EVALUATION OF SPEECH DISORDERS BY MEANS OF LONG-TIME-AVERAGE-SPECTRA

Børge Frøkjær-Jensen and Svend Prytz[1]

1. Introduction

As part of a project dealing with acoustic parameters for objective estimation of voice disorders (Buch and Frøkjær-Jensen 1972) some pilot registrations have been carried out by means of long-time-average-spectra (LTAS) (Winckel 1967, Winckel and Krause 1965, Blomberg and Elenius 1970, Jansson and Sundberg 1973, Fritzell, Halén and Sundberg 1974) of 20 patients with unilateral paralyses of the recurrens nerve.

Our LTAS-analyses had two purposes: (1) to test the applicability of a new commercially available "Real-time Narrow Band Analyzer" (type 3348, Brüel & Kjær)[2] in speech research, and (2) to study the changes in the spectral distribution of speech intensity which occur during a period of voice therapy.

2. Instrumentation

To our knowledge, long-time-average-spectra of human speech have only been employed sporadically in the last 25 years, partly due to the complicated instrumentation needed for these analyses, especially when performed in real time. Only a few laboratories have been able to set up their own LTAS-analyzing system.

---

1) Svend Prytz is a doctor at the ENT-clinic, University Hospital, Copenhagen.

2) We thank the firm Brüel & Kjær, Nærum, Denmark, for kindly having placed a type 3348 analyzer, a type 2307 Level Recorder, and a type 7003 DC tape recorder at our disposal.

This paper does not attempt to describe the new instrumentation in detail. We only intend to demonstrate the utility of the equipment when applied to speech research.

The analyzer is a 400 channel hybrid measuring system which consists of (1) a spectrum analyzer, (2) an averager and interface control, and (3) a 12" CRT display unit. Furthermore, (4) a level recorder is used for the paper curve recordings.

The LTAS-spectra shown in this paper have been analyzed by means of the spectrum analyzer which was adjusted to a squared cosine filter slope (Hanning weighting which gives the best selectivity) with a 3 dB bandwidth of 18.75 Hz and a resolution (distance between the centre frequencies of adjacent filters) of 12.5 Hz. A full spectrum analysis in the frequency range 0-5000 Hz was carried out every 80 ms.

In the averager and interface control the spectra have been averaged over a period of 45 seconds with a 50 dB dynamic range of the resulting average spectrum. Since there are pauses in the read text, whose effect are to lower the average level, it has been necessary to subject the averaged spectrum to +12 dB "post averaging gain" before displaying it. This suffices to give a 40 dB dynamic range in the long-time-averaged-spectra for normal subjects if there are only few pauses in the reading.

The averager and interface control unit contain digital memories for storing two spectra. The contents of either memory may then be continuously displayed, or the contents of both may be displayed alternately. In this way comparisons of LTAS before and after speech therapy are easily made.

In the display unit (the CRT) each of the 400 channels are updated 22 times per second. A special feature is a 4-digit display showing the signal level in dB (resolution 0.2 dB) of any preselected channel. Outputs are available at the rear of the cabinet for read-outs in BCD code for further dataprocessing or in analog form for a YT-recorder.

3. The Real-Time-Narrow-Band-Analyzer used for long-time-averaging of speech spectra

3.1 Test of the instrumentation with speech input

The instrumentation has been tested with speech input from 30 different speakers, some of these speaking slowly with long pauses, some speaking fast without pauses, some having a good voice quality and some having a weak voice with few harmonics.

To sum up our results: The analyzer seems to be an expedient instrumentation for LTAS analyses of speakers not making too long pauses in the speech flow. For a pathological speaker who must often talk with long pauses, the dynamic range of the averaged spectrum is too reduced.

If 50 % of the averaging time is used for pauses the dynamic range is reduced with a factor 2 which equals 6 dB. Furthermore, the mean level of sustained vowels is about half of the peak level, i.e. 6 dB below the peak level, and the consonants are of considerably lower mean level. The mean level of the speech flow is 9-18 dB below the peak level when we do not take the pauses into account.

If we average over a read text including pauses the mean level will be about 12-24 dB below the peak levels, depending upon the pause/speech-ratio and the spectral composition of the speaker's voice.

Fig. 1 shows an LTAS of a female patient with unilateral recurrens paralysis. The voice sounds dull and weak with a reduced ability to modulate the pitch and intensity.

