AN EXPERIMENTAL SEGMENT SPECTROGRAPH BASED ON SOME NOTES ON FREQUENCY ANALYSIS OF SPEECH SEGMENTS

by Ton Wempe

SUMMARY

Commonly used displays of signal segment spectra like 'line spectra' mostly offer poor readability of various speech signal parameters.

Some ideas for obtaining spectra with greatly improved readability are described.

Experiments using hardware were carried out for implementation of these methods. For estimation of inaccuracy in practice, artificial 'speech signals' with known parameters were used. Some of their spectra are shown as examples. Additionally, spectra of some natural speech segments are displayed which resulted from the methods as described.

The construction of a segment spectrograph based on one of the methods was finished in 1976 and has been in regular use since.

1. INTRODUCTION

In general, the analysis of segments of speech sound is not fundamentally different from a description using a set of universal parameters. However, parameters suitable for the investigation of certain segments of speech (which is non-stationary) maybe unsuitable for other segments, depending on the location and duration of the selected segment. Besides, the choice of parameters depends on the aim of the analysis. A particular set of parameters may result in an analysis which could be completely satisfying for a mathematician but not necessarily for a physicist.

To meet the acoustic phonetician's requirements for an analysis we have to answer the question: what do we want to measure? Or in other words: when selecting a speech segment, what are we looking for?

When speaking of frequency analysis we naturally concentrate on the more stationary speech sections in order to give the term 'frequency' some physical relevance.
So we are particularly concerned with the voiced segments.

The determination of formant frequencies in the description of this class of speech sound still plays an important part.

Defining formants as the maxima in the system function of the vocal tract confronts us at once with the problem of extracting formants from the speech signal.
In the speech production model the excitation source (glottal pulse) drives the linear filter system consisting of the vocal tract filter and the radiation filter.
The formants as defined above, are system parameters, whereas, from the output, only signal parameters can be extracted. This would not be a problem if the excitation source could be considered to deliver a suitable test signal or, at least, a signal with known parameters. The fact that this is not the case is the main reason for the difficulties in formant extracting.

From a signal analysis point of view the problem doesn't exist: if a formant is undetectable owing to the properties of the excitation source, the signal present contains insufficient information for extracting that formant.

Therefore we contend ourselves with frequency distribution diagrams which are valid for the selected speech segments. Like many others, we will call the peaks of these diagrams 'formants' for the sake of convenience...

Although theoretically frequency analysis is applicable to any length of the signal segment, it is mostly short segments we are dealing with in order to be able to make a more detailed analysis by means of many contiguous short time frequency spectra along the time axis rather than a long time average spectrum of a large signal portion.

A rapidly varying formant could change with a velocity of 30 Hz/msec (according to an example from Markel, 1972), which means that the frequency shift between two sequential periods of the fundamental frequency of a male voice could be about 300 Hz!

In our effort in making the analysis more detailed, we would like to shorten the segments to be analysed. However, if we continue to shorten the signal time intervals, the term 'frequency' becomes gradually more meaningless. According to the uncertainty principle the inaccuracy of the determination of a frequency component is about $1/T_s$ Hz where T_s represents the length of the segment analysed.
In cases of rather steady parts of vowel sounds, it seems therefore advisable to take segments of longer duration in order to be able to measure frequency peaks with greater accuracy. Because of the periodicity of the signal within the steady segment, we might just as well store one F_0-period in a signal memory, which is afterwards scanned repetively so that a continuous periodic signal arises of any desired duration.

The frequency domain representations of periodic signals however, are 'line spectra' with constant line distances equal to F_0 Hz.

A spectrum of this kind is to be considered as a sampled version of a continuous spectrum of one F_0-period. It will be clear that as far as the accuracy of formant extracting goes, there is no point in taking segments of longer duration than one F_0-period.

The frequency components in the frequency domain representation of signals are of infinite duration whereas the signal segments are definitely not so.

If we think of a signal segment as being the result of a multiplication of a continuous time function with a rectangular 'window' function of unit amplitude and of the same duration as the signal segment, the spectrum of the segment is found by convolving the individual spectra of the multiplied functions. The spectrum of a rectangular 'window' function is a $(\sin\pi fT_0/\pi fT_0)$ function and has zero components at multiples of $1/T_0$. The result after the convolution is a spectrum with additional maxima and minima ('lobes').

Because the readability of formants can be poor owing to the occurrence of these 'lobes', a number of different 'windows' have been applied (Gauss, Hanning, Hamming) in order to eliminate these 'lobes' or reduce their amplitudes. In addition various (computerized) smoothing techniques have been developed to smoothe this 'F_0-ripple' or to convert the 'spectral line' output of the FFT-program into a continuous graph.

From a physical point of view we should prefer perhaps the use of time-limited sinusoidal components, each with a length equal to the speech segment, rather than continuous ones. Apart from the accuracy of frequency determination, the different lengths of the segments then have no further consequences as regards the shape of their spectra.

Anyway, our aim was to try to develop relatively simple hardware which should:

1. approximate the amplitude spectrum of one F_0-period with <u>op-</u>

timum readability of the formants and with known inaccuracy. Information about the formant amplitudes and bandwidths are thus included;

2. handle segments of various lengths (i.e. different F_0-periods) at optimum resolution;
3. be easy to operate.

2. FIRST APPROXIMATION

The insertion of time (T_a) into each period (T_0) of the fundamental frequency of a periodic vowel sound was the first idea in approximating the continuous spectrum.

The effect of this is that the distances between the spectral lines is decreased because the modified fundamental (F_0') equals $1/(T_0+T_a)$. The more T_a is increased, the better the continuous spectrum approximation becomes.

To get some insight in the shape of the continuous spectrum of one F_0-period of a vowel sound, we define a simplified speech segment as a damped sinusoid, which is started at t=0 and truncated at $t=T_0$:

$$(1) \qquad g(t)=e^{-\alpha t} \sin\omega_1 t \qquad\qquad 0 < t < T_o$$

Using Euler's relation g(t) changes to:

$$g(t)= \frac{1}{2j}\left[e^{-(\alpha-j\omega_1)t} - e^{-(\alpha+j\omega_1)t}\right] \qquad\qquad 0 < t < T_o$$

Its Fourier transform is:

$$G(\omega)= \frac{1}{2j}\left[\int_o^{T_o} e^{(-\alpha+j\omega_1-j\omega)t}\,dt - \int_o^{T_o} e^{(-\alpha-j\omega_1-j\omega)t}\,dt\right]$$

Evaluating leads to:

$$(2) \qquad G(\omega)=\frac{1}{2j}\left[\frac{1-e^{-[\alpha+j(\omega-\omega_1)]T_o}}{\alpha+j(\omega-\omega_1)}-\frac{1-e^{-[\alpha+j(\omega+\omega_1)]T_o}}{\alpha+j(\omega+\omega_1)}\right]$$

This can be written as the subtraction of two functions:

$$(3) \qquad G(\omega)=G_e(\omega) - G_m(\omega) \text{ , where:}$$

$$G_e(\omega)=\frac{1}{2j}\left[\frac{1}{\alpha+j(\omega-\omega_1)}-\frac{1}{\alpha+j(\omega+\omega_1)}\right]$$

$$G_m(\omega)=\frac{e^{-(\alpha+j\omega)T_o}}{2j}\left[\frac{e^{j\omega_1 T_o}}{\alpha+j(\omega-\omega_1)}-\frac{e^{-j\omega_1 T_o}}{\alpha+j(\omega+\omega_1)}\right]$$

Or, in trigonometric form:

$$(4) \qquad G_e(\omega)=\frac{\omega_1}{(\alpha+j\omega)^2+\omega_1^2}$$

$$(5) \qquad G_m(\omega)=e^{-(\alpha+j\omega)T_o}.\frac{(\alpha+j\omega)\sin\omega_1 T_o+\omega_1\cos\omega_1 T_o}{(\alpha+j\omega)^2+\omega_1^2}$$

The latter function becomes zero if T_0 tends to infinity, so that $G_e(\omega)$ represents the Fourier transform of an untruncated damped sinusoid.

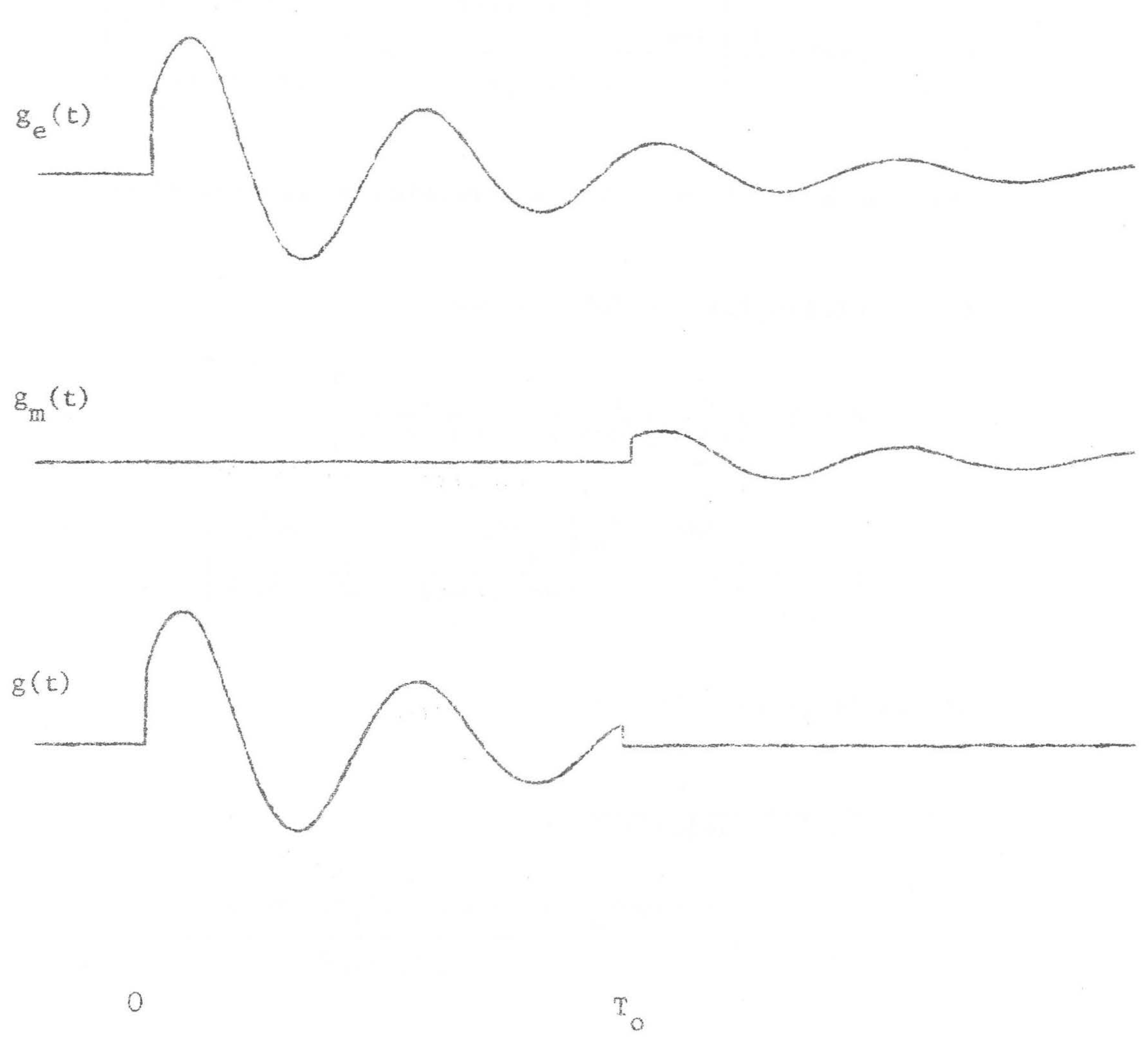
From the superposition property of the Fourier transform it follows that:

$$g(t)=g_e(t) - g_m(t)$$

which causes $g_m(t)$ to be:

$$g_m(t)=e^{-\alpha t}\sin\omega_1 t \qquad\qquad T_o < t < \infty$$

(See fig. 1)



$$g(t) = g_e(t) - g_m(t)$$

Fig.1

The function $g_m(t)$ can be thought of as a time-shifted version of $g_e(t)$ with an initial amplitude of $e^{-\alpha T_o}$ and a phase-shifted origin.

The amplitude spectrum of $G_e(\omega)$ is determined by:

(6) $$|G_e(\omega)| = \frac{\omega_1}{\sqrt{(\alpha^2+\omega_1^2-\omega^2)^2+4\alpha^2\omega^2}}$$

This function exposes maxima at

(7) $$\omega=\omega_m = \pm\sqrt{\omega_1^2-\alpha^2}$$

(8) The magnitudes of these maxima are $1/2\alpha$.
To derive the 3 dB bandwidth of the peaks we solve the equation:

$$\frac{\omega_1}{\sqrt{(\alpha^2+\omega_1^2-\omega^2)^2+4\alpha^2\omega^2}} = \frac{1}{2\alpha\sqrt{2}}$$

which delivers:

$$\omega_h = \sqrt{\omega_1^2-\alpha^2+2\alpha\omega_1}$$

$$\omega_l = \sqrt{\omega_1^2-\alpha^2-2\alpha\omega_1}$$

As long as $2\alpha^2 << \omega_1^2 - 2\alpha\omega_1$ we may write:

$$\omega_h = \sqrt{\omega_1^2+\alpha^2+2\alpha\omega_1} = \pm(\omega_1+\alpha)$$

$$\omega_l = \sqrt{\omega_1^2+\alpha^2-2\alpha\omega_1} = \pm(\omega_1-\alpha)$$

which means that $\omega_h - \omega_1 = 2\alpha$

(9) or: $B = \frac{\alpha}{\pi}$ Hz

The inaccuracy of the approximation in cases of practical values of ω_1 and α is rather small: for example in the unfavourable case if $f_1=300$ Hz and $\alpha=100\ \pi$, the deviation amounts to less than 3%. The shape of $|G_e(\omega)|$ is outlined in fig. 2.

The shape of the function $|G_m(\omega)|$ depends on the phase angle $\omega_1 T_0$.

If $\omega_1 T_0 = n\pi$(where n is an integer):

(10) $$|G_{m1}(\omega)| = e^{-\alpha T_o} |G_e(\omega)|$$

and in this case $|G_m(\omega)|$ is simply an attenuated version of $|G_e(\omega)|$.

If $\omega_1 T_0 = \pi(n+\frac{1}{2})$:

(11) $$|G_{m2}(\omega)| = \frac{\sqrt{\alpha^2+\omega^2}}{\omega_1} |G_{m1}(\omega)|$$

$|G_{m2}(\omega)|$ only equals $|G_{m1}(\omega)|$ if $\omega^2 = \omega_1^2 - \alpha^2$, viz at the maxima of $|G_{m1}(\omega)|$.

For values of ω where $\omega^2 >> \alpha^2$ the spectrum $|G_{m2}(\omega)|$ compared to $|G_{m1}(\omega)|$ increases linearly with increasing ω, according to 6 dB/ octave.

Maxima occur at

(12) $$\omega_{max} = \sqrt{\omega_1 \sqrt{\omega_1^2 + 4\alpha^2} - \alpha^2}$$

or: $$\omega^2_{max} = \sqrt{(\omega_1^2 + 2\alpha^2)^2 - 4\alpha^4} - \alpha^2$$

For practical values of ω_1 and α, we may neglect the term $4\alpha^4$ in respect of $(\omega_1^2+2\alpha^2)^2$ so that

$$\omega_{\max}^2 = \omega_1^2+\alpha^2 \quad \text{and}$$

$$(13) \qquad \omega_{\max} = \pm\sqrt{\omega_1^2+\alpha^2}$$

The height of the maxima is:

$$|G_{m2}(\omega)|_{\max} = \sqrt{\frac{1}{2\omega_1[\sqrt{\omega_1^2+4\alpha^2}-\omega_1]}}$$

Again we may write:

$$|G_{m2}(\omega)|_{\max}^2 = \frac{1}{2\sqrt{\omega_1^4+4\alpha^2\omega_1^2+4\alpha^4-4\alpha^4}-2\omega_1^2} =$$

$$= \frac{1}{2(\omega_1^2+2\alpha^2)-2\omega_1^2} = \frac{1}{4\alpha^2}$$

$$(14) \qquad \text{Hence:} \qquad |G_{m2}(\omega)|_{\max} = \frac{1}{2\alpha}$$

Which is equal to the maxima of $|G_e(\omega)|$.

