly presumed slope of -6 dB per octave of the radiated sound wave (voice source + radiation). As may be expected, we notice that the slope of the averaged speech spectra is steeper than -6 dB per octave.

---

1) During the presentation of this paper, all illustrations were shown as slides, and sound samples from all voices shown on the slides were replayed from tape.

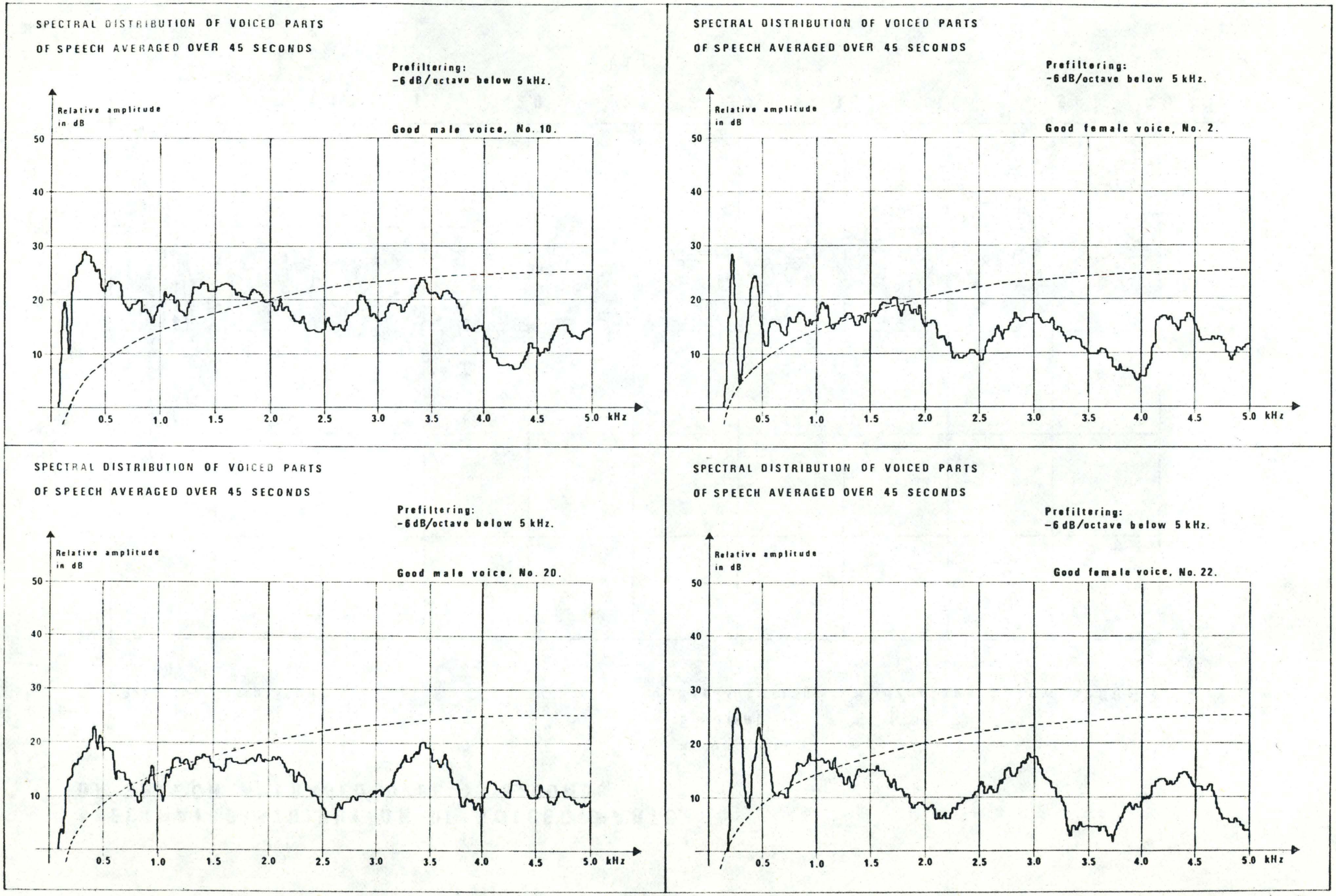



Fig. 1

LTAS-analyses of good and healthy voices.