The dynamic range of the instrumentation was not sufficient for registration of amplitude levels above 1200 Hz. If the dynamic range of the LTAS could have been expanded down to -20 dB most of the relevant information in the spectrum could have been displayed.

3.2 Changes in LTAS during speech therapy

Pilot analyses have been made in order to test the validity of LTAS analyses as an objective means for the evaluation of the spectral changes that occur during the recovery period in patients suffering from speech disorders.

Typically, the LTAS analyses of recurrens paralyses can be divided into two groups, based upon the spectral changes:

(1) Patients who get a reduced energy spectrum above the first formant region during the recovery, and

(2) patients who get an increased energy spectrum above the first formant region during the recovery.

Both groups consist of patients having a dyscoordination between the subglottal airpressure and the medial compression of the vocal folds. In the first group the result is a phonatory hyperfunction which emphasizes the higher part of the spectrum and diminishes the stability of the fundamental frequency (rough voice quality), often to such a degree that diplophonia occurs.

The second group, on the contrary, is characterized by a phonatory hypofunction where the vocal folds cannot close very well, which results in an acoustic spectrum with weak higher harmonics. Often some breathy noise is heard, caused by turbulent air passing through the glottis.

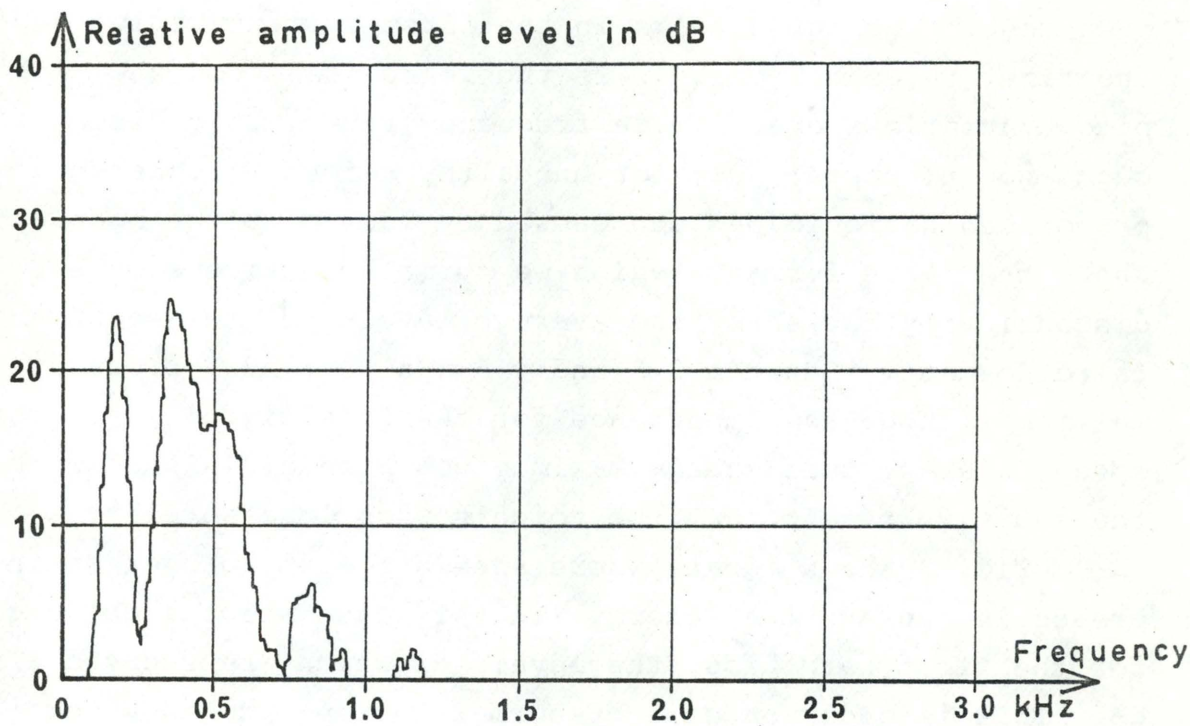

Figure 1

Spectral distribution of speech amplitude levels averaged over 45 seconds for a female patient with unilateral paralysis of the recurrens nerve.