The 3 dB bandwidth can be found by solving the equation:

$$\frac{\alpha^2+\omega^2}{(\alpha^2+\omega_1^2-\omega^2)^2+4\alpha^2\omega^2} = \frac{1}{4\omega_1[\sqrt{\omega_1^2+4\alpha^2}-\omega_1^2]}$$

which delivers:

$$\omega_h^2 = 2p^2-\omega_1^2-\alpha^2+\sqrt{\omega_1^4+3p^4-4\omega_1^2p^2}$$

$$\omega_l^2 = 2p^2-\omega_1^2-\alpha^2-\sqrt{\omega_1^4+3p^4-4\omega_1^2p^2}$$

$$(15) \qquad \text{where} \qquad p^2 = \omega_1\sqrt{\omega_1^2+4\alpha^2}$$

Let be $\omega_h^2 = x+y$ and $\omega_1^2 = x-y$, then:

$$\omega_h - \omega_1 = \sqrt{x+y} - \sqrt{x-y}$$

$$(\omega_h - \omega_1)^2 = 2x - 2\sqrt{x^2 - y^2}$$

so that $(\omega_h - \omega_1)^2 = 4p^2 - 2\omega_1^2 - 2\alpha^2 - 2\sqrt{(2p^2 - \omega_1^2 - \alpha^2)^2 - \omega_1^4 - 3p^4 + 4\omega_1^2 p^2}$

Substituting of (15) and some minor manipulations yield:

$$(\omega_h - \omega_1)^2 = 4\omega_1\sqrt{\omega_1^2 + 4\alpha^2} - 2\omega_1^2 - 2\alpha^2 - 2\sqrt{(\omega_1^2 + \alpha^2)^2 + 4\alpha^2\omega_1^2 (1 - \sqrt{1 + 4\frac{\alpha^2}{\omega_1^2}})}$$

For practical values of ω_1 and α applies:

$$4\alpha^2\omega_1^2 (1 - \sqrt{1 + 4\frac{\alpha^2}{\omega_1^2}}) << (\omega_1^2 + \alpha^2)^2$$

This allows the simplification:

$$(\omega_h - \omega_1)^2 = 4\omega_1\sqrt{\omega_1^2 + 4\alpha^2} - 4\omega_1^2 - 4\alpha^2$$

Again we write:

$$(\omega_h - \omega_1)^2 = 4\sqrt{\omega_1^4 + 4\alpha^2\omega_1^2 + 4\alpha^4 - 4\alpha^4} - 4\omega_1^2 - 4\alpha^2$$

and neglect $-4\alpha^4$: $(\omega_h - \omega_1)^2 = 4(\omega_1^2 + 2\alpha^2) - 4\omega_1^2 - 4\alpha^2 = 4\alpha^2$

(16) Hence: $\omega_h - \omega_1 = 2\alpha$ or $B = \frac{\alpha}{\pi}$ Hz

So, in practice the bandwidth of the peak is the same as in the case of $|G_e(\omega)|$. (See fig. 3 for outlines of $|G_{m1}(\omega)|$ and $|G_{m2}(\omega)|$). Other values of the phase angle $\omega_1 T_0$ result in amplitude spectra somewhere between these extreme curves.

FIG. 2

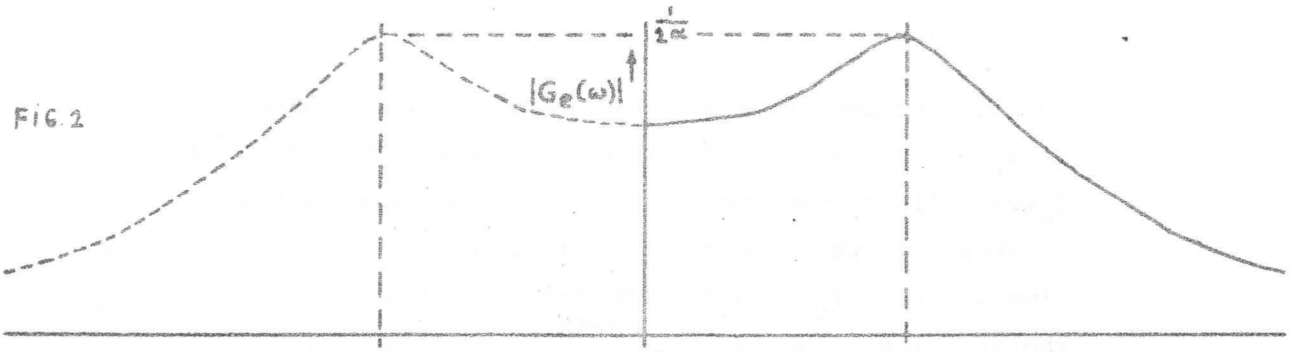


FIG. 3

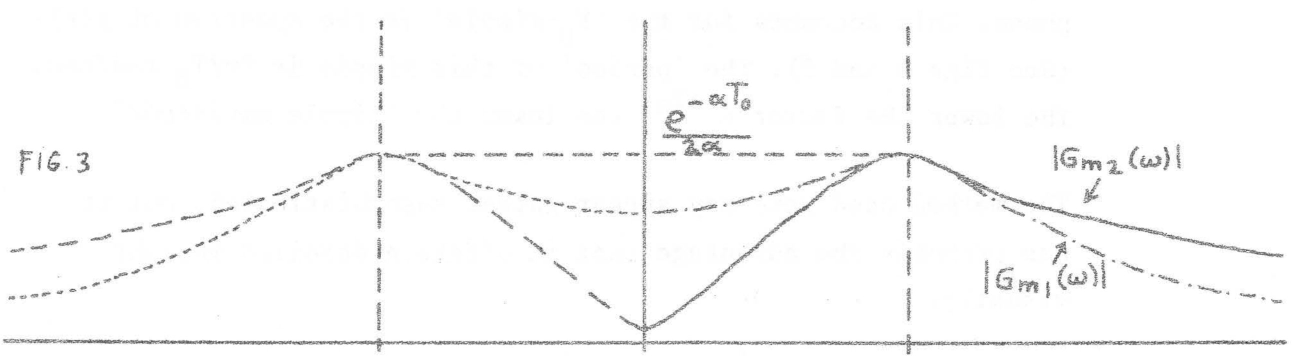


FIG. 4

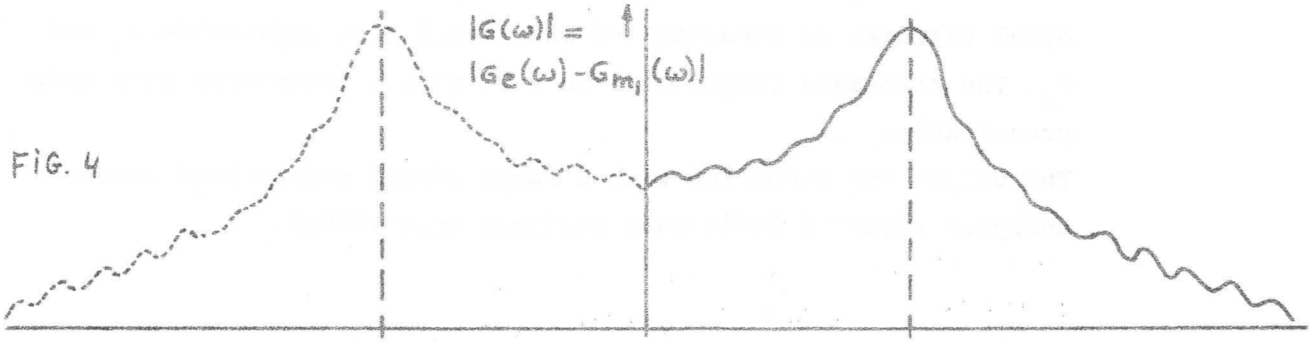
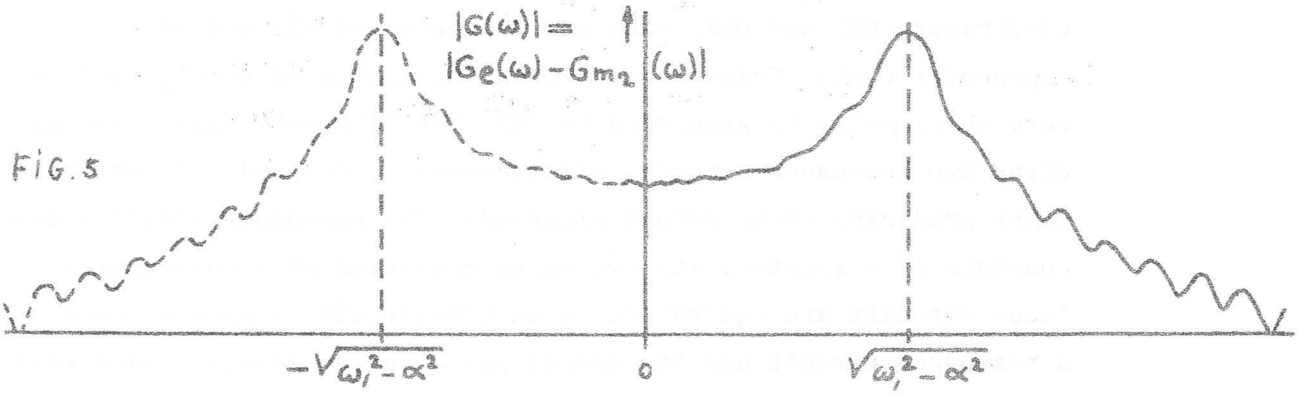


FIG. 5



To find the amplitude spectrum of g(t), each frequency component of $G_m(\omega)$ has to be subtracted from the corresponding component of $G_e(\omega)$ while accounting for the relative phase shifts.
According to the time shift property of the Fourier transform a time shift of T_0 seconds corresponds to a phase shift of $-\omega T_0$ radians (hence the factor $e^{-j\omega T_o}$ in equation 5).
Thus when subtracting $G_m(\omega)$ from $G_e(\omega)$, maximum deviations from $|G_e(\omega)|$ occur when corresponding components have equal or opposite phase. This accounts for the 'F_0-ripple' in the spectrum of g(t). (See figs 4 and 5). The 'period' of this ripple is $2\pi/T_0$ rad/sec. The lower the factor $e^{-\alpha T_o}$, the lower the 'ripple amplitude'.

The method used here may appear rather unsophisticated, but it has probably the advantage that it offers a detailed insight visually.

To test this time insertion in practice and experience the effect of the 'F_0-ripple' on the readability of formants, a simple artificial vowel generator was made which produced two damped sinusoidal voltages as formants and was fitted with adjustable T_0 and T_a. The resonance frequencies as well as the bandwidths were made presettable.
The output was connected with a swept signal narrow band spectrum analyzer (General Radio wave analyzer type 1900A).

Brief description of the artificial vowel generator (see fig. 6).

A pulse generator is made by means of two coupled monostable multivibrators OS1 and OS2, each one triggered at the end of its opponent's cycle. Coinciding with the beginning of the T_0-cycle a very short pulse is generated by OS3 ($T\delta$= 125µsec). This pulse excites two resonance circuits simultaneously, GC1 and GC2, which start producing their damped sinusoids. The resonance circuits each consists of a gyrator, its two gates connected with capacitors. These circuits are equivalent to an LCR-circuit. A gyrator used as a resonance circuit has the advantages of being stable, being easy

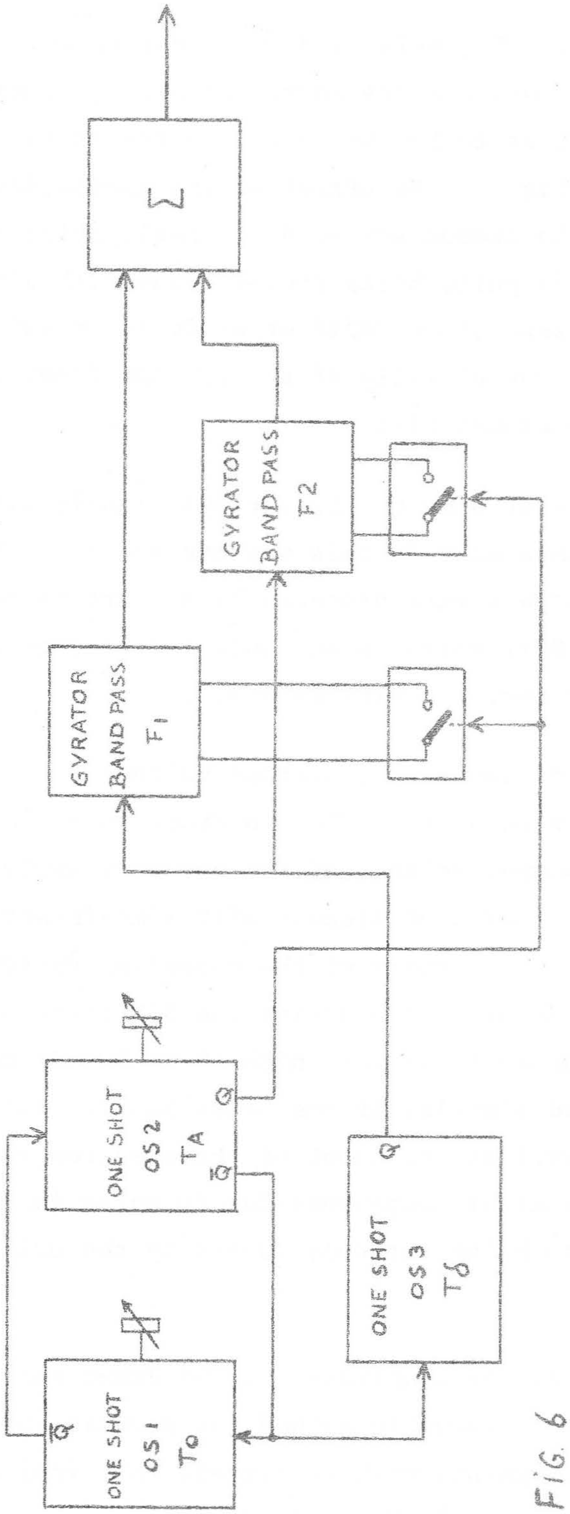

ONE SHOT
OS1
To
Q̄
ONE SHOT
OS2
TA
Q̄
Q
ONE SHOT
OS3
Tδ
Q
GYRATOR
BAND PASS
F1
GYRATOR
BAND PASS
F2
Σ

FIG. 6

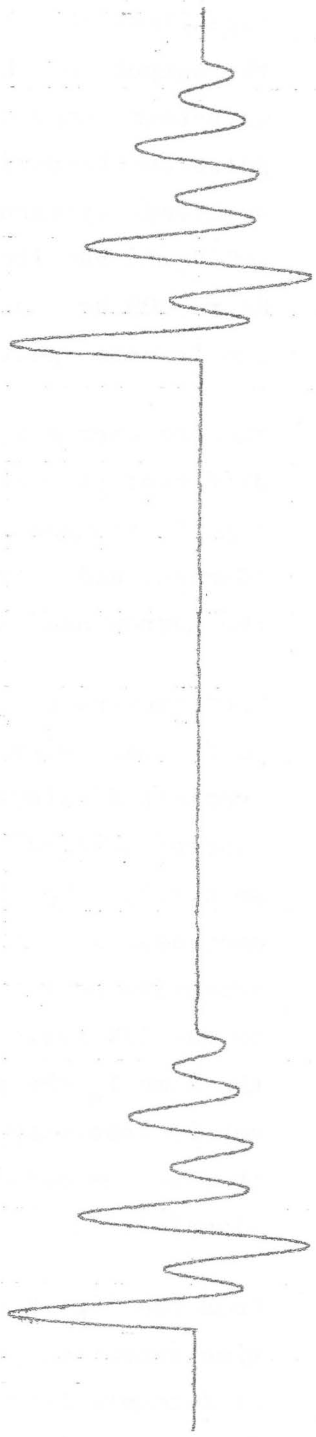


FIG. 7

to construct and has its resonance energy contained in only two capacitors. (See chapter 4 for a description of the gyrator principle).
At the end of a T_0-cycle, a T_a-cycle is started during which the capacitors of the gyrator devices are short-circuited, causing the outputs of these devices to become zero. The result is a wave-form somewhat like fig. 7. The effect of the excitation pulse on the purity of the damped sinusoids is negligible: the amplitude spectrum of this pulse being proportional with sin $\frac{1}{2}\omega T\delta/\frac{1}{2}\omega T\delta$ has its first zero at $\omega = 2\pi/T\delta$ or at $f = 1/T\delta = 10^6/125$ Hz = 8000 Hz, so that in the vicinity of the formant frequencies the 'source spectrum' is almost flat.

The corresponding spectra of some of the various signals with different parameters we measured in this way are shown in the figs 8 through14. The outputs were produced by a chart recorder (General Radio type 1521BQ1) which is mechanically coupled with the narrow band spectrum analyzer already mentioned.

The time-insertion process causes the average voltage of a signal to decrease with a factor equal to T_0'/T_o. In order to achieve convenient displays, the output voltage of the spectrum analyzer was increased accordingly in cases of signals with time-insertion. Naturally, the signal to noise ratio of the measuring system decreases with the same amount; in practice the S/N ratio of selective measurements however, is very high (here 80 dB) compared to the S/N ratio of taped signals. If one bears in mind that during the time T_a the noise level at the input of the spectrum analyzer can be kept very low, eventual components due to noise in the spectrum can be considered as being entirely caused by the original signal itself.

From the figs 8 through 14 the conclusion can be drawn that the time-insertion process is a valuable method for accurate measuring of formants from signal segments with relatively low final amplitudes (i.e. F_0-periods of male vowels). For this last class of signals the forms 7 and 9 are quite applicable.