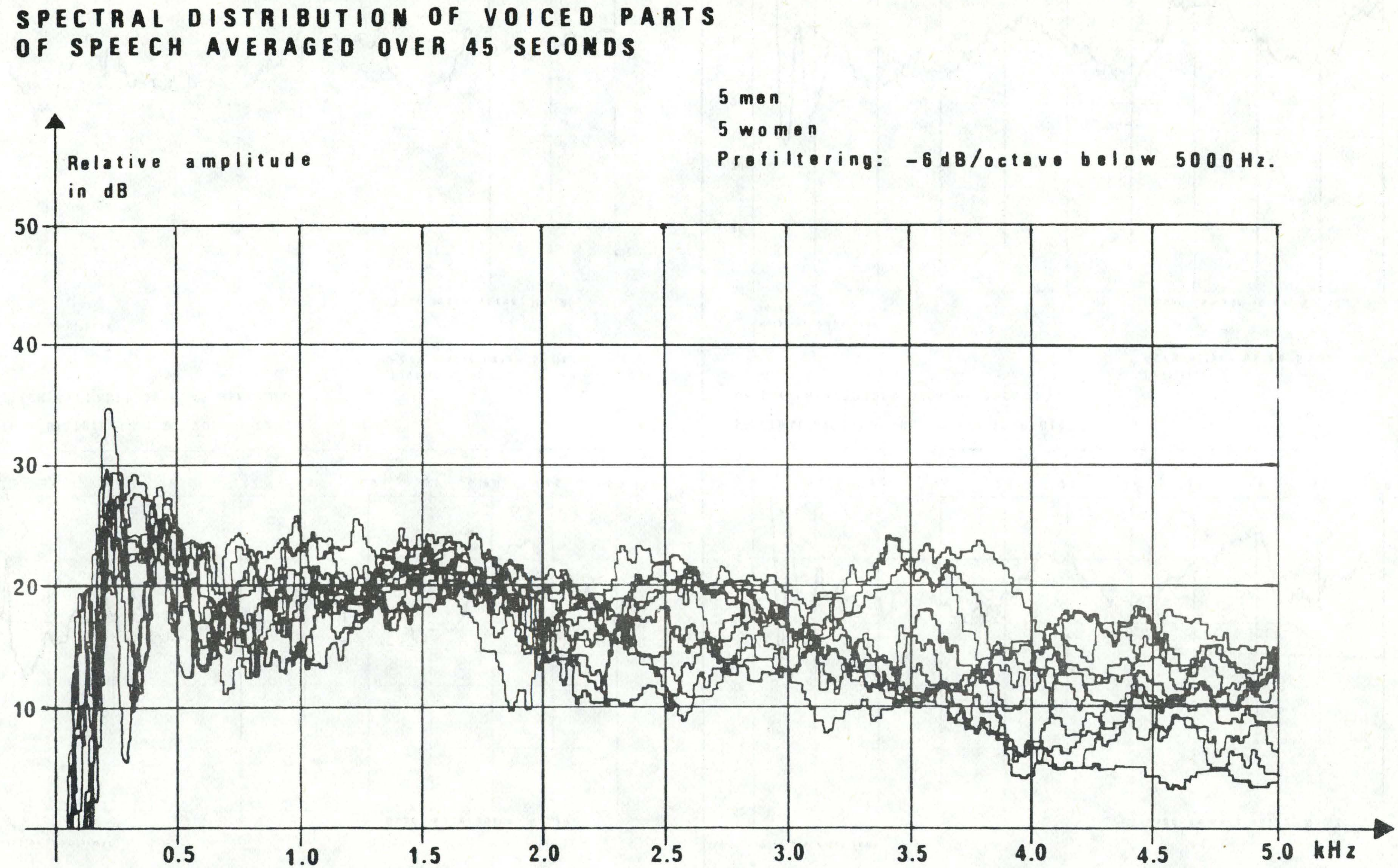

SPECTRAL DISTRIBUTION OF VOICED PARTS
OF SPEECH AVERAGED OVER 45 SECONDS
5 men
5 women
Prefiltering: -6dB/octave below 5000Hz.
Relative amplitude
in dB
50
40
30
20
10
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0 kHz