Fig. 2 and Fig. 3 are illustrations of the two types of changes in the LTAS. In fig. 2 (female voice suffering from recurrens paralysis recorded before and during speech therapy) we observe that the acoustic spectrum below 1000 Hz has not been changed very much during therapy. Above 1000 Hz, however, the higher harmonics of the spectrum are weaker than before treatment. The three peaks in the first formant region depict the first three harmonics. If the voice is a good one with pitch variations over a wide frequency range, this discontinuity ought not to appear, but for unhealthy voices we observe a pronounced harmonic pattern caused by lack of pitch modulations. Above the first formant region we normally notice some spectral discontinuity caused by the average levels of the second and third formant. Between 3.5 and 4.5 kHz we find a broad energy maximum without any importance for the intelligibility. During speech therapy this energy maximum has been diminished with the auditive result that the voice sounds less shrilly.

Fig. 3 shows a case where speech therapy causes an increase in the spectral energy (in this case about 4 dB) except for the first harmonics, the levels of which are reduced a few dB. This is a common case of normal therapy progress for recurrens paralysis.

4. Further investigations

4.1 Averaging the LTAS analyses

Some preliminary attempts have been made with computer averaging of the LTASes.[1] The result of such averagings will be information concerning the mean distribution of the long-time averaged spectra for a certain group of speakers.

---

1) The averagings of long-time-average-spectra (recorded on a DC tape recorder) have been carried out on a PDP 8/lab computer at the ENT-clinic, University Hospital of Copenhagen.