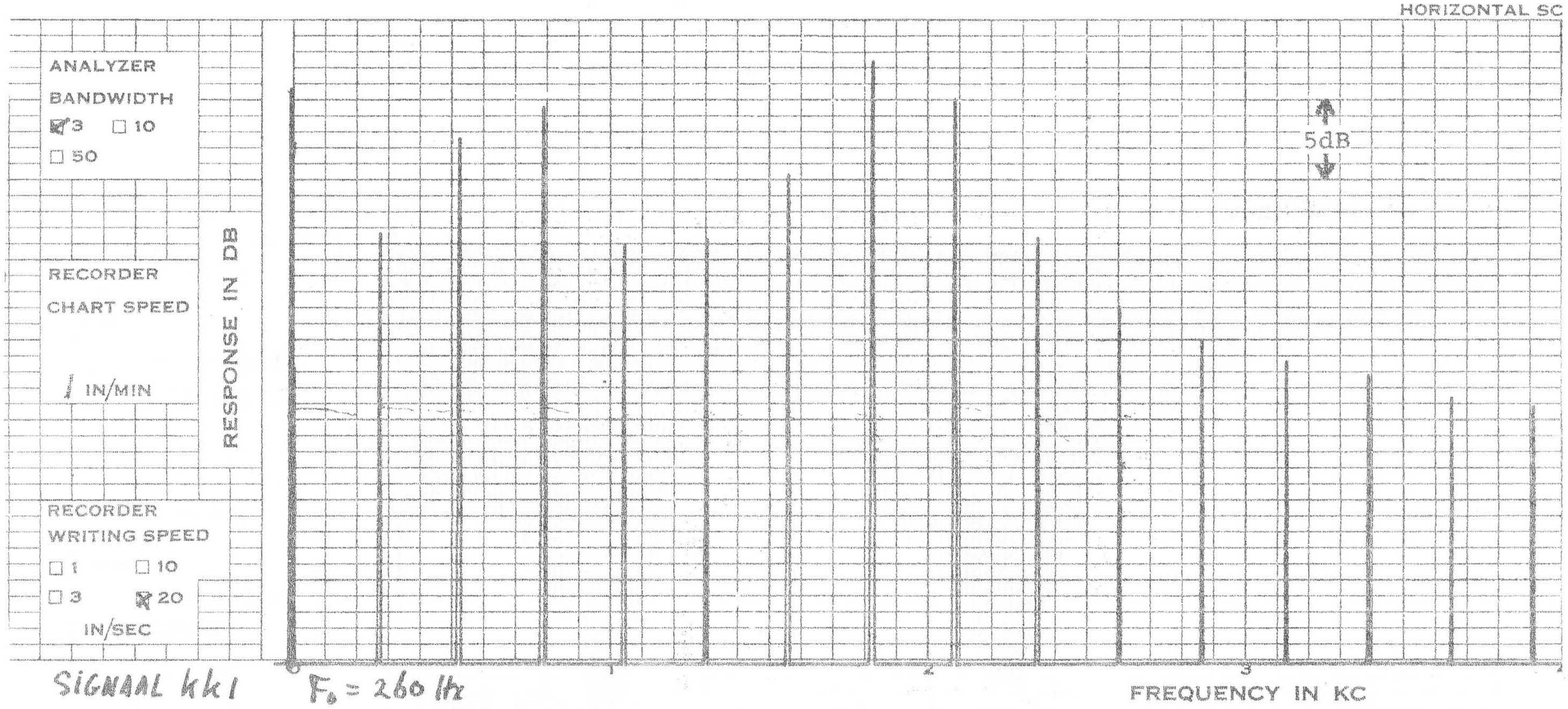


Fig.8 Spectrum of signal KK1, F_o=260Hz, F_1=700Hz, F_2=1950Hz

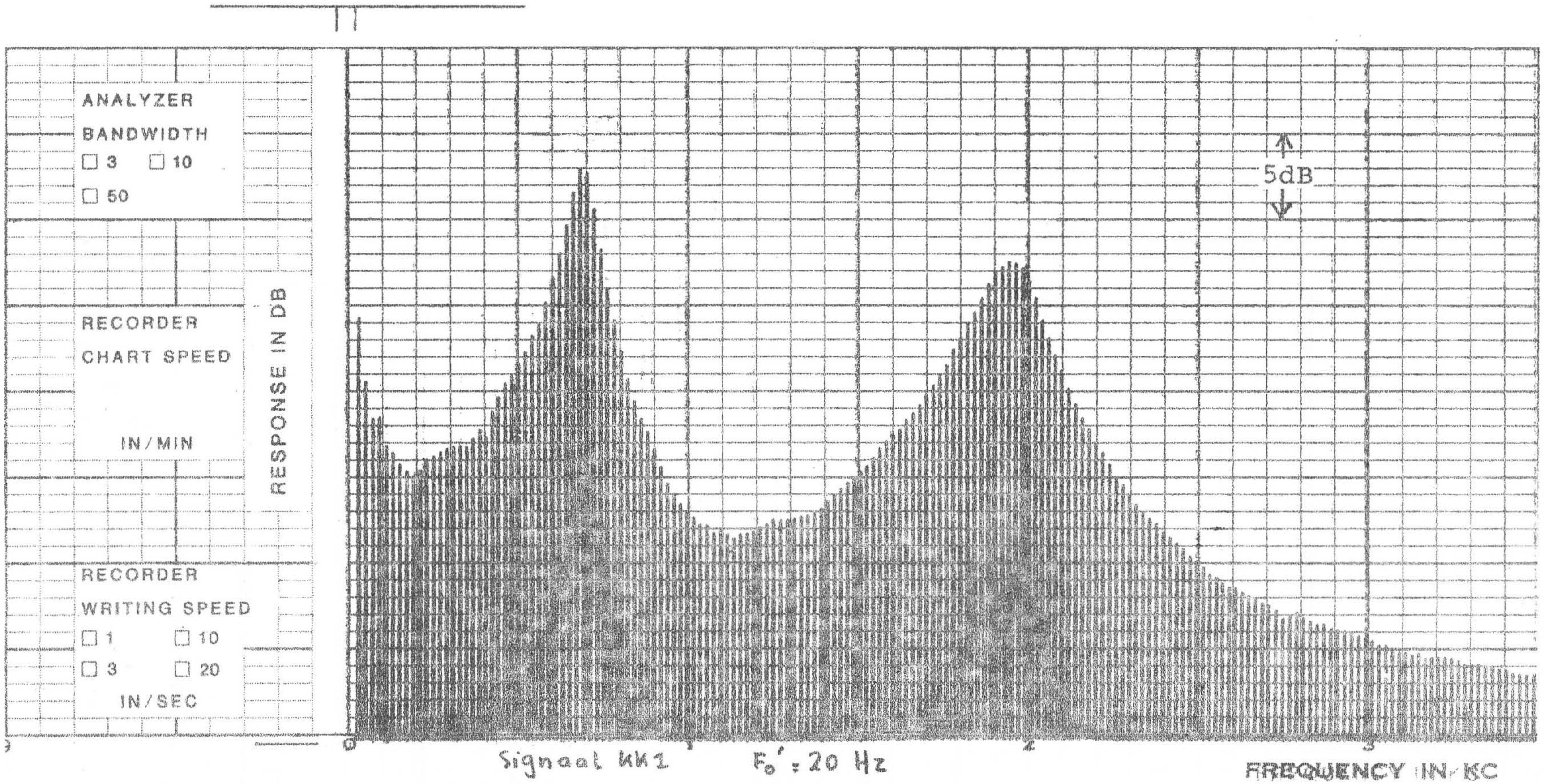


Fig.8a Spectrum of the signal from fig.8 with time-insertion, F'_o=20Hz

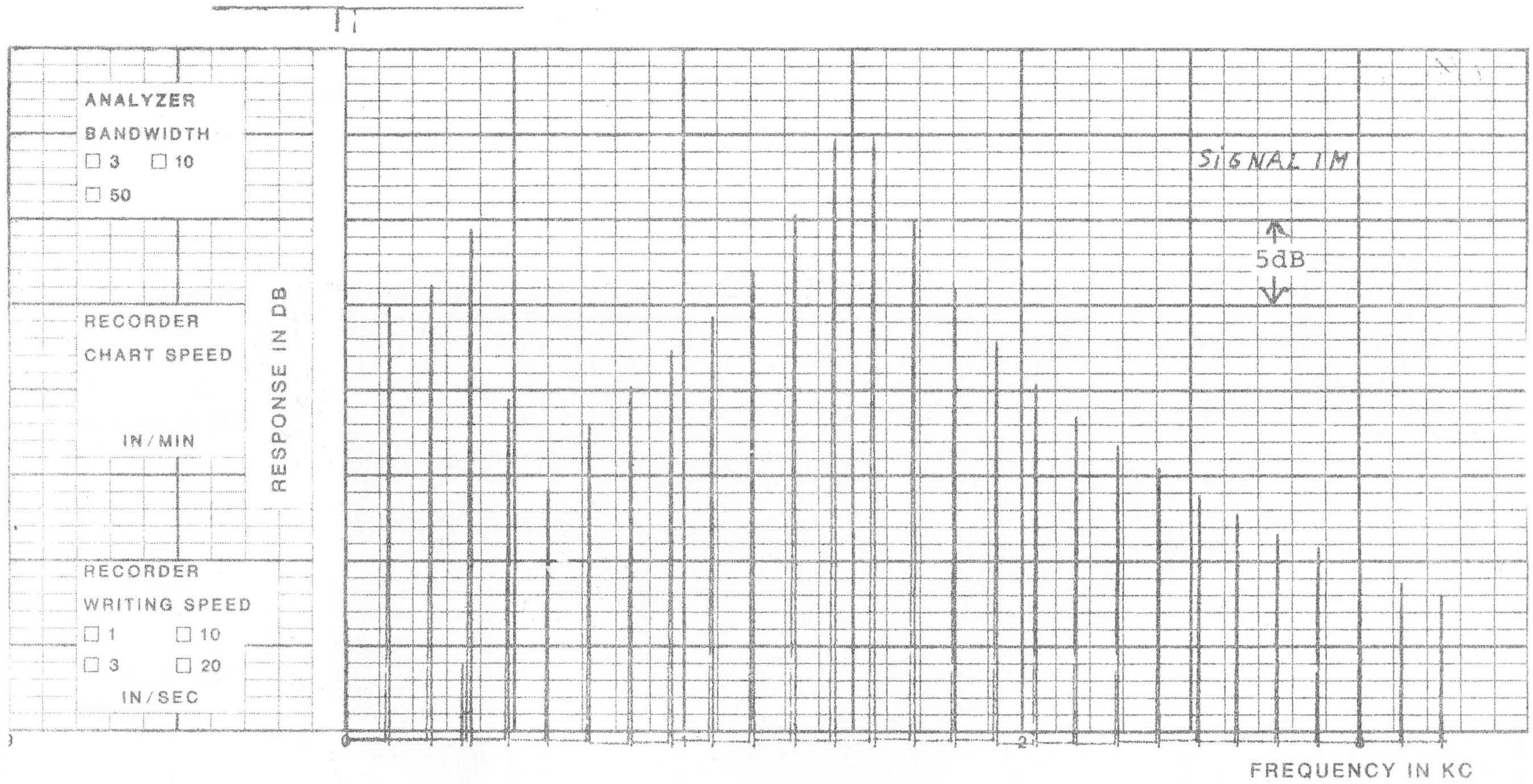


Fig.9 Spectrum of signal IM, F_o=120Hz, F_1=400Hz, F_2=1500Hz

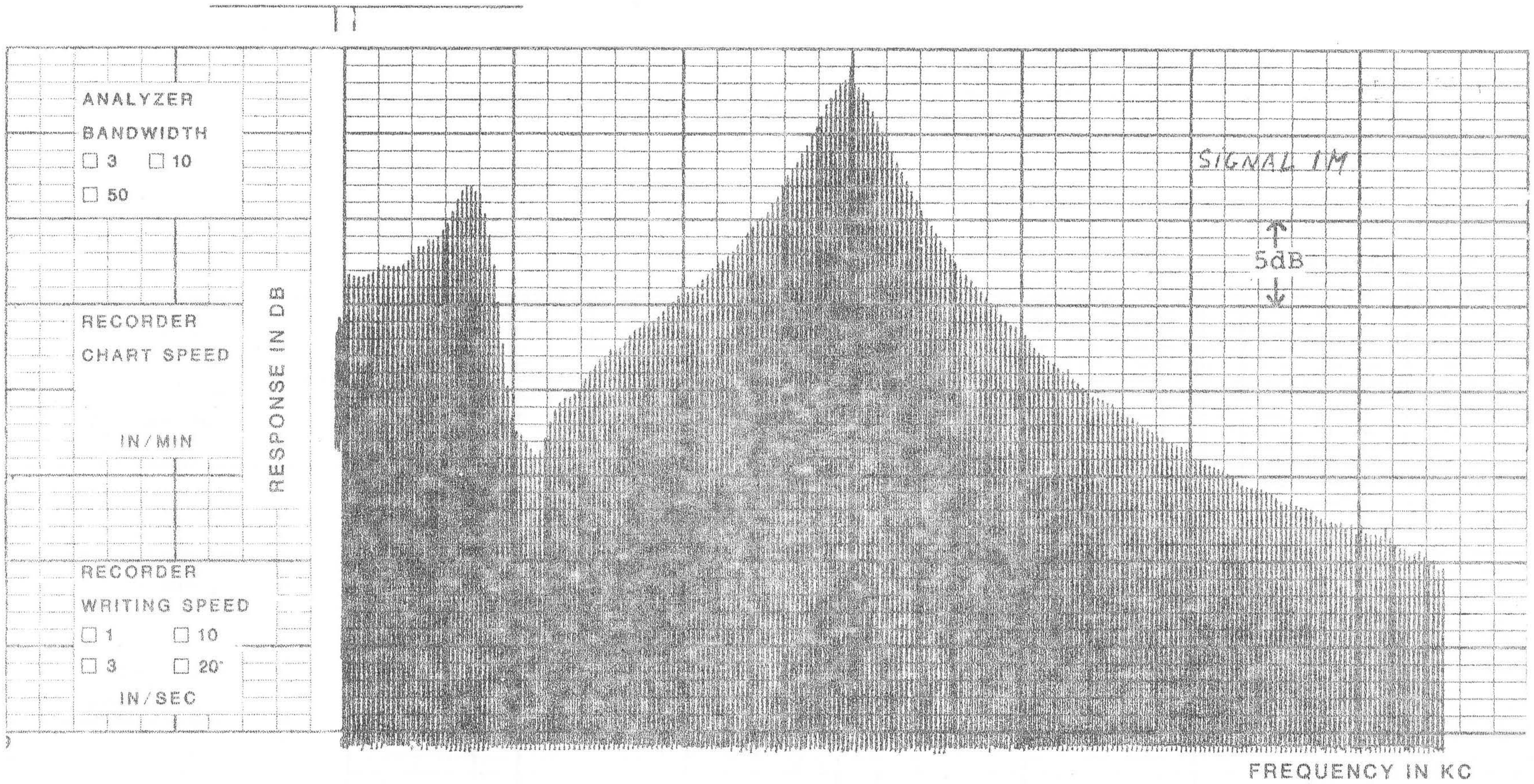