Fig. 2

SPECTRAL DISTRIBUTION OF VOICED PARTS
OF SPEECH AVERAGED OVER 45 SECONDS
FOR 10 GOOD VOICES

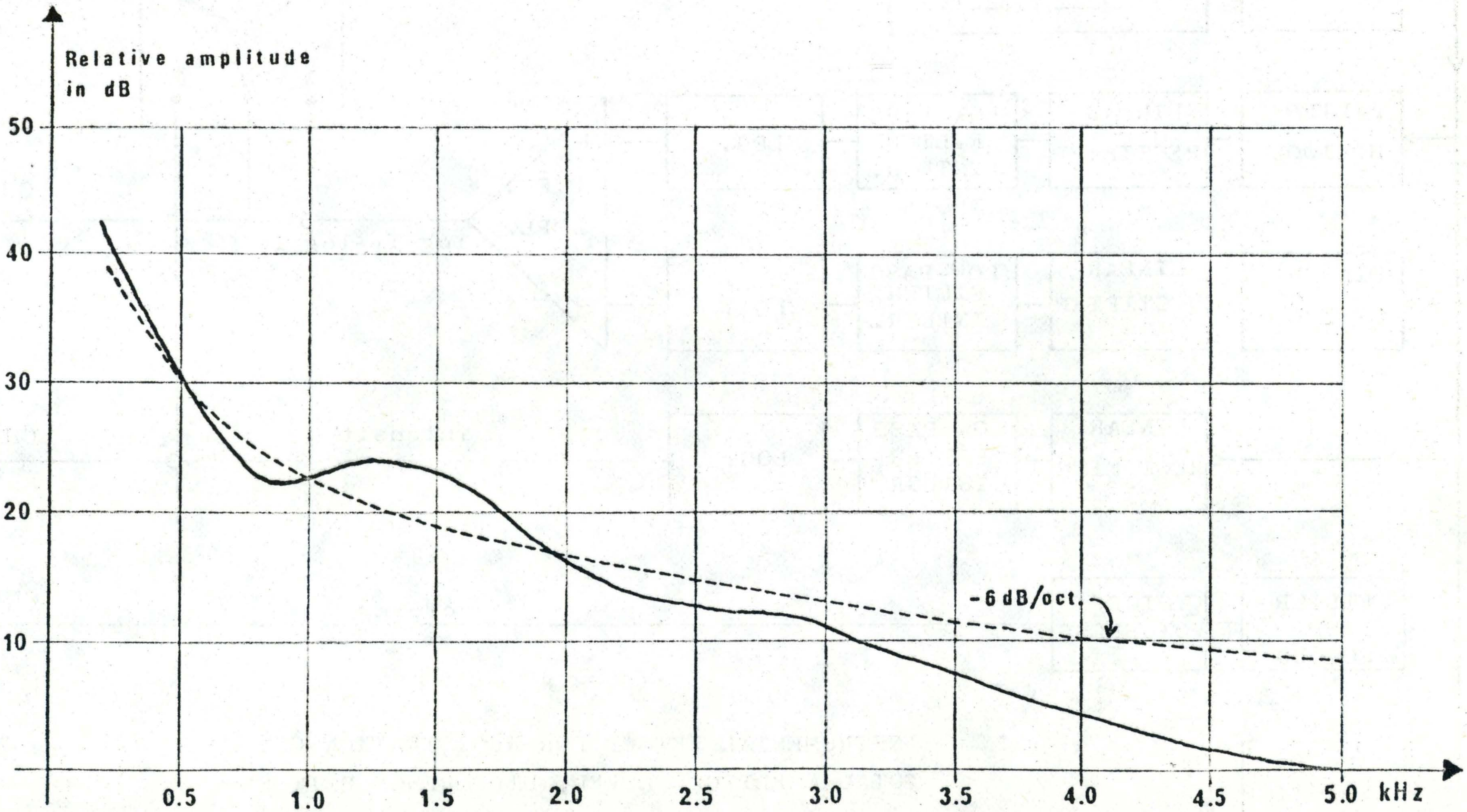
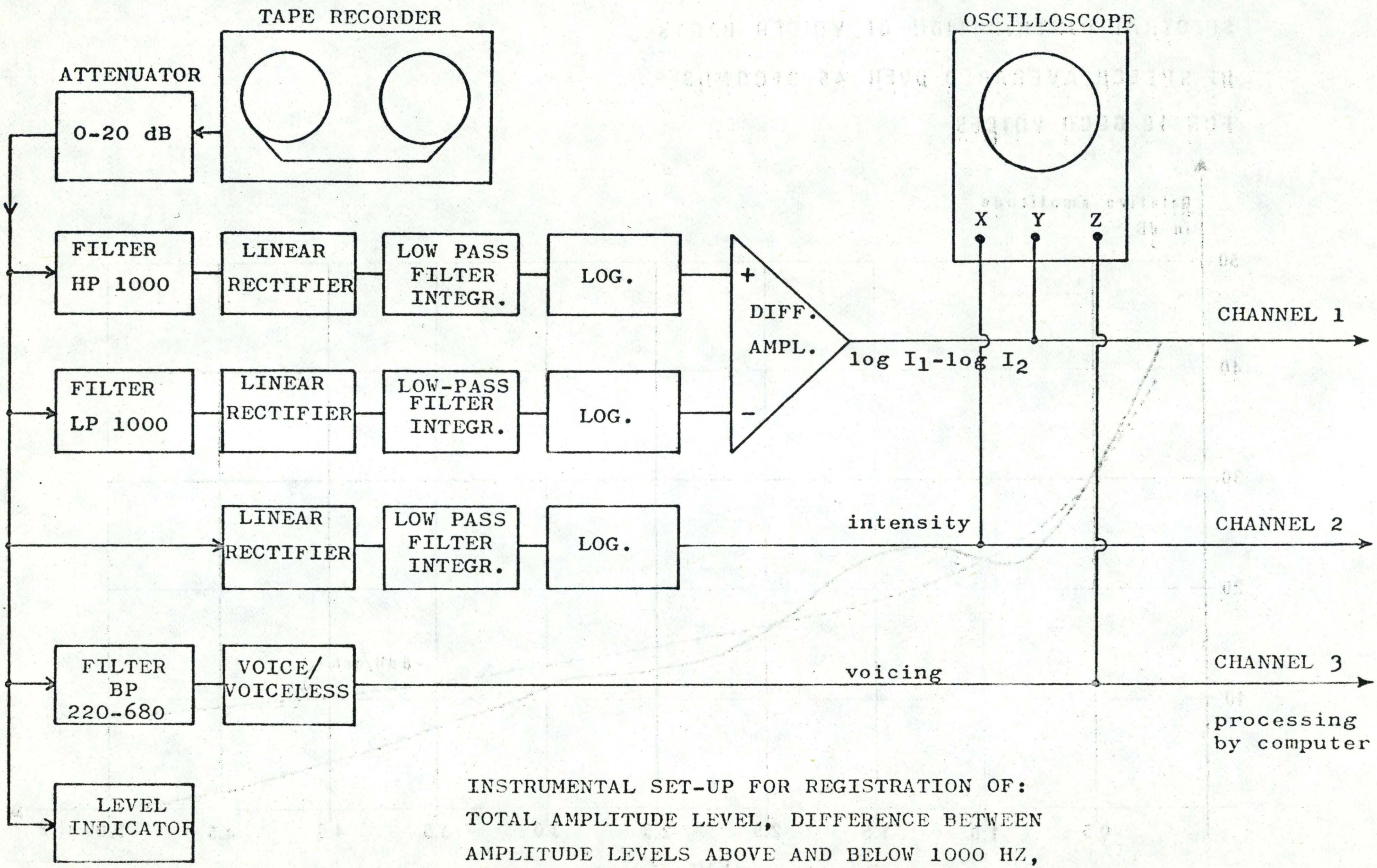