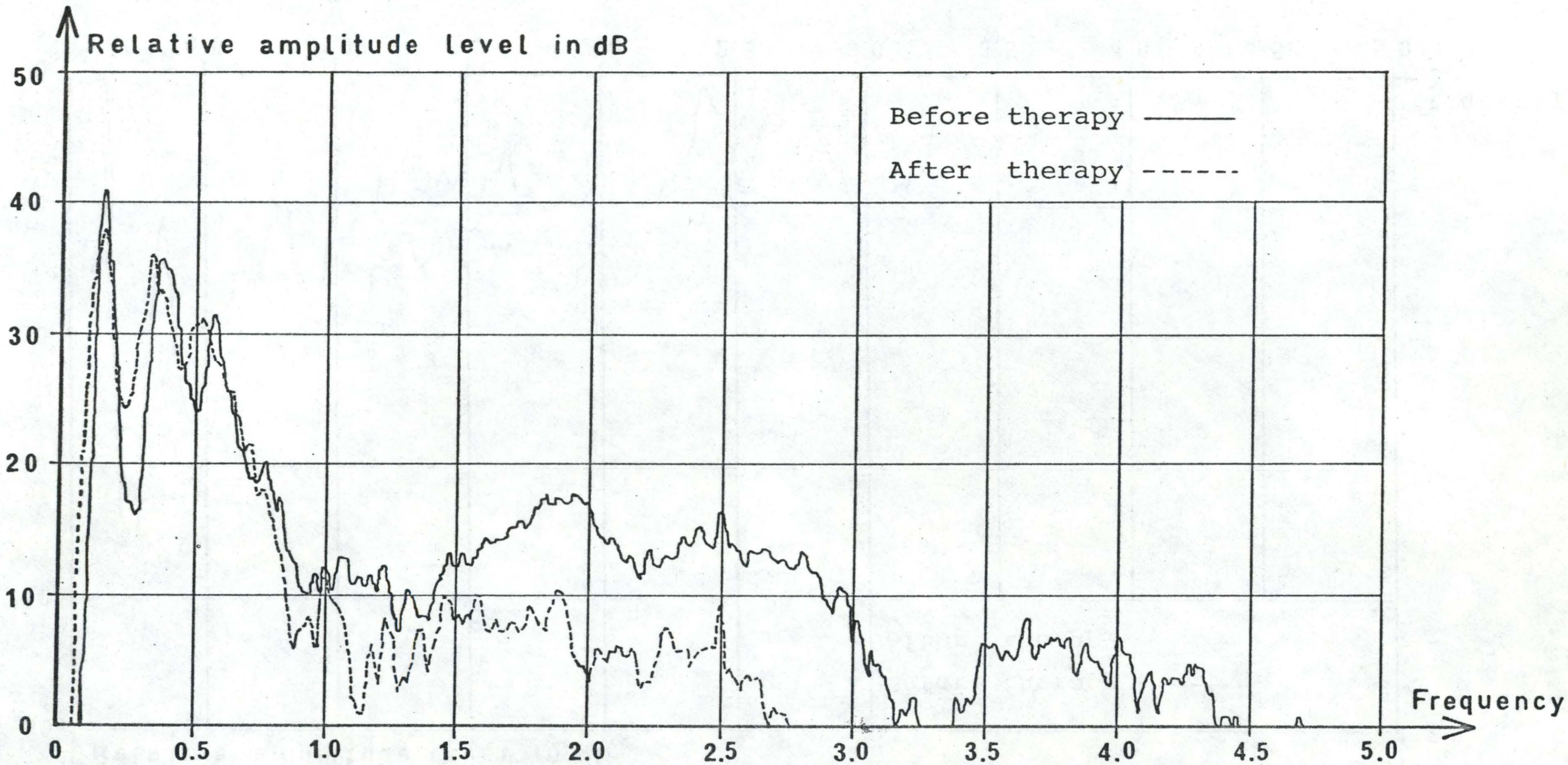

Figure 2

Spectral distribution of speech amplitude level averaged over 45 seconds. Female patient with unilateral recurrens paralysis.