Fig.9a Spectrum of the signal from fig.9 with time-insertion, F_o'=14.4Hz

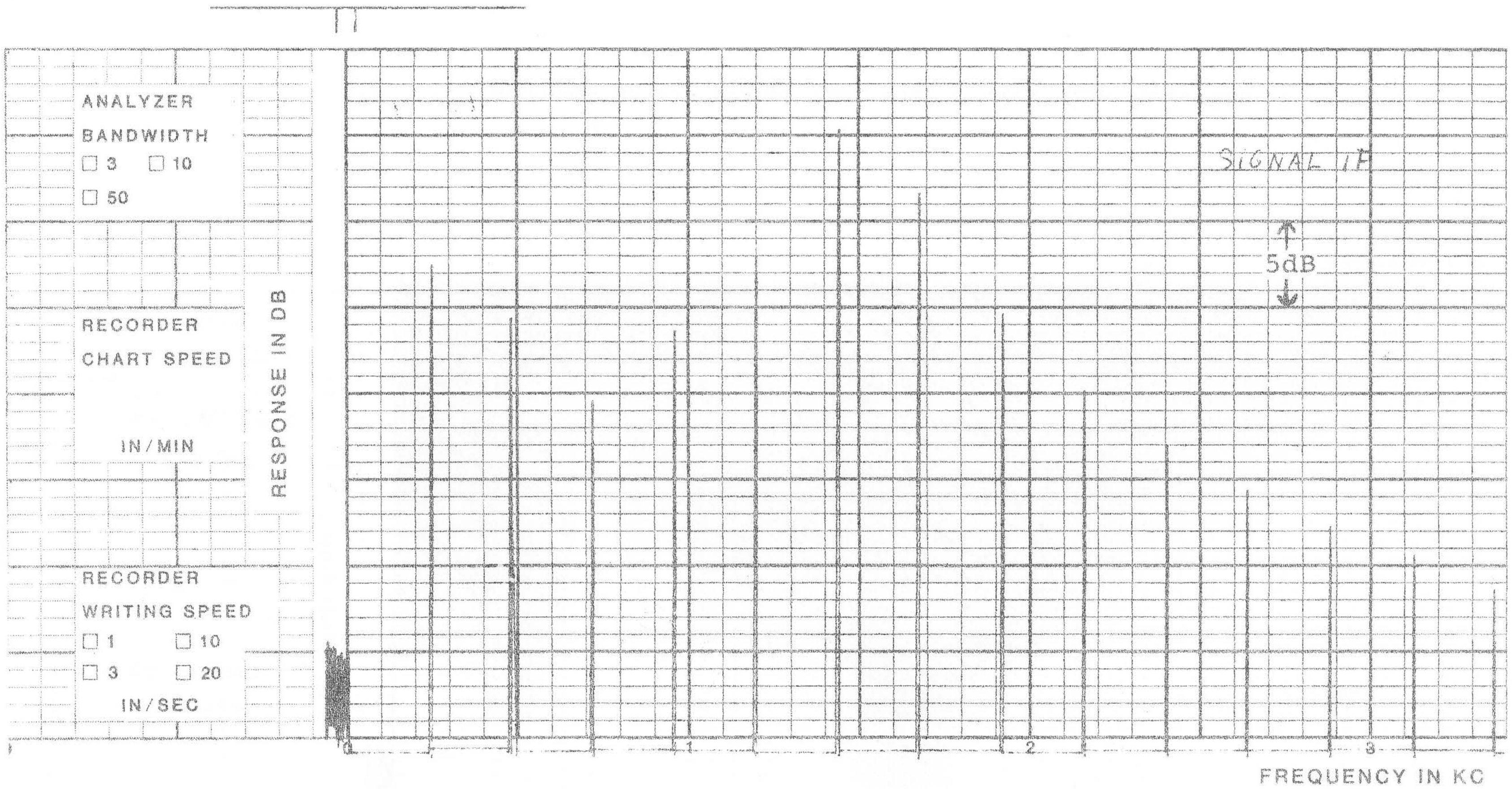


Fig.10 Spectrum of signal 1F, F_o=240Hz, F_1=400Hz, F_2=1500Hz

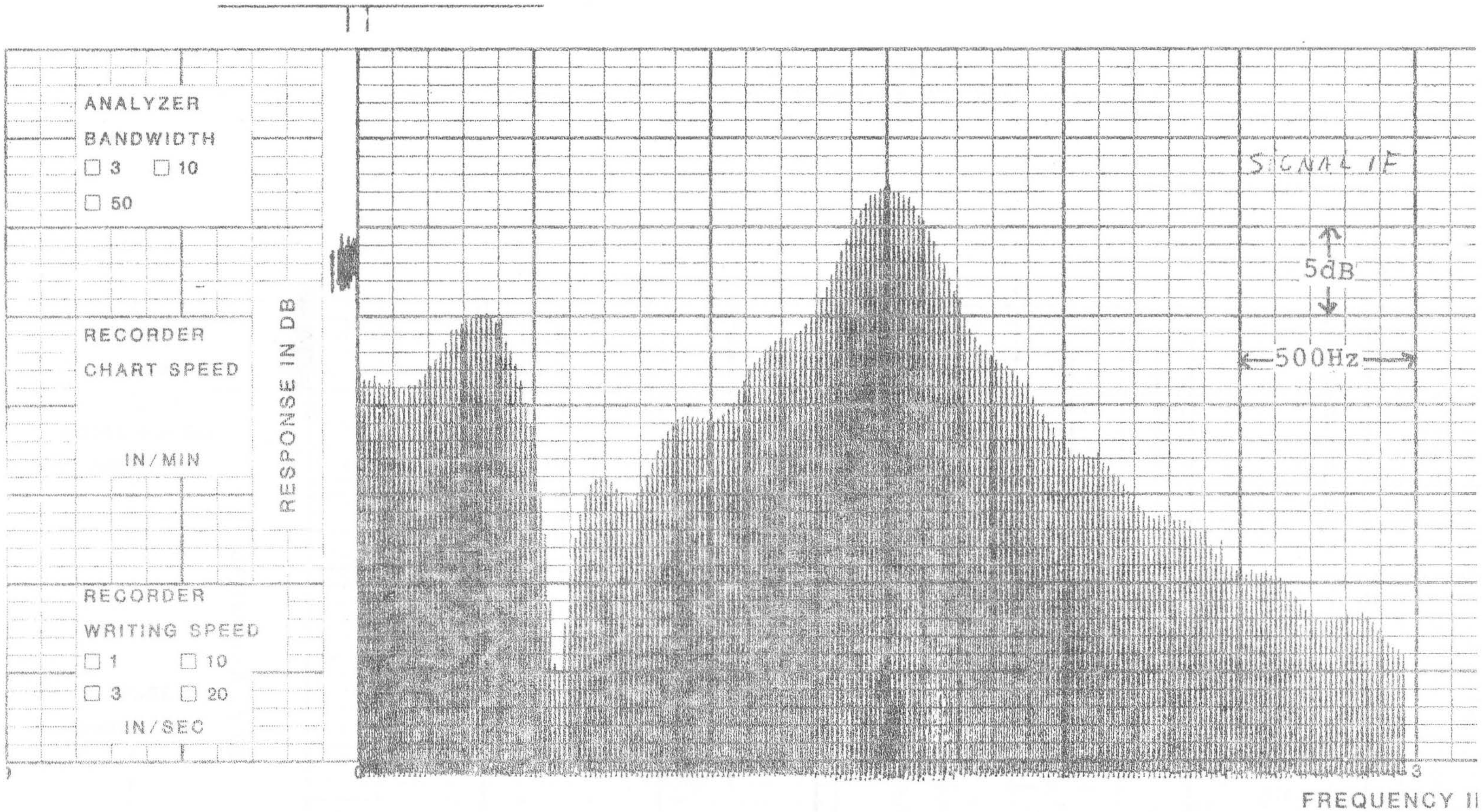
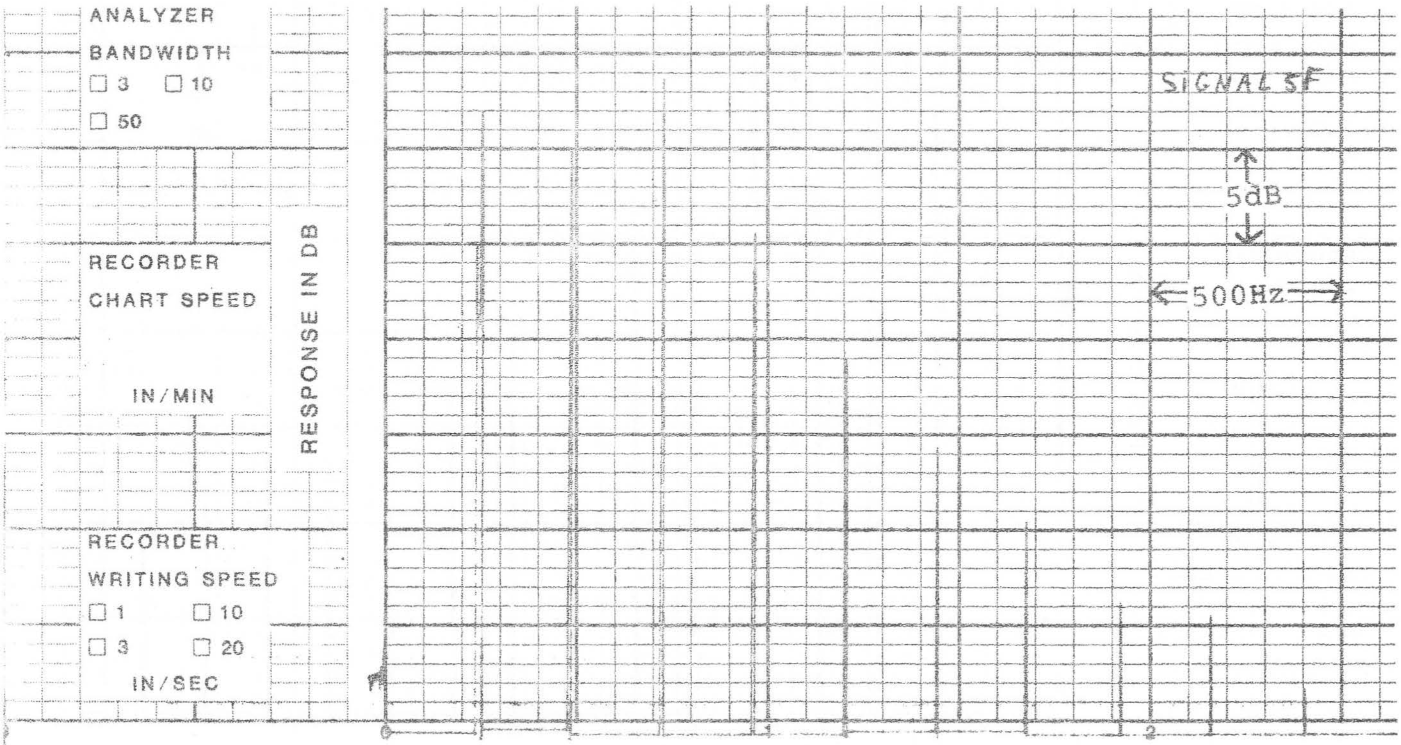


Fig.10a Spectrum of the signal from fig.10 with time-insertion, F_o'=14.4Hz



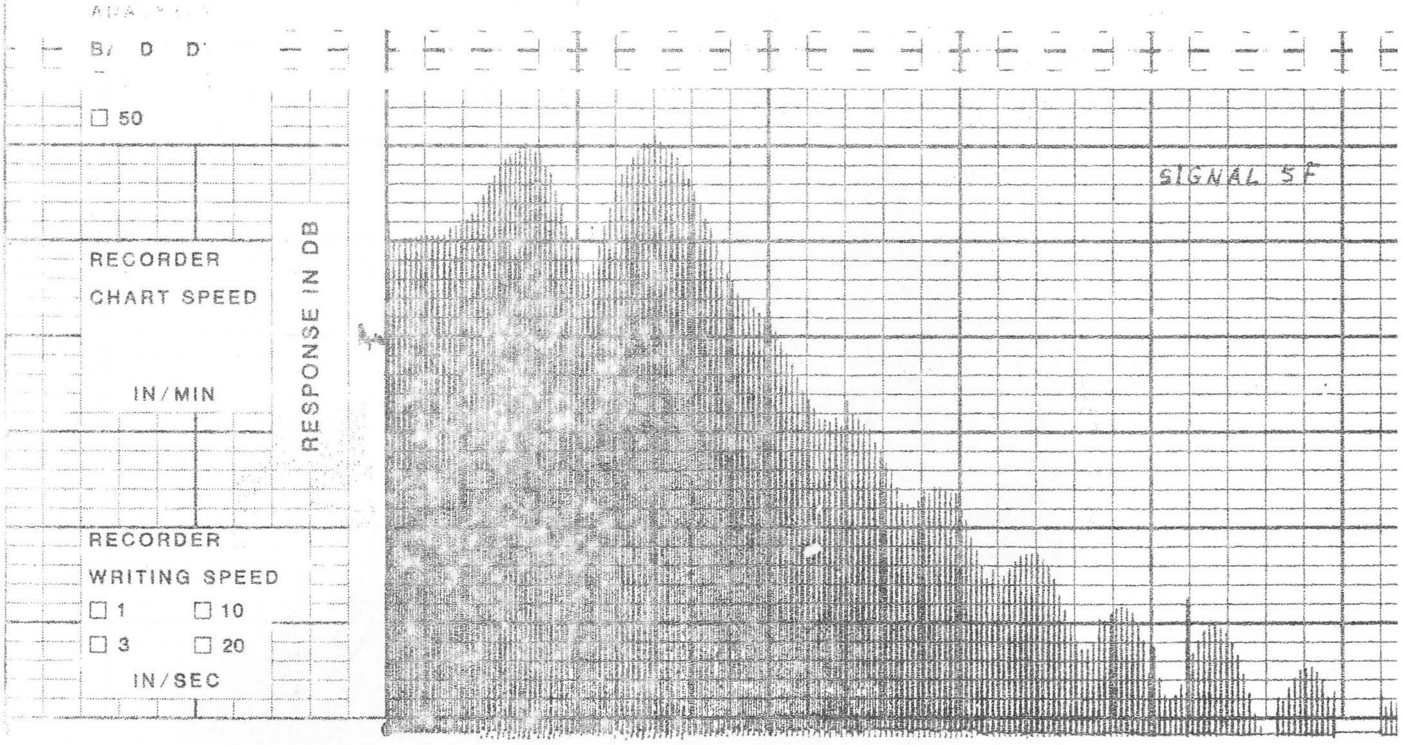
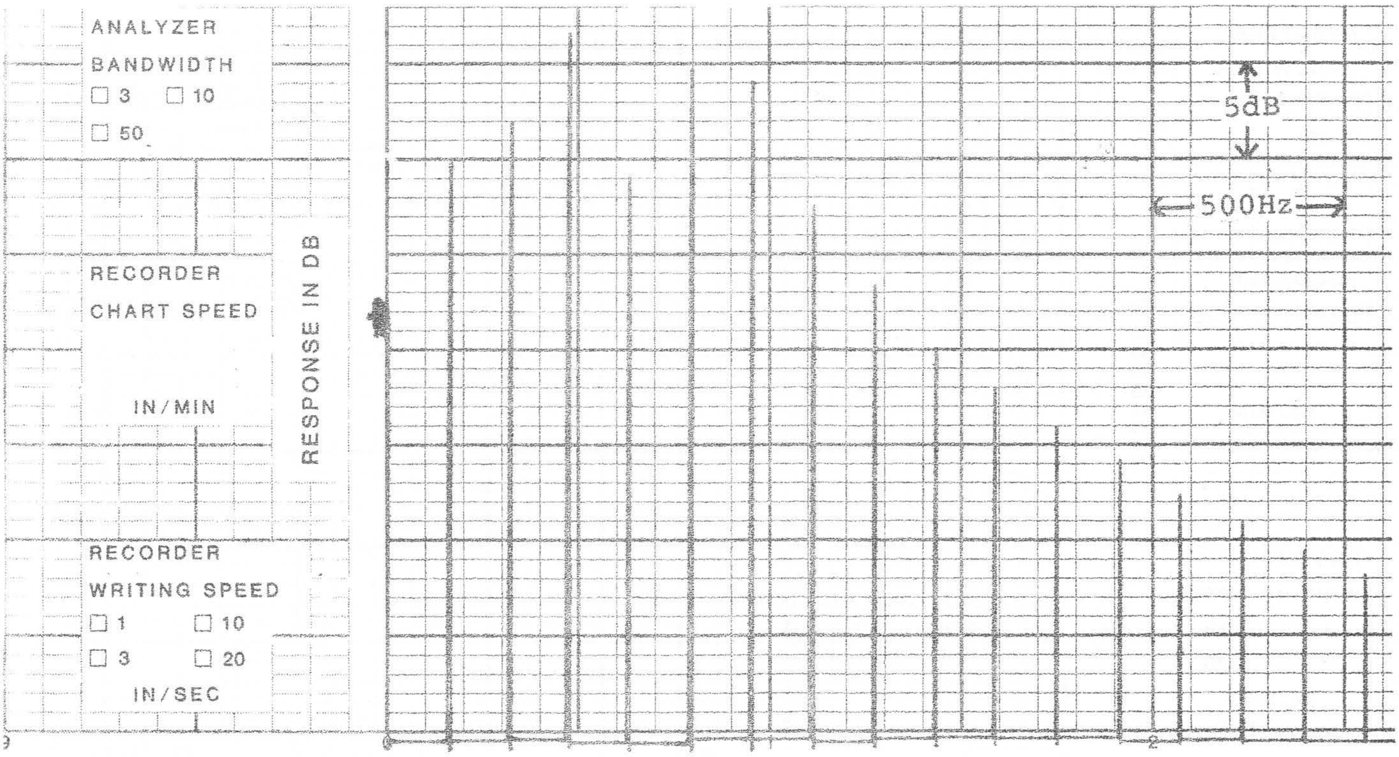