Fig. 3




INSTRUMENTAL SET-UP FOR REGISTRATION OF: TOTAL AMPLITUDE LEVEL, DIFFERENCE BETWEEN AMPLITUDE LEVELS ABOVE AND BELOW 1000 HZ, AND DURATION.

Fig. 4

Just around 800 Hz we find a zero in the radiated sound spectrum, but we cannot, based upon these recordings, decide whether this is due to less frequent occurrence of formant energy around this frequency, or whether it is due to a zero in the voice source.

In the previous illustrations we have shown some analyses of normal voices. The following graphs depict comparisons of voices suffering from unilateral recurrent paralysis before and after therapy - not for the purpose of showing what happens during the treatment of a given disorder, but merely to show how these analyses could be used for comparisons of the voice qualities.

For these comparisons we have tried to set up a new parameter, which we have called $\alpha$.

We have defined

$$\alpha = \frac{\text{amplitude level above 1000 Hz}}{\text{amplitude level below 1000 Hz}}$$

$$\log\alpha = \log A\ (\text{above 1000 Hz}) - \log A\ (\text{below 1000 Hz})$$

Because the amplitude above 1000 Hz is normalized relative to the amplitude below 1000 Hz, $\alpha$ is independent of the microphone distance, amplitude levels, etc.

Fig. 4 shows how the $\alpha$-parameter is extracted from the tape recordings. In a differential amplifier we get the difference between the logarithmic voltages proportional to the intensity levels above and below 1000 Hz.

It does not matter whether we use intensities or amplitude levels, it will only be a question of calibration, because the intensities are proportional to the square of the sound pressure level.

The set-up includes a voice/voiceless indicator based upon a sensing of the energy in the $F_1$-region, and a full frequency logarithmic intensity channel.

The $\alpha$-parameter is displayed on a storage oscilloscope as a function of the total intensity, where the light intensity of the oscilloscope is switched off and on by the voice/voiceless indicator.

$\alpha$ may also be recorded automatically 25 times per second and represented as a histogram by the computer.

Fig. 5 shows the LTAS-analysis and $\alpha$-histograms of a patient with a phonatory hypofunction caused by recurrent paralysis before and after speech therapy. The graph shows how much the spectral amplitude has increased at different frequencies in the spectrum during treatment. - LTAS-graphs of phonatory hypofunctions often show that during speech therapy the energy is increased, except for the lowest part of the spectrum.

Examinations of LTAS-graphs from the voices of more than 50 patients and several normal subjects reveal that 1000 Hz seems to be a reasonable cut-off frequency for the above-mentioned comparisons between the higher and the lower part of the spectrum. This is in agreement with Ilse Lehiste, Gordon Peterson and Svend Smith.

The histograms of $\alpha$ before and after treatment in this illustration show an increase of about 4 dB for $\alpha$.

In the next illustration (fig. 6), we notice an increase of about 3 dB during the speech training.

Fig. 4 above showed the instrumental set-up for recording the $\alpha$-parameter. As it appears from that illustration, the -parameter could also be shown on an oscilloscope as a function of the total intensity. This is illustrated in fig. 7. The photos of the storage screen of the oscilloscope depict the $\alpha$-parameter as a function of the total intensity, averaged over 20 seconds. In this illustration we have given three healthy and three pathological voices for demonstration purposes only.