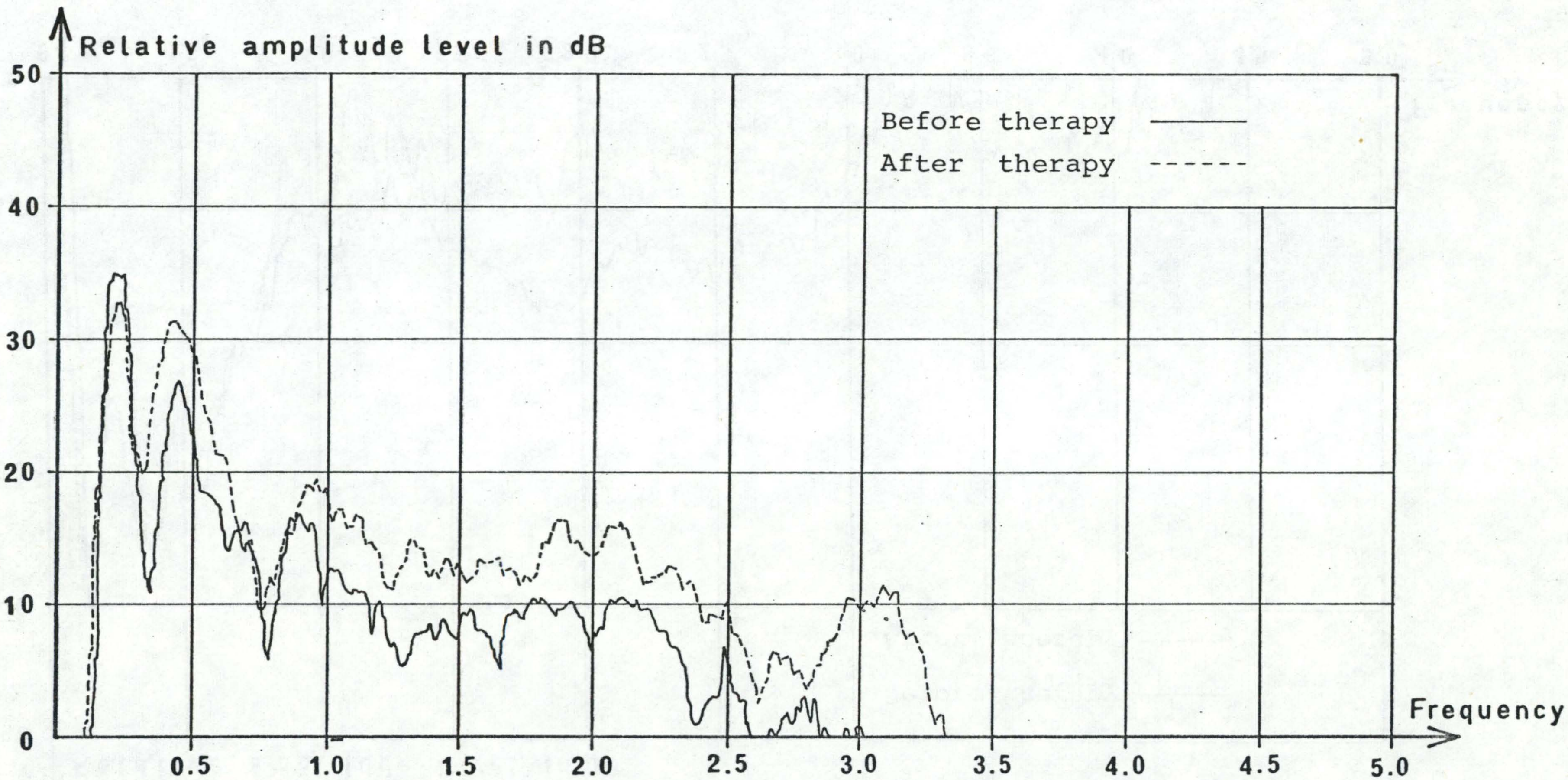

Figure 3

Spectral distribution of speech amplitude level averaged over 45 seconds. Female patient with unilateral recurrens paralysis.

In general, these LTAS-averagings show for the normal voice:

(1) that the "normal" amplitude spectrum of speech has a slope of about 12-14 dB per octave, and

(2) that the first formant emphasizes the spectrum round 400-500 Hz an extra couple of dB,

both phenomena have been demonstrated previously by other authors (Dunn & White 1940, Winckel & Krause 1965, Fritzell, Hallén and Sundberg 1974).

In particular, the LTAS-averagings of a homogeneous and well documented group of patients suffering from a certain disorder yield information concerning the general spectral tendencies for this disorder, and may thus be useful in establishing criteria for the evaluation of the deviations from the normal spectrum.

For example: during a period of voice therapy of voices suffering from recurrens paralysis the spectral changes in the LTAS-averagings show that the therapy mainly affects the level of harmonics above 1000 Hz (increase or decrease as explained above, see Buch & Frøkjær-Jensen 1972).

The contribution of the unvoiced sounds to the total spectral energy is nearly negligible as far as we have experienced. However, more detailed studies of the spectral contribution from the unvoiced and from the voiced sounds will be carried out separately. A phonation detector (sensing when the first formant is present) with a gate has been constructed. This circuit will be used for gating out the unvoiced and voiced sounds, respectively, before making the LTAS analyses.

4.2 Variations in the averaging time

All our analyses have been made with an averaging time of 45 seconds. A few LTAS have been made with shorter averaging times, and it seems that even with an averaging time as short as 20 seconds the average spectrum will differ only slightly from average spectra with a longer averaging time.

Future investigations will, among other things, concern the relation between the averaging time and the variations of the LTASes.

5. Summary

The LTAS analysis can now be performed easily in real time with a new type "Real-Time-Narrow-Band-Analyzer". The instrument has been tested on speech material. Preliminary results show that the LTAS analysis seems to be a good way of obtaining information about the acoustic changes in voice quality, e.g. as a diagnostic equipment in speech clinics or as an objective means for the evaluation of voice quality.

References

Blomberg, M. and K. Elenius 1970: "Statistical analysis of speech signals", QPSR 4, p. 1-8

Buch, Nils H. and B. Frøkjær-Jensen 1972: "Some remarks on acoustic parameters in speech disorders", ARIPUC 6, p. 245-259

Dunn, H.K. and S.D. White 1940: "Statistical measurements on conversational speech", JASA 11, p. 278-288

French, N.R. and J.C. Steinberg 1947: "Factors governing the intelligibility of speech sounds", JASA 19, p. 90-119

Fritzell, Bjørn, O. Hallén, and Johan Sundberg 1974: "Evaluation of teflon injection therapy for paralytic dysphonia", QPSR 1, p. 18-24

Jansson, E.V. and Johan Sundberg 1973: "Organ stops analyzed by means of Long-Time-Average-Spectra", QPSR 4, p. 48-53

Winckel, F. 1967: "Darstellung des Sprechverhaltens als Statistische Tonhöhen und Formantverteilung mittels "Langzeitsanalyse"", Proc.Phon. 4, p. 1031-1035

Winckel, F. and M. Krause 1965: "Ermittlung von Spracheigenschaften aus Statistischen Verteilungen von Amplitude und Tonhöhe", Proc. Acoust. 5 (Liege) paper A41.