Fig.11 Spectrum of signal 5F, F_o=240Hz, F_1=400Hz, F_2=700Hz

Fig.11a Spectrum of signal 5F with time-insertion, F'_o=14.4Hz



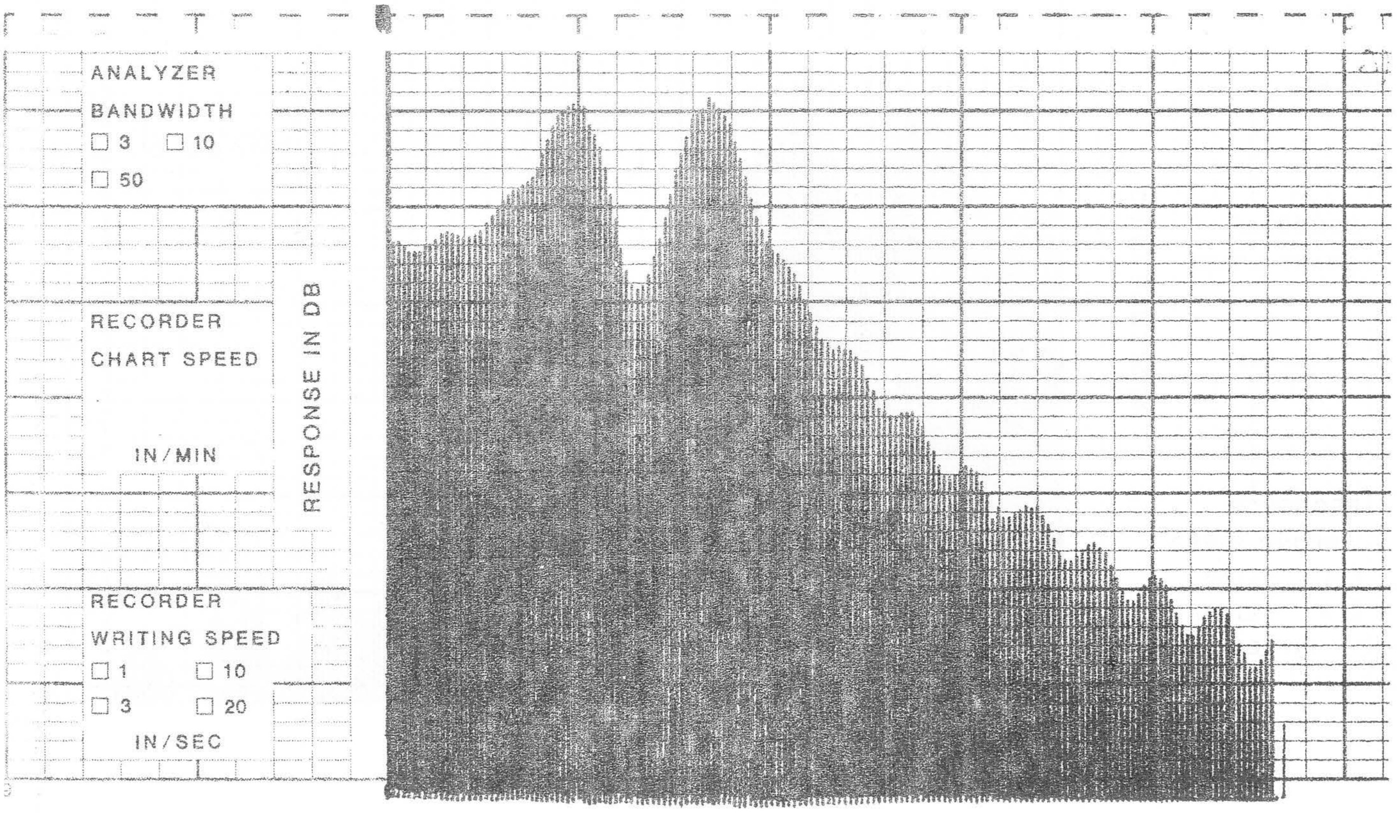
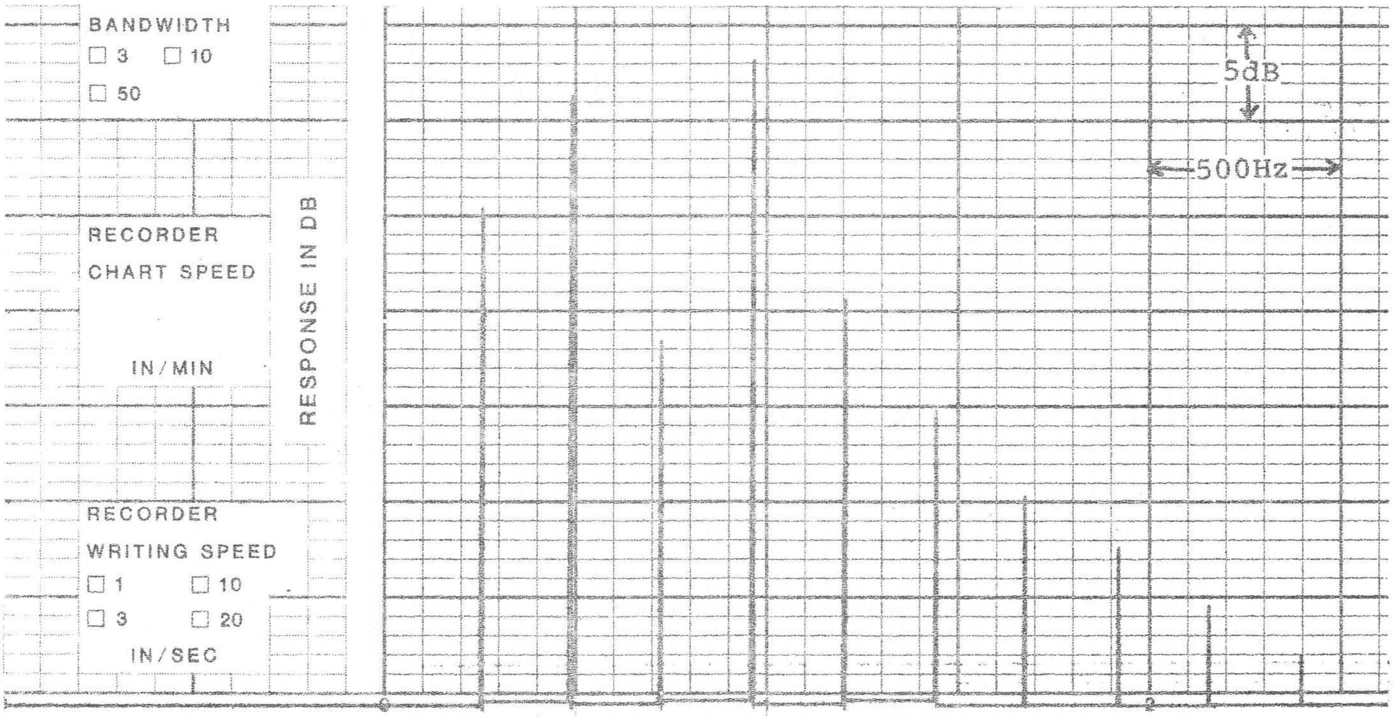


Fig.12 Spectrum of signal 2M, F_o=160Hz, F_1=540Hz, F_2=860Hz

Fig.12a Spectrum of signal 2M with time-insertion, F_o'=14.4Hz



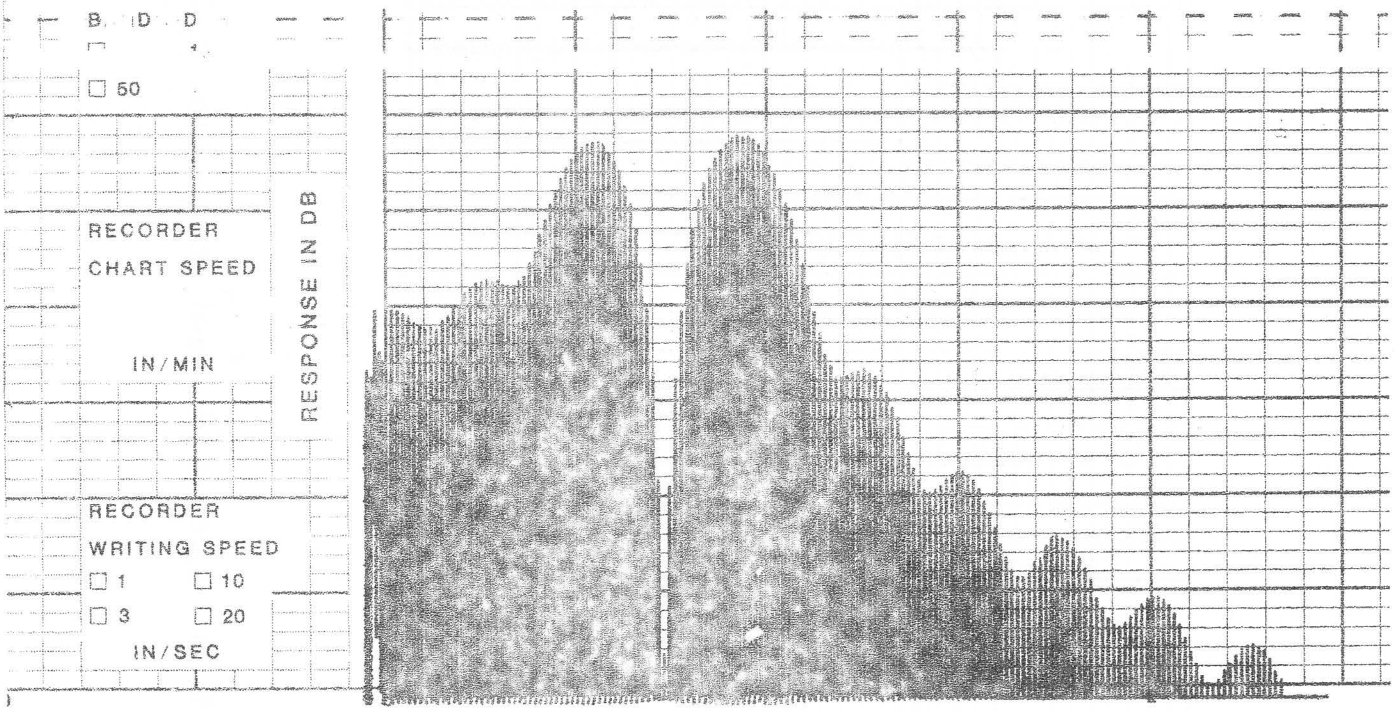


Fig.13 Spectrum of signal 7F, F_o=240Hz, F_1=580Hz, F_2=930Hz

Fig.13a Spectrum of signal 7F with time-insertion, F_o'=14.4Hz

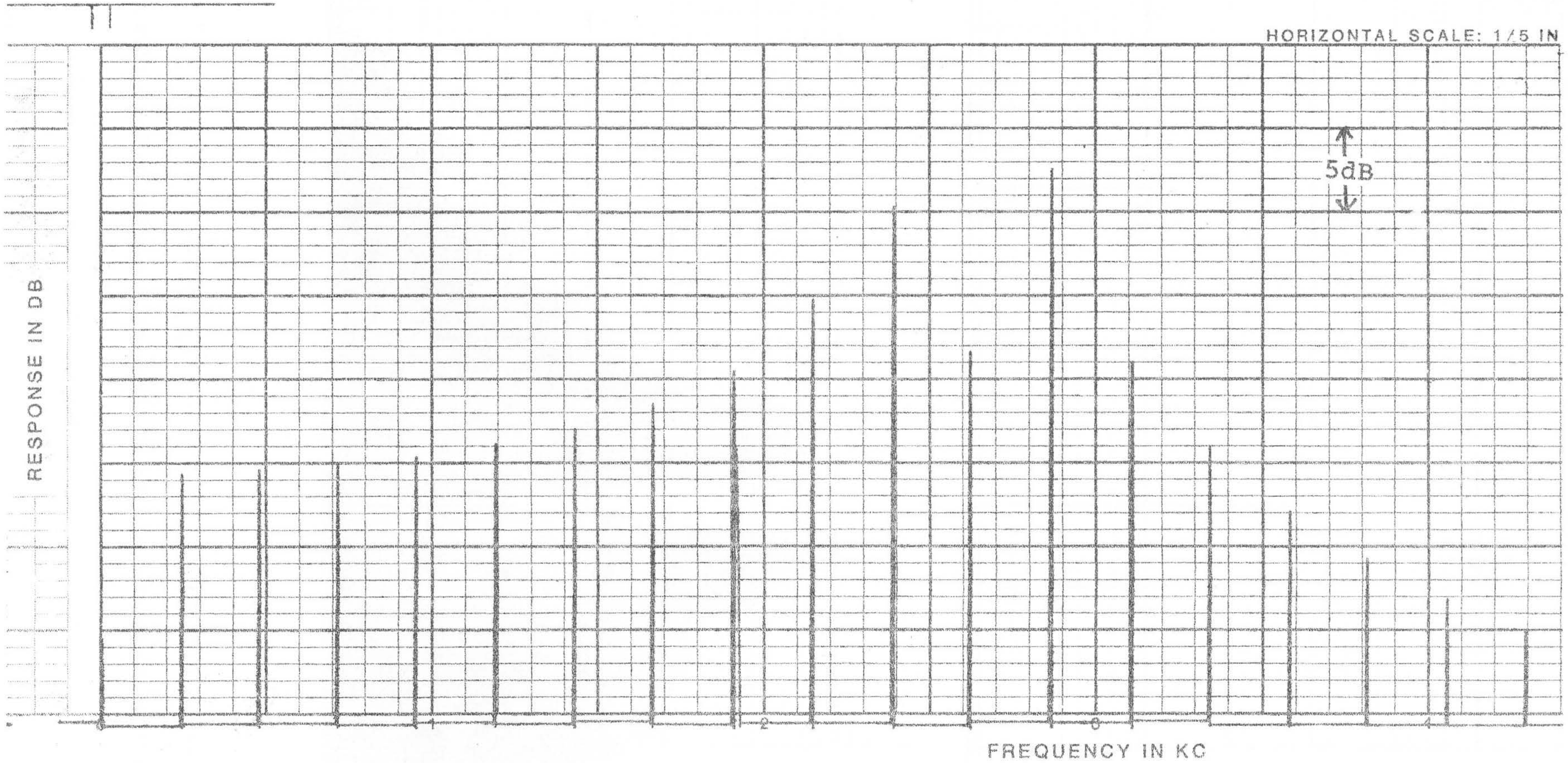


Fig.14 Spectrum of signal 6F, F_o=240Hz, F_1=2350Hz, F_2=2800Hz

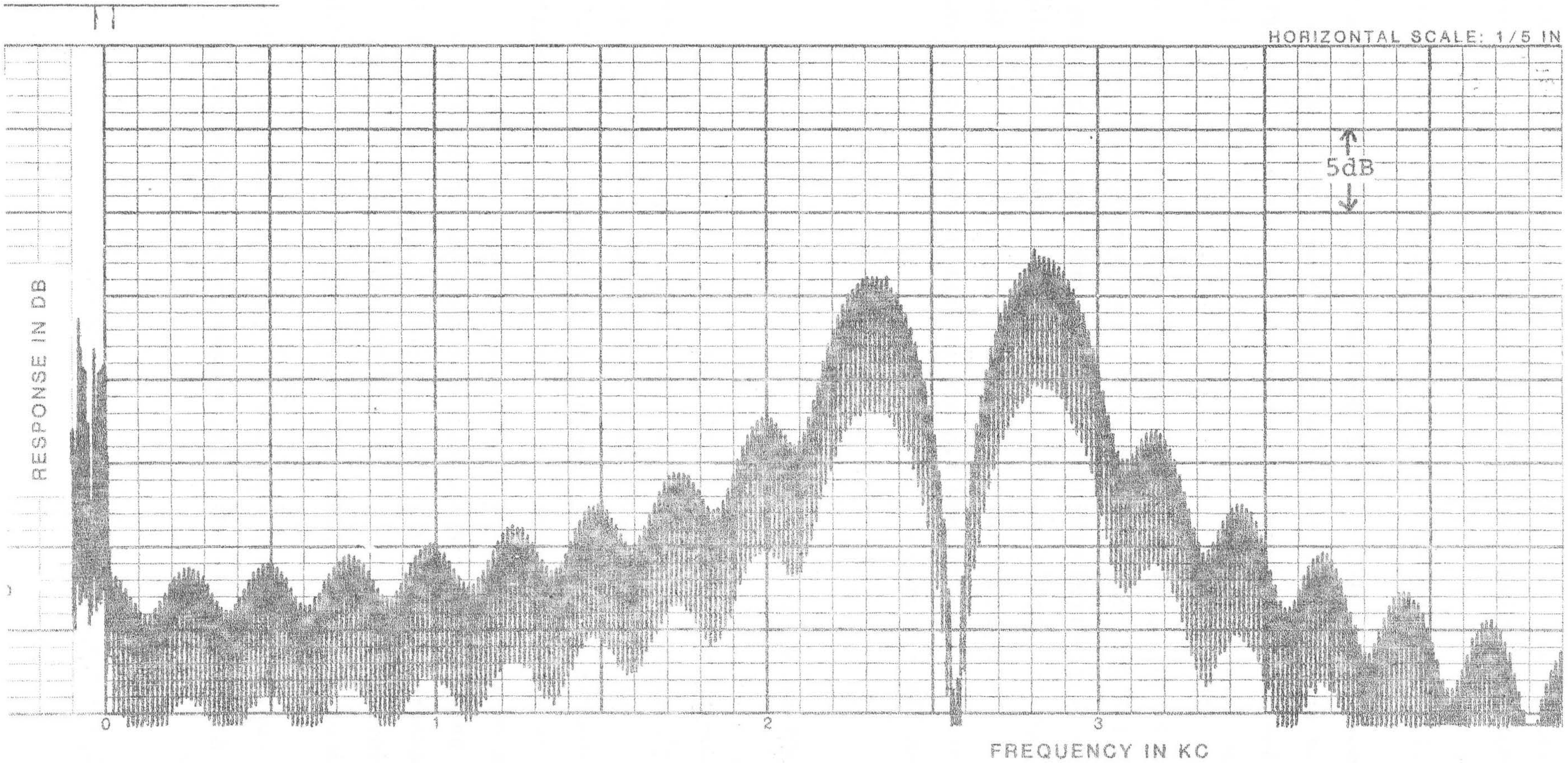


Fig.14a Spectrum of the signal from fig.14 with time-insertion, F_o'=14.4Hz

The F_0-ripple of spectra from signal segments with considerable final amplitude (figs 10a and 14a) could occur to such an extent that the readability of the formants would be rather poor, especially when more than two significant formants are present.

Furthermore, the limitations of the method for formant extracting from line spectra using the principle of a weighted average of maxima and their neighbouring components (Potter and Steinberg, 1950) are clearly displayed particularly in the figs 13 and 13a.

The high-frequency behaviour of spectra of the type as shown in the figs 11a and 12a is caused by the final amplitude step present in the corresponding signals. In practice, when gating a signal segment, amplitude steps can easily be avoided by taking care that the segment starts and ends at zero-crossings, so that the spectra of figs 11a and 12a are in fact academic ones (and with them the function $G_{m2}(\omega)$).

3. SECOND APPROXIMATION

The commonly used method for eliminating or attenuating the 'interval ripple' in spectra of time-limited signals is multiplication of the interval with some 'window function' prior to the transformation into the frequency domain. However, if this is applied to the short intervals of the F_0-periods, large signal portions of these are greatly attenuated, especially in the initial parts, where the signals have their major amplitudes. From chapter 2 we know that the 'ripple amplitude' is proportional to the initial amplitude of the modification function $g_m(t)$. The obvious way to attenuate the 'F_0-ripple' is decreasing this initial amplitude (which equals $e^{-\alpha T_0}$ times the initial amplitude of $g(t)$) by multiplying the signal with an exponential function:

$$(17) \quad g_h(t) = e^{-\alpha_a t} \qquad [0 \leq t]$$

Now, α could be thought of as the sum of the natural and artificial damping:

$$g(t) = e^{-(\alpha_n+\alpha_a)T_0} \sin \omega_1 t$$

The amplitude of the modification function alters to:

$$e^{-(\alpha_n+\alpha_a)T_o}|G_e(\omega)|$$

If α_a could be given an individual value for each segment to be measured, it would be possible to keep the factor

$e^{-(\alpha_n+\alpha_a)T_o}$ constant for all different segments.

If $e^{-(\alpha_n+\alpha_a)T_o} = k$, then the spectrum occurs between $(1+k)|G_e(\omega)|$ and $(1-k)|G_e(\omega)|$. A logarithmic amplitude scale causes a constant 'peak-to-peak ripple value':

$$(18) \qquad V_{rpp} = 20 \log \frac{1+k}{1-k} \text{ dB}$$

Choosing $k = 1/20$ yields the ripple to be within 1 dB.

$$(19) \qquad \text{Therefore } e^{-(\alpha_n+\alpha_a)T_o} = \frac{1}{20}$$

If for most vowels α_n is valued at 250 and for T_0 is written $1/F_0$, then from (19) the rule-of-thumb can be obtained:

$$(20) \qquad \alpha_a = 3F_0 - 250$$

Of course the accuracy of the displayed formant decreases with increasing α_a and formula (20) therefore offers a compromise between ripple smoothing and accuracy of the frequency measurement.
For a 'worse case' example, assume: $F_0 = 240$ Hz, then:

$$\alpha = \alpha_n+\alpha_a = 3F_0 = 720$$

Substitution of this into (7) together with the 'worse' value $f_1 = 300$ Hz as used before gives:

$$\omega_{max} = \sqrt{(2\pi.300)^2 - 720^2} = 1742$$

or: $f_{max} = 277.25$ Hz.