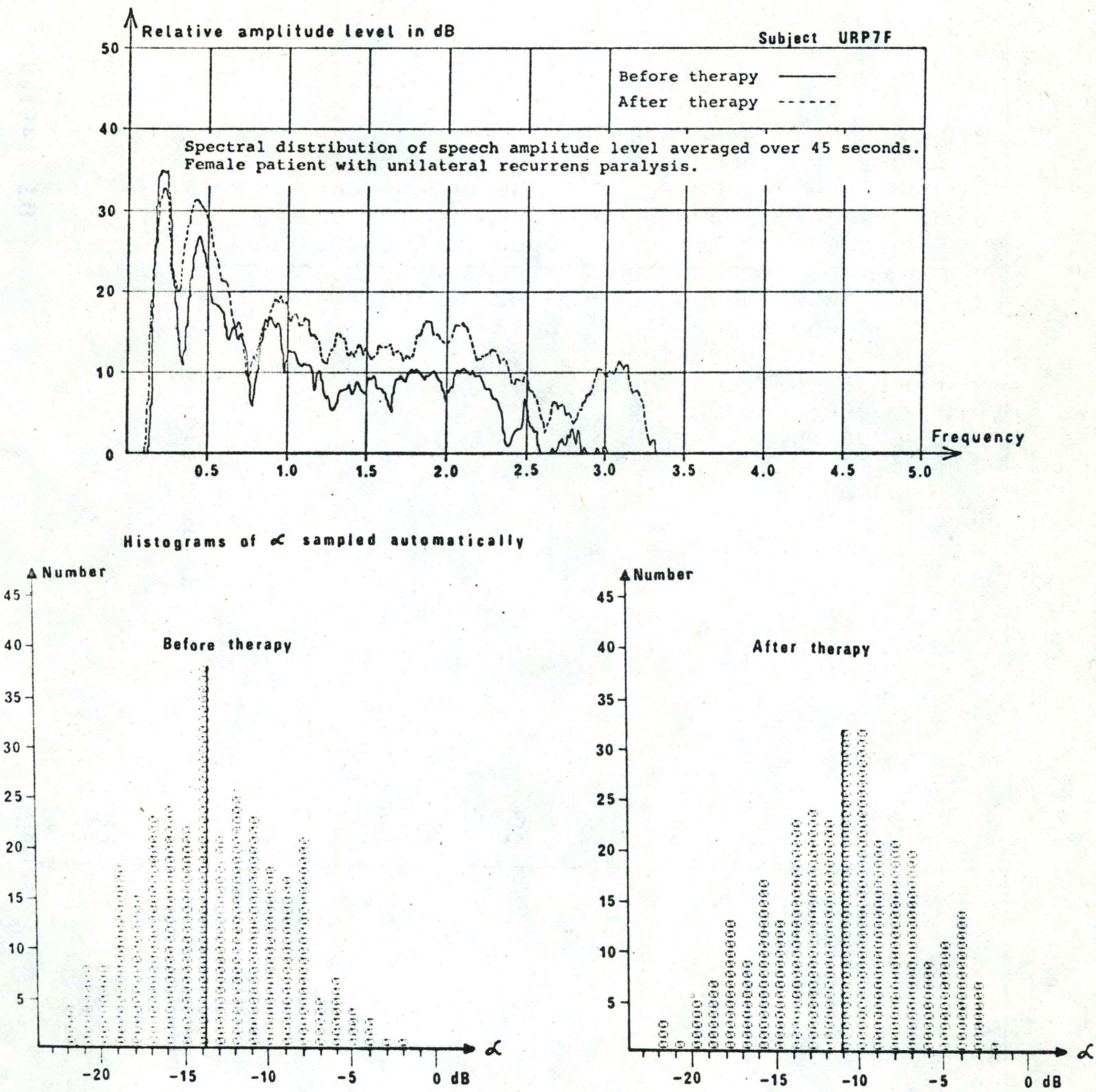



Fig. 5

Upper graph: LTAS-analysis of a patient suffering from unilateral paralysis before and after treatment.

Lower graphs: Distribution of the amplitude above 1000 Hz relative to the amplitude below 1000 Hz, sampled automatically 25 times per second.