The error is about 7.5%, which is low compared to the familiar formant extraction errors of this type of signal. Besides, because the deviation from f_1 is known, a correction is possible if desired.

Of course the 3 dB bandwidth is considerably increased.

In this example it amounts to about 230 Hz ($B = \frac{\alpha}{\pi}$).

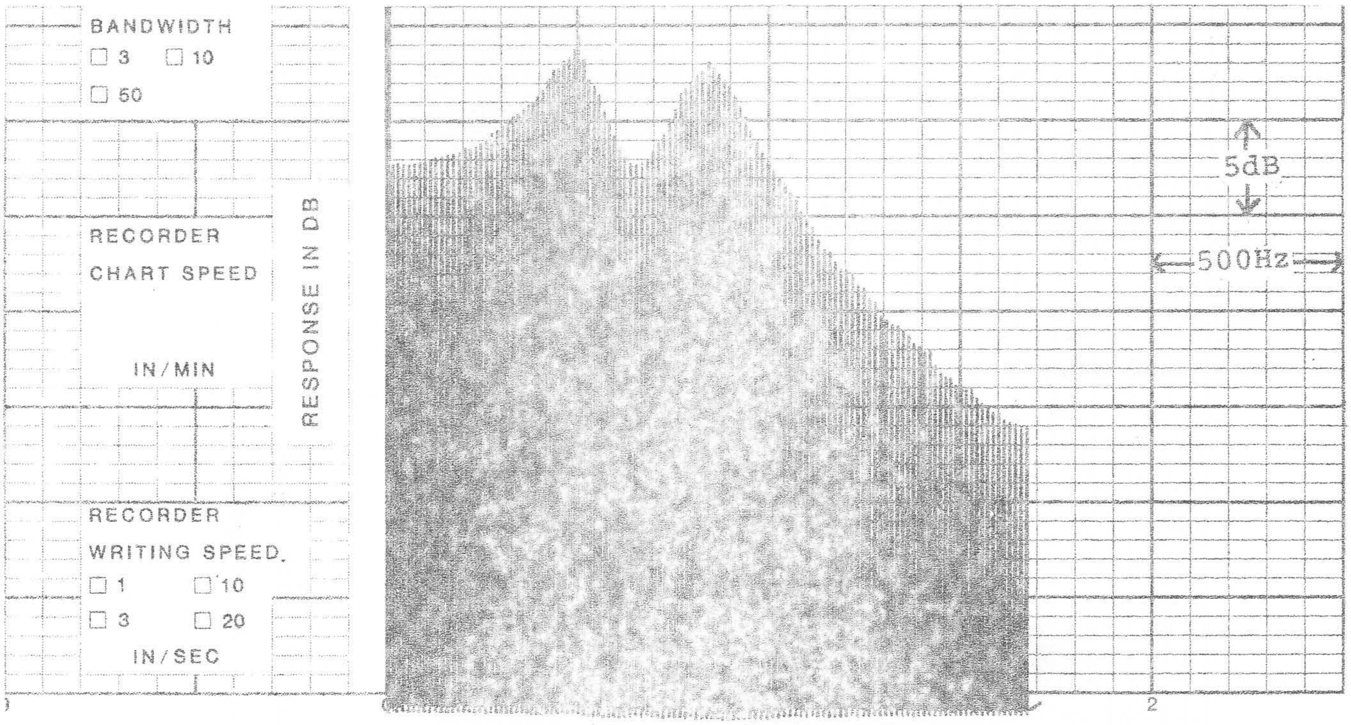
Evidently, to check this method in practice, the same test set-up as described above is applicable, if the damping of the sinusoids is adjusted according to:

$$\alpha = 3F_0 = \frac{3}{T_o}$$

Some spectra obtained from measurements as carried out in this manner are displayed in the figs 15, 16 and 17 , whereas the spectra of the corresponding signals with 'normal' values for α (i.e. without additional damping) are depicted in the figs 12a, 13a and 14a. As a result of applying this 'exponential window', the readability of the 'formants' is improved considerably.

Using the convolution theorem it can be stated that multiplication of the time functions results in the convolution of their individual spectra. Though this is not exactly true for amplitude spectra, the error could be significant only in the vicinity of zero frequency (Randall, 1977).

The process of a measuring filter scanning the entire frequency range can be considered as the practical analogy of the process of convolving the input spectrum with the measuring filter characteristic. Then if the filter pass band curve could be made equal to the spectrum of the exponential window, this method would give results very similar to the exponential window method.



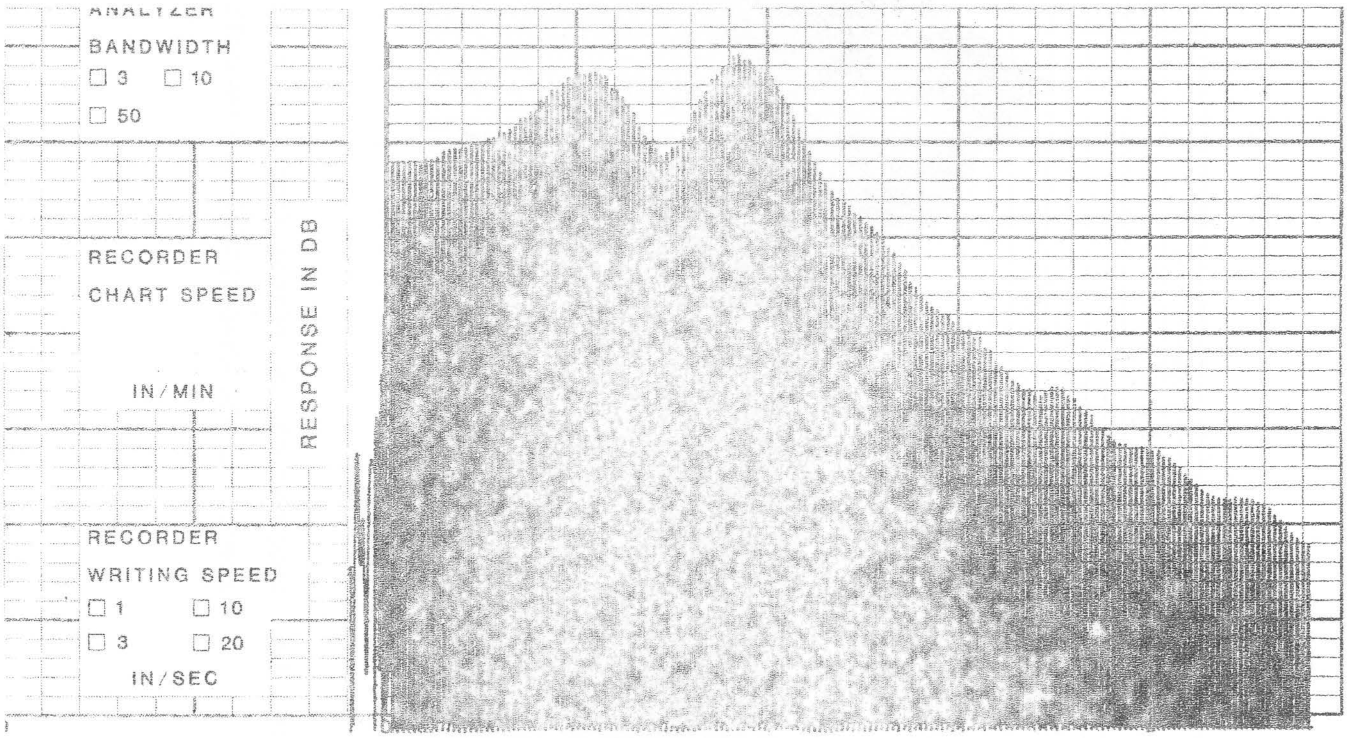


Fig.15 Spectrum of the signal from fig.12 (signal 2M) with time-insertion and exponential window multiplication, $\alpha_a \approx 200$

Fig.16 Spectrum of the signal from fig.13 (signal 7F) with time-insertion and exponential window multiplication, $\alpha_a \approx 350$

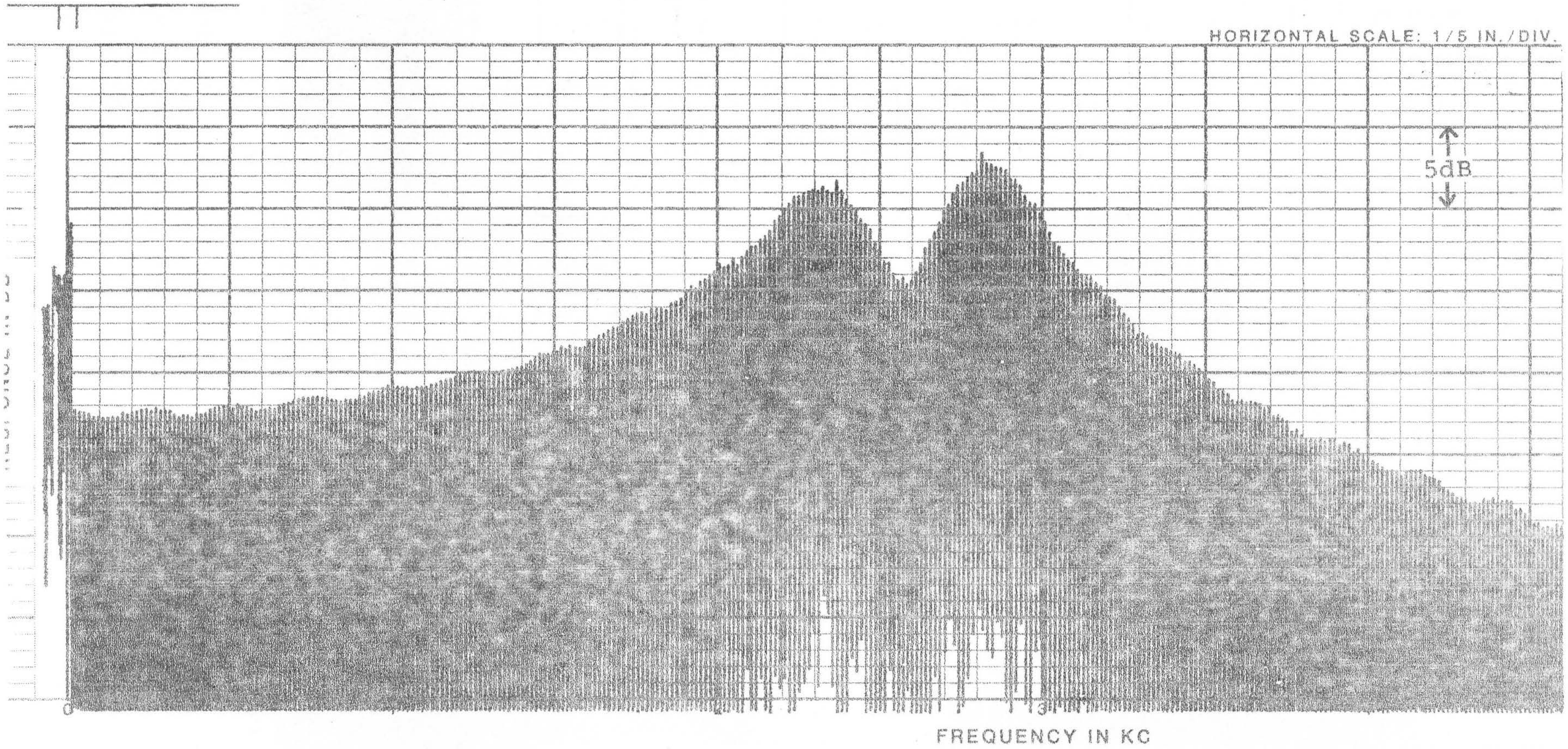


Fig.17 Spectrum of the signal from fig.14 (signal 6F) with time-insertion and exponential window multiplication, $\alpha_a \approx 400$

The amplitude spectrum of the function $e^{-\alpha_a t}$ $[0 \leqslant t]$ is equal to:

$$\frac{1}{\sqrt{\alpha^2+\omega^2}}$$

A simple second order resonance circuit with a bandwidth of $\frac{\alpha_a}{\pi}$ Hz matches this function apart from the asymmetry when plotted on a linear frequency scale.
Although this method seems quite realisable, we thought it easier at the time to apply the exponential window method since in that case we could use the narrow band spectrum analyzer again whereas we otherwise had to construct a spectrum analyzer with a continuous adjustable bandwidth.

The time insertion process and the exponential window method were also applied to natural vowel signals. For this purpose the following measuring set-up was made (see fig.18):
From the vowel signal to be measured one period of the fundamental frequency was isolated with the aid of the 'Precision Gate' (Wempe, 1976) and stored in a digital signal memory (1000 x 8 bit). The memory length was made 20 msec which is sufficient to store any F_0-period of vowel signals. Thus after the signal period zeroes were stored.
The contents of the memory were repetitively scanned at the same rate as during the recording, controlled by an external timing circuit.
The binary equivalents of the sequential samples were converted into their original values by a multiplying digital-to-analog convertor. The DAC's reference input was controlled by a device which produced the function $e^{-\alpha_a t}$. Its α_a was made adjustable. The timing circuit caused the function $e^{-\alpha_a t}$ to start from its origin $(t = 0)$ at the beginning of each memory scanning cycle. Besides, this circuit also activated an analog switch in such a way that the analog output of the DAC was connected to the spectrum analyzer only during one out of three scanning cycles, which caused an additional time insertion.

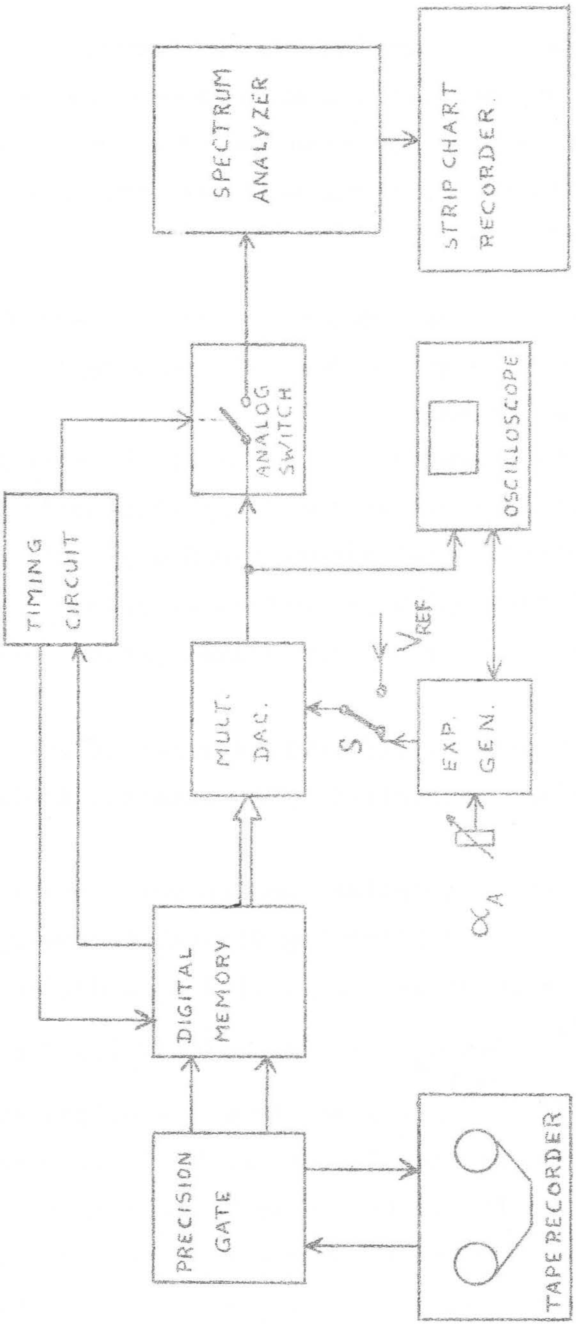
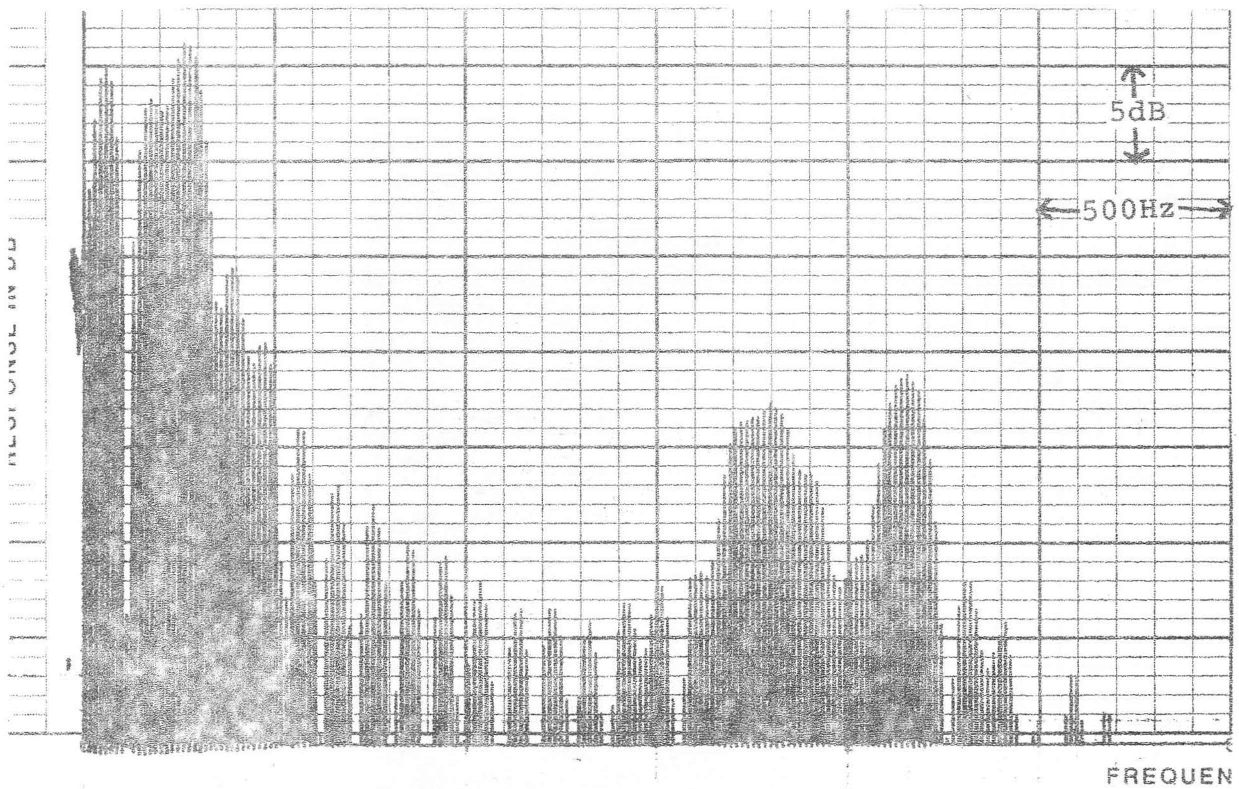

PRECISION GATE
TAPE RECORDER
DIGITAL MEMORY
TIMING CIRCUIT
MULT. DAC.
EXP. GEN.
S
V_{REF}
α_A
ANALOG SWITCH
OSCILLOSCOPE
SPECTRUM ANALYZER
STRIP CHART RECORDER.