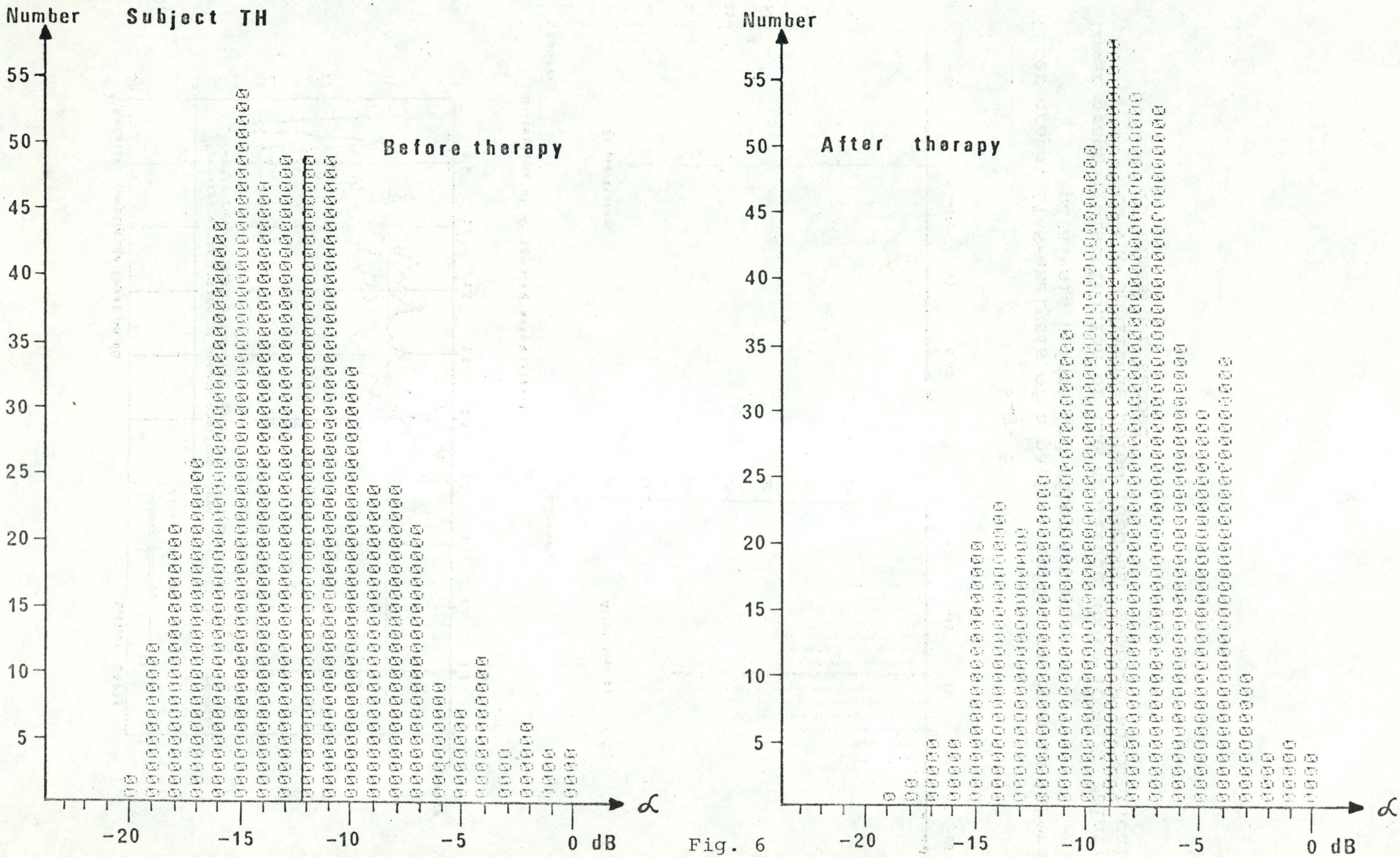



Fig. 6

Sample of $\alpha$-graphs before and after speech training, showing an increase of about 3 dB of the spectral amplitude above 1000 Hz relative to the spectral amplitude below 1000 Hz

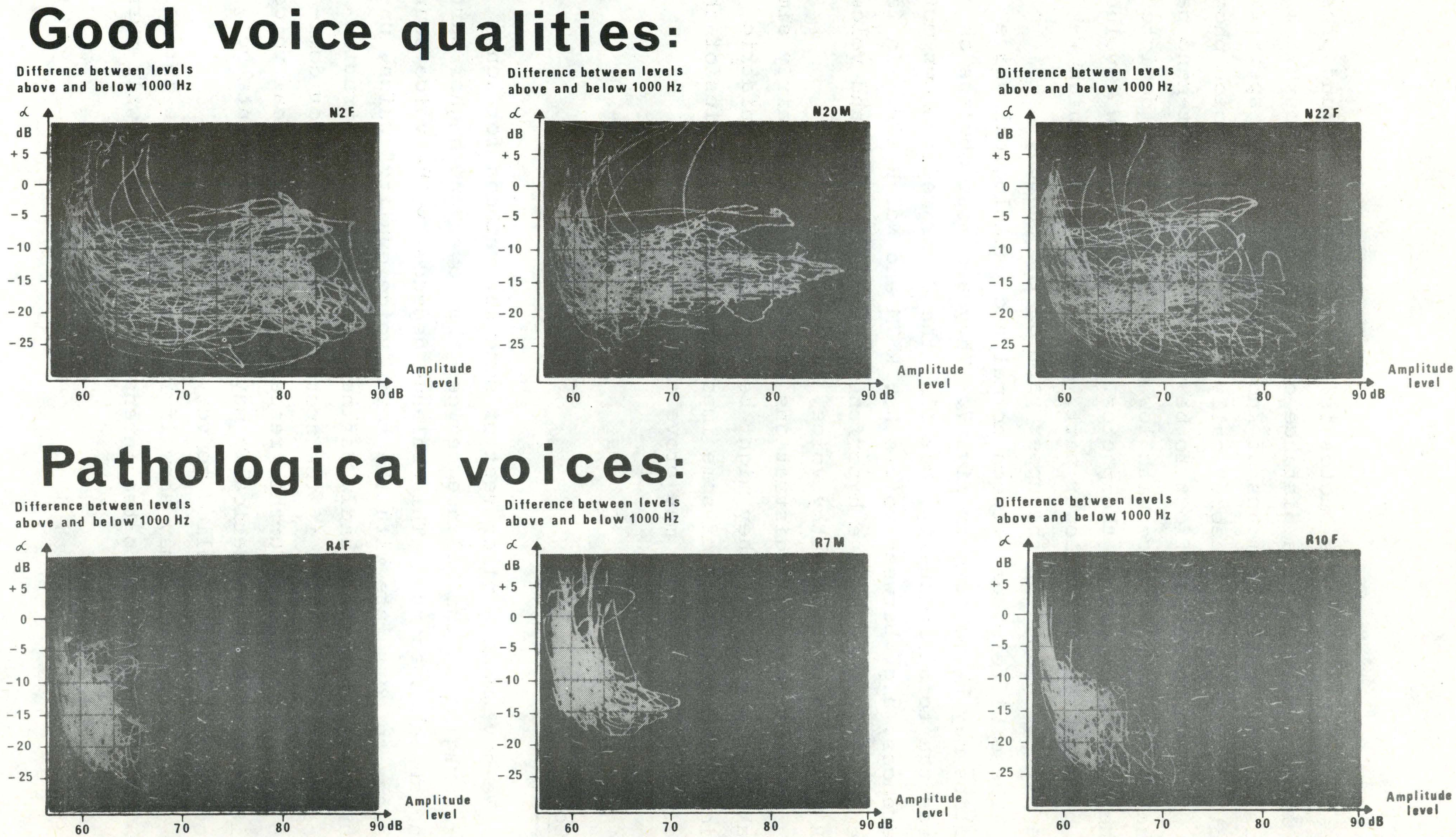



Fig. 7

Oscilloscope displays of total amplitude level (in dB SPL) and relative amplitude above 1000 Hz.

The X-axis in these photos is calibrated in dB sound pressure level, measured at a distance of 30 cm from the mouth. The Y-axis is calibrated by means of known synthetic vowel spectra and attenuated tones. Theoretically we may expect a voice phonating with a high voice effort to be placed in the right-hand part of the photos, and a voice with low voice effort to be depicted in the left-hand part of the photos. Voices with a low $\alpha$-value will be depicted in the lower part, and voices with a high $\alpha$-value will be placed in the upper part of the photos.

These differences between normal and pathological voice qualities may be noticed in fig. 7. They are most obvious as regards the total intensity, but also the $\alpha$-parameter shows mutual differences, e.g. between voices No. R 7M and No. R 10F. In fact, voice No. R 7M sounds as a hyperfunctional dysphonia, and voice No. R 10F as a weak, breathy voice.

As oscilloscope registrations of this kind are fairly simple to make, it seems that they might be useful in the phoniatric clinic as a quick check of some important characteristics of the voice. Further research may prove this.

## 3. Conclusion

We have made a pilot test of some new methods for long-time-averaging of the balance between the lower and higher parts of the speech spectrum. This balance depends on the voice source and seems to correlate with the term "voice quality" which, unfortunately, is still a badly defined term. The registration methods seem to be valid if the acoustic spectrum is not dominated by white noise. Further research with pathological voice qualities produced synthetically may show to what extent the method is valid when applied to very noisy voices.

The coming years may show if registrations of the intensity above 1000 Hz relative to the intensity below 1000 Hz will turn out to be a useful aid in the phoniatric and logopedic routine diagnosis, as well as a tool for voice evaluation during speech therapy.

References


Buch, N.H. and B. Frøkjær-Jensen 1972: "Some Remarks on Acoustic Parameters in Speech Disorders", ARIPUC 6, p. 245-259

Frøkjær-Jensen, B. and S. Prytz 1974: "Evaluation of Speech Disorders by Means of Long-time-average-spectra", ARIPUC 8, p. 227-237

Hecker, M.H.L. and E.J. Kreul 1971: "Descriptions of the Speech of Patients with Cancer of the Vocal Folds", JASA 49, 4, p. 1275-1282

Iwata, S. and Hans von Leden 1970: "Pitch Perturbations in Normal and Pathological Voices", Folia Phoniatrica 22, p. 413-424

Leden, Hans von and Y. Koike 1970: "Detection of Laryngeal Disease by Computer Technique", Arch. Otolaryng. 91, p. 3-10

Lehiste, I. and G. Peterson 1959: "The Identification of filtered Vowels", Phonetica 4, p. 161-177

Smith, S. 1961: "On Artificial Voice Production", Proc. Phon. IV, Helsinki, p. 96-110

Smith, W.R. and P. Lieberman 1969: "Computer Diagnosis of Laryngeal Lesion", Computers & Biomedical Research II, p. 291-303

Wendahl, R.W. 1966: "Some Parameters of Auditory Roughness", Folia Phoniatrica 28, p. 26-32

Winckel, F. 1967: "Darstellung des Sprechverhaltens als Statistische Tonhöhen und Formantverteilung mittels "Langzeitsanalyse"", Proc. Phon. VI, Praha, p. 1031-1035