Fig. 18



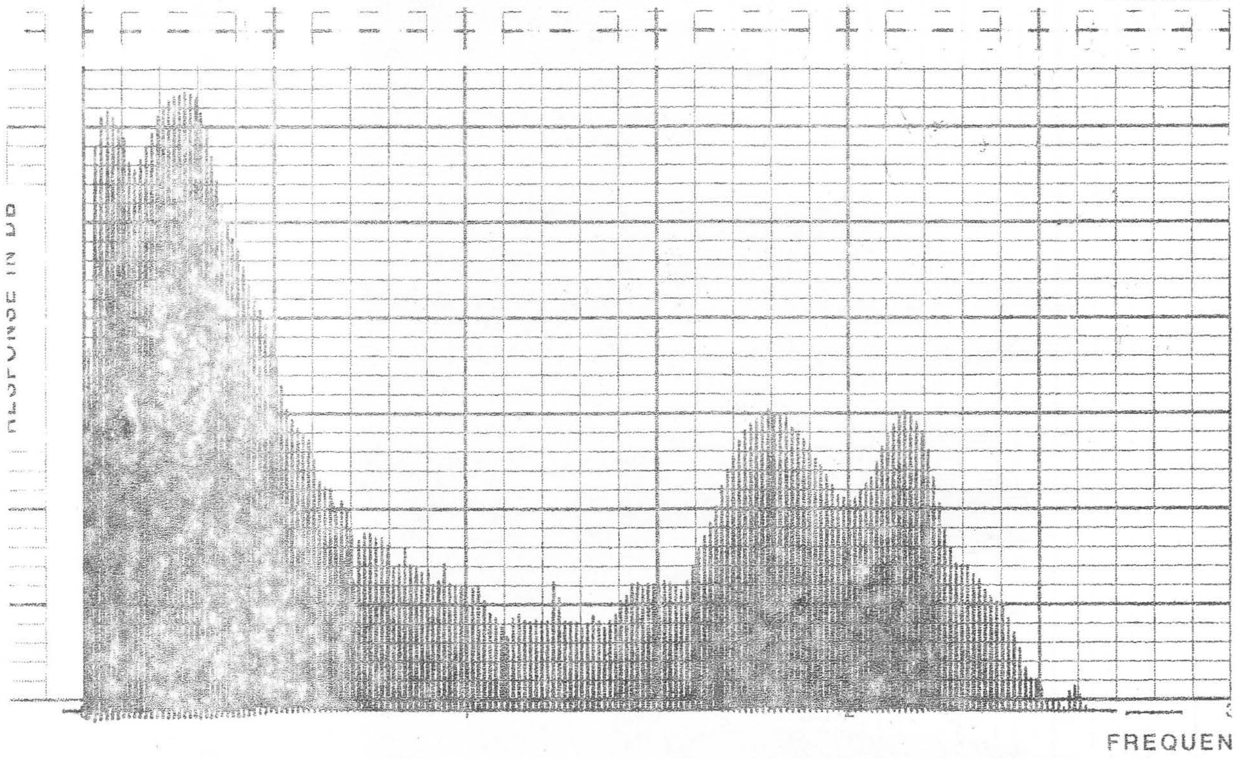


Fig.19 Spectrum of male [y] with time-insertion, F_o=95Hz, F'_o=14.4Hz

Fig.19a Spectrum of the same vowel with time-insertion and exponential window multiplication, $\alpha_a \approx 50$

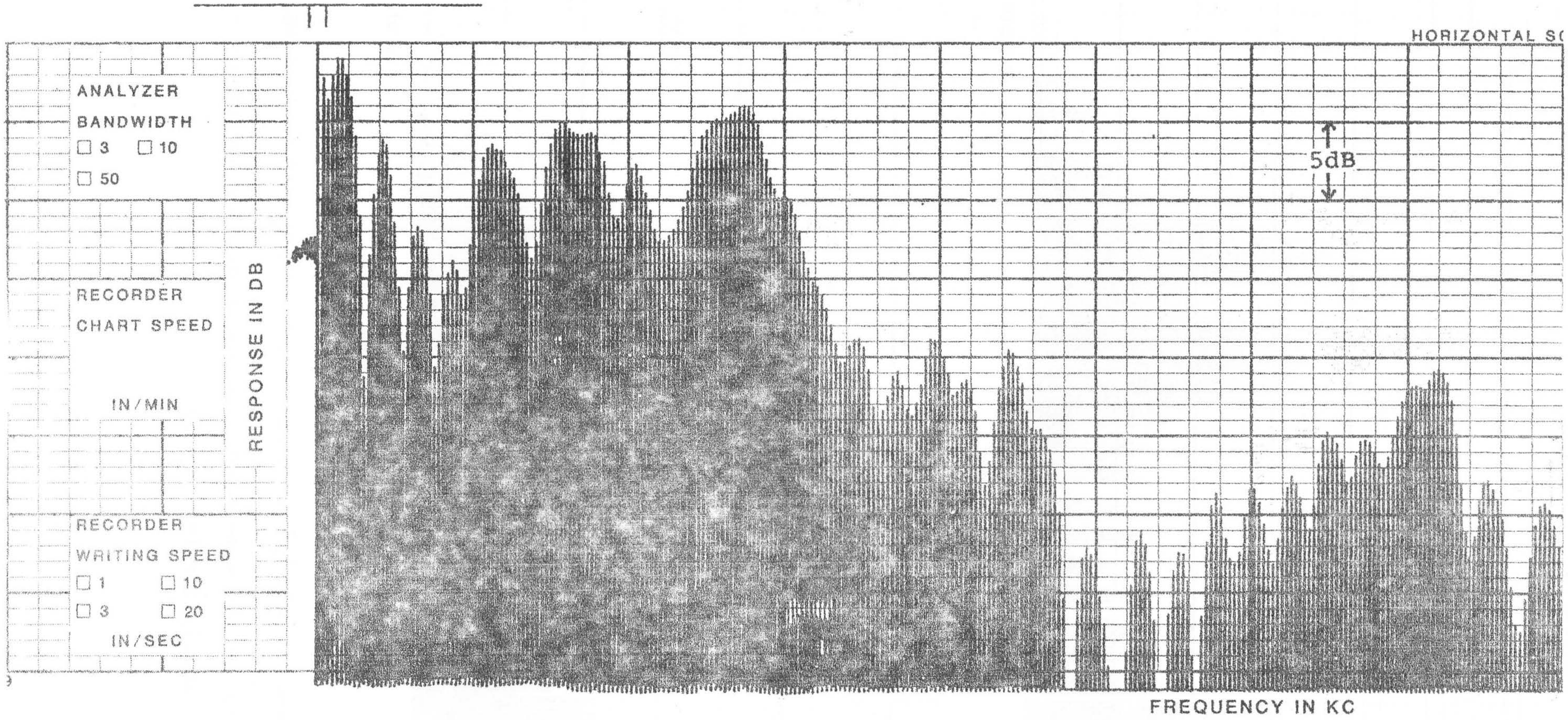


Fig.20 Spectrum of male [a] with time-insertion, F_o=120Hz, F'_o=14.4Hz

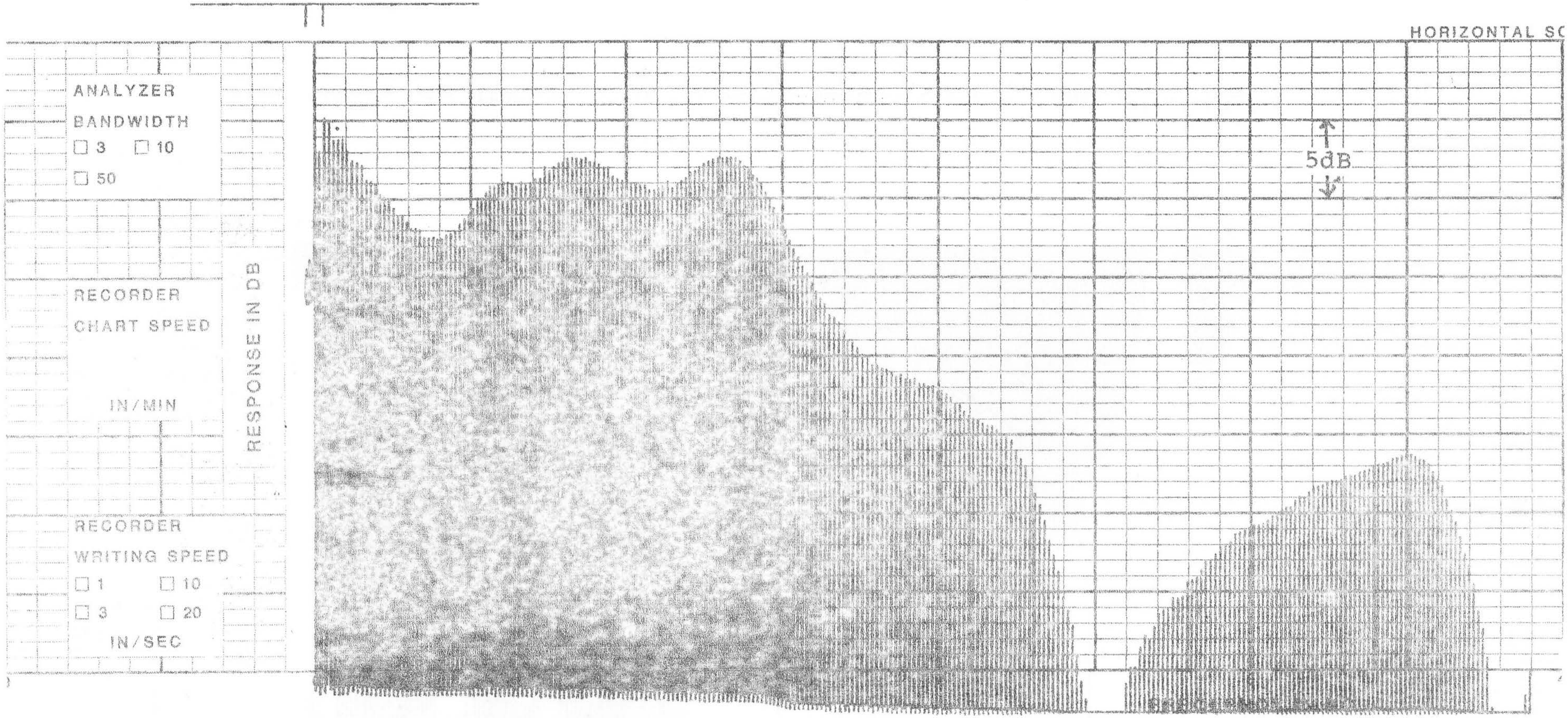


Fig.20a Spectrum of the male [a] from fig.20 with time-insertion and exponential window multiplication, $\alpha_a \approx 100$

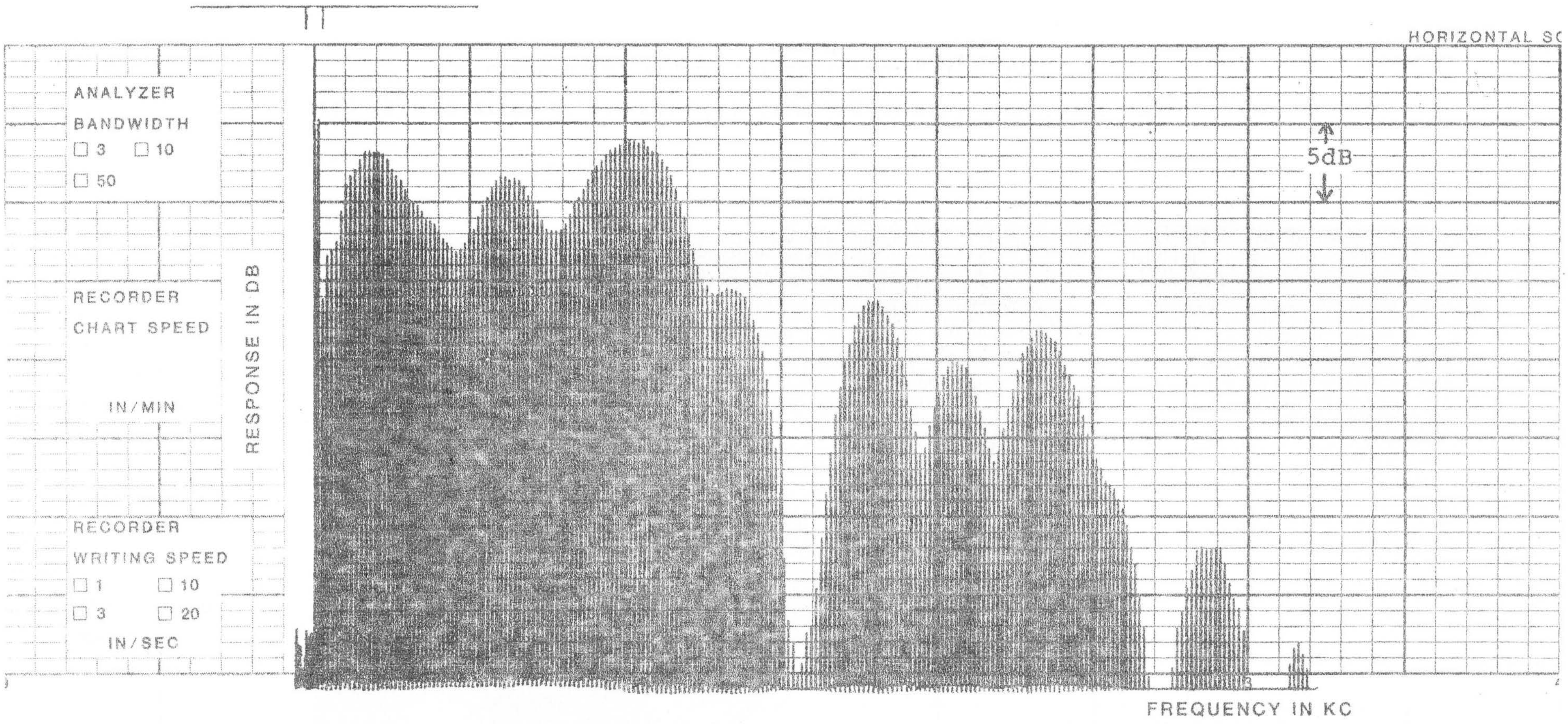


Fig.21 Spectrum of female [a] with time-insertion, F_o=270Hz, F_o'=14.4Hz

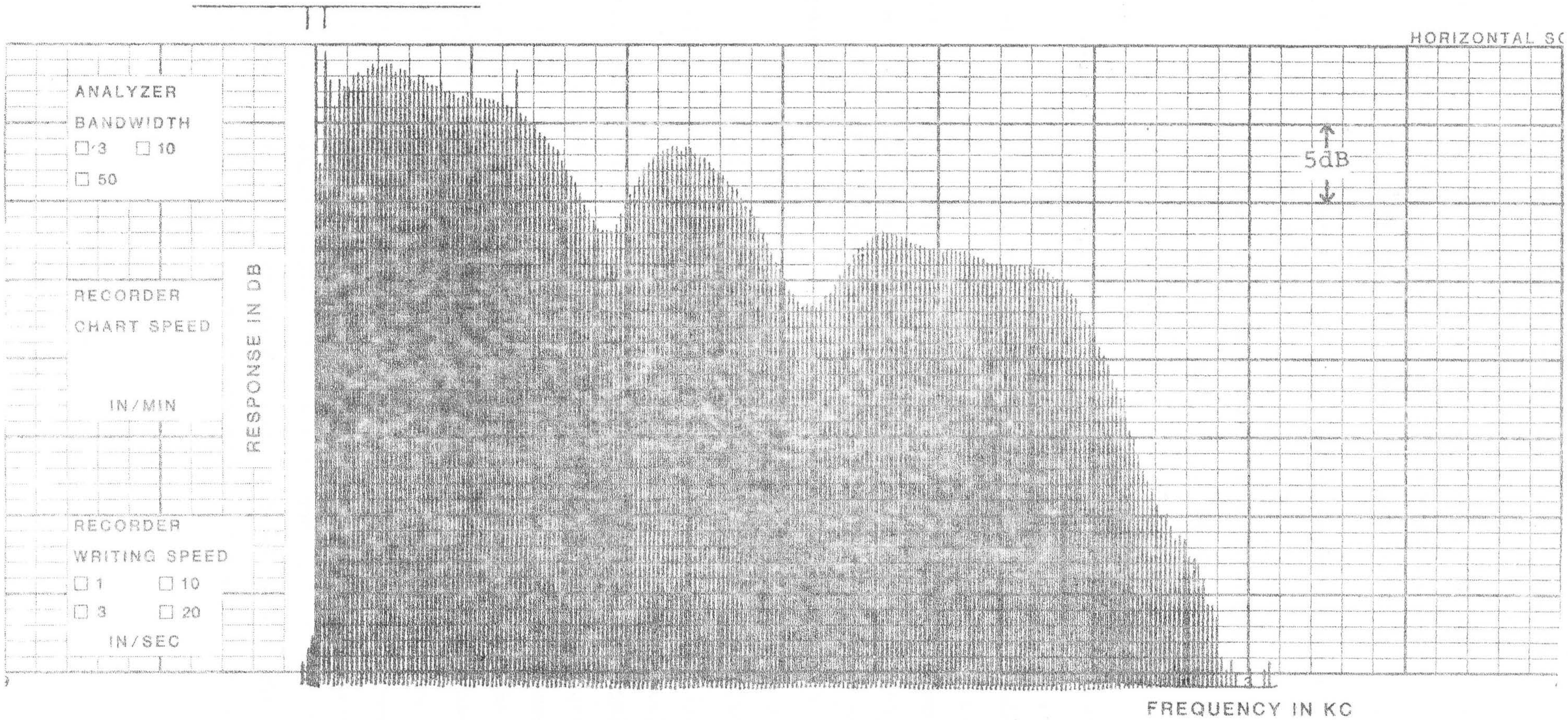


Fig.21a Spectrum of the female [a] from fig.21 with time-insertion and exponential window multiplication, $\alpha_a \approx 400$

(The resulting fundamental period T'_o was slightly more than three scanning cycles owing to a time delay between consecutive scanning cycles. The final F'_o therefore was equal to 14.7 Hz).
The effect of the adjustable factor α_a on the time function of the speech interval was made visible on an oscilloscope. Cancelling of the multiplication function was possible with the switch S.

From the signals we measured in this manner a few are displayed in the figs 19 through 21 . The figs 19 , 20 and 21 were made without the use of the exponential window, whereas the effect of the exponential window is shown in the figs 19a,20a and21a respectively.

4. THIRD APPROXIMATION

The considerable amount of time necessary for plotting a complete spectrum and the optimum adjustment of α_a sometimes being difficult were the major disadvantages of the previous method. Althou both problems can be practically eliminated by respectively adapting high speed analysis and determining α_a automatically from the length of the segment, at this stage an experiment of a different method for frequency analysis was started.
The following may explain the principle of this method.
Consider the practical case of a measuring filter responding to a time-limited signal portion.
From a time domain point of view the influence of the signal length on the response of the filter only starts after the moment the signal segment ceases. In fact the sudden change at the end of the segment contributes to the filter response only from that moment on. The 'formant masking' side lobes in the continuous spectrum (or in its practical approximation) are therefore solely caused by this sudden change at the segment end. Now the idea was to modify the time inserting method from chapter 2 by 'enabling' the measuring filter only during the occurrence of the signal segment within each period of the filter input signal. 'Disabling' of the filter was intended to be achieved by withdrawing the energy from it so that

the filter response should start from zero at the beginning of each signal period.
Using narrow band filtering implies that in this way the filter output voltage is kept considerably below its steady state value. The spectral information however, can be extracted from the various peak voltages of the filter output at different center frequencies.

Realization.

For the practical implementation of this idea the construction of a special purpose spectrum analyzer was necessary owing to the unusual requirements. The measuring procedure could be roughly described as follows:
From a signal, an interval with 20 msec length is recorded into the signal memory in such a way that the fundamental period or other segment to be analysed forms part of the memory contents. During repetitive replay of the memory contents, the wanted segment can be selected by means of controls which cause the passing of the signal via an electronic switch only during the occurrence of the segment involved. (An oscilloscope serves as a monitor necessary for adjustment of the controls).
The signal passing through the electronic switch is fed to the input of the measuring filter of the analyzer. After each replay cycle of the memory, the central frequency of the measuring filter is increased one step. The control voltage for the electronic switch serves also as a control signal for the 'enabling' and the 'disabling' of the measuring filter. (Disabling occurs by short-circuiting the voltages of the 'memory components' within the filter via electronic switches).
The filter output is connected to a full-wave peak rectifier which is reset before each new replay cycle is started. The output voltage from the peak rectifier is fed to a strip chart recorder through a logarithmic convertor. After a number of replay cycles of the memory, the filter has swept over the frequency range of interest and the plotting is completed.

Bandwidth.

In order to attain an output voltage of some importance from the interrupted filter, its response time should not be too long compared with the segment duration (T_o), i.e. narrow band analysis cannot be applied. However, this doesn't imply a loss of resolution: the highest possible resolution is limited by the inherent 'pulse bandwidth' of $1/T_o$ Hz.
Obviously during one spectrum plotting a constant bandwidth is required. Although the choice of the bandwidth isn't critical here, it should preferably be made proportional to $1/T_o$ in order to avoid the spectral amplitudes being dependent on the segment length.
As the bandwidth is constant, a linear frequency scale is justified. (The possible reasons for applying a logarithmic frequency sweep are completely gone when analysing short signal segments).
Tuning could be done by scanning the memory contents at different speeds and leaving the filter unaltered. Although this could be realized very easily and cancels the problem of making a tunable filter, it implies a measurement with a resulting constant *percentage* bandwidth instead of a constant bandwidth. The problem of stepwise correcting the bandwidth being comparable with the problem of making a tunable filter, we decided to solve the latter, particularly on account of the taylor-made properties of the gyrator band pass principle.
This may be explained by the following description of the gyrator band pass filter principle:

A gyrator (Tellegen, 1948) basically consists of two mutually coupled 'transconductance amplifiers' (see fig. 22). The output current of an ideal amplifier of this sort is proportional to the input voltage, whereas the input impedance is infinite.
One of the amplifiers is an inverting type which thus causes a negative feedback. Each amplifier output forms a terminal (gate) with regard to the common connection.
Let both amplifiers have equal transconductances (g). If one gate is loaded with an impedance Z_1, the effected impedance (Z_t)

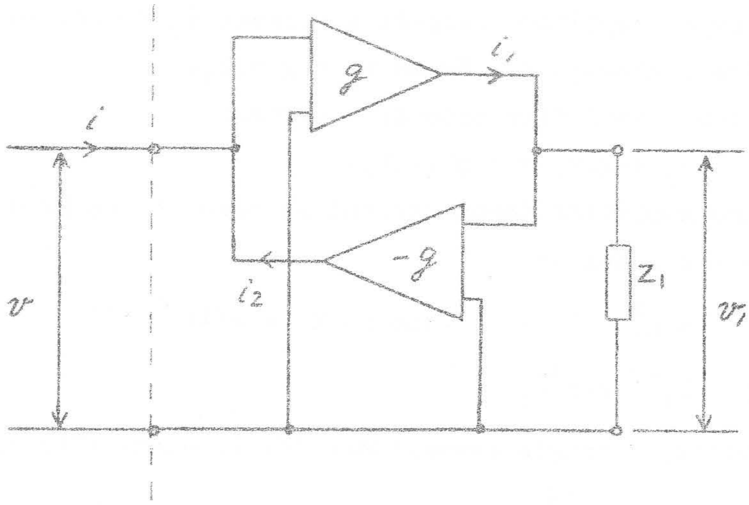


Fig. 22

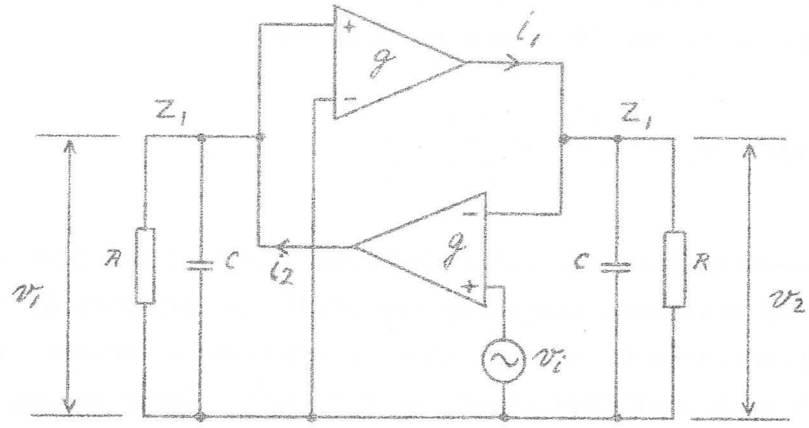


Fig. 23

at the other gate can be computated as follows:

A current i flowing through the effected impedance Z_t causes a voltage v occurring across it: $Z_t = v/i$.

The upper amplifier outputs a current $i_1 = g.v$, which is flowin entirely through Z_1. Hence $v_1 = g.v.Z_1$.

The lower amplifier outputs a current

$$i_2 = -g.v_1 = -g^2.v.Z_1.$$

As the amplifier input current is zero it can be stated that $i_2 = - i$. Hence:

$$-g^2.v.Z_1 = - i \text{ and as } Z_t = v/i:$$

$$Z_t = 1/g^2Z_1 \tag{21}$$

If for Z_1 a single capacitance (C) is used, (21) changes to:

$$Z_t = j\omega\frac{C}{g^2}$$

which behaves like a self-inductance:

$$L = \frac{C}{g^2} \tag{22}$$

It will be clear that a simple resonant circuit can be made by connecting the left gate with a capacitance as well.

Its resonance frequency results from:

$$\omega_r = \sqrt{\frac{1}{LC}} = \sqrt{\frac{g^2}{C^2}} = \frac{g}{C} \tag{23}$$

assuming that the two capacitances have equal values.

This resonance frequency can easily be adjusted by varying g, which could be accomplished by means of a variable attenuator at each amplifier input. In this way, linear tuning is very simple to realize.

If Z_1 consists of a capacitance and a resistance (R) in parallel, Z_1 changes to:

$$Z_1 = \frac{R}{1+j\omega CR}$$

and: $Z_t = \frac{1}{g^2 Z_1} = \frac{1+j\omega CR}{g^2 R}$.

Now Z_t can be considered as consisting of two components connected in series, namely: a resistance $r = \frac{1}{g^2 R}$ and a self-inductance according to (22).

The application of this means that the filter bandwidth can be determined by rating the resistors connected across the capacitances.

A schematic set-up of this is shown in fig. 23.

The following equations are valid:

$$i_1 = g.v_1$$

$$i_2 = g(v_i - v_2)$$

$$v_1 = i_2.Z_1$$

$$v_2 = i_1.Z_1$$

Some substitutions result into:

$$(24) \quad \frac{v_2}{v_i} = \frac{g^2 Z_1^2}{1+g^2 Z_1^2}$$

and $$\frac{v_1}{v_i} = \frac{g\ Z_1}{1+g^2 Z_1^2}$$

where $Z_1 = \frac{R}{1+j\omega CR}$

Writing $\omega_r = \frac{g}{C}$ and α for $\frac{1}{RC}$ changes this to:

$$(25) \quad Z_1 = \frac{1}{C(\alpha + j\omega)}$$

Substituting (25) into (24) yields:

$$\frac{v_2}{v_i} = \frac{\omega_r^2}{\omega_r^2+(\alpha+j\omega)^2}$$

or, slightly modified:

$$(26) \qquad \frac{v_2}{v_i} = \frac{g}{C} \frac{\omega_r}{\omega_r^2+(\alpha+j\omega)^2}$$

In a similar way can be found:

$$(27) \qquad \frac{v_1}{v_i} = \frac{g}{C} \frac{\alpha+j\omega}{\omega_r^2+(\alpha+j\omega)^2}$$

Apart from the term $\frac{g}{C}$ these results are equal to the complex spectra of damped sinusoids with initial phases of 0 and $\frac{\pi}{2}$ radians respectively, as calculated in chapter 2.
If the input voltage of the gyrator filter is divided by g, its amplitude response is proportional with the amplitude spectra of the damped sinusoids. Therefore what is found in chapter 2 concerning the maxima and bandwidth is completely valid for the filter response.
Moreover, if this type of band pass filter is used, the 'inaccuracy' of the spectral peak locations as calculated in that chapter is decreased and even completely corrected when the filter bandwidth equals 2α.
If the signal gating is carried out according to the final remark in chapter 2, optimum correction occurs when v_2 is considered as the filter output voltage.

With reference to the functional diagram (fig.24) the working principle of the segment spectrograph can now be described in some detail.
After the recording of a part of a signal into the signal memory ('transient recorder') as already described, the signal is replayed repetitively at a constant rate. The analog output (Y) of the

Segment spectrograph system
with interrupted filtering
Functional diagram

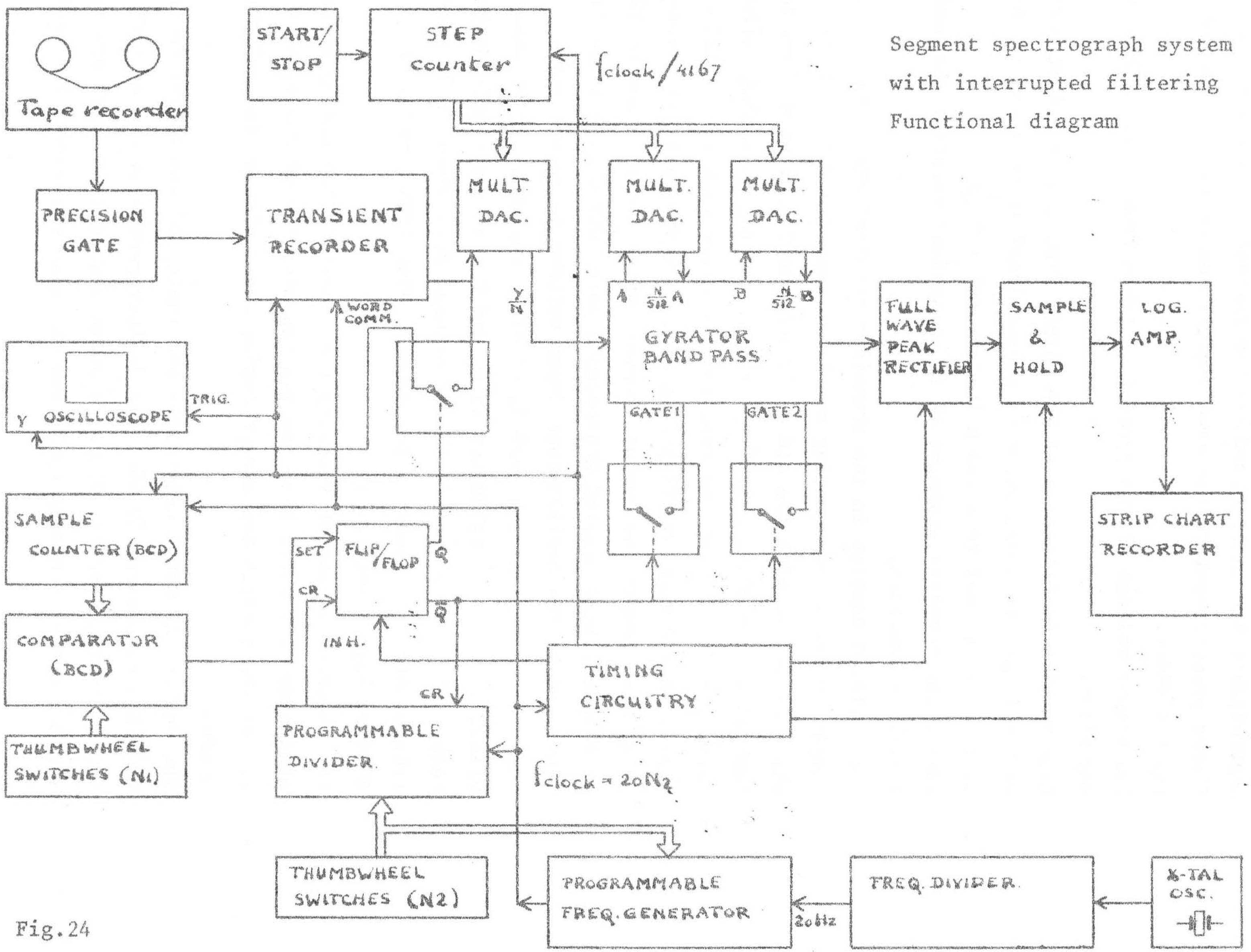


Fig.24

transient recorder is attenuated by a digitally controllable divider (using a multiplying digital-to-analog convertor) and then led to the input of the gyrator band pass filter.

Two multiplying digital-to-analog convertors serve as digitally controllable attenuators, making the central frequency of the filter proportional to the same digital number (A) as used for the input divider.

The output voltage of the gyrator is taken from v_2 via a buffer amplifier.

This results in filtering of the output from the transient recorder with a filter response which is proportional to equation (6) where ω_1 is proportional to A and α is constant ($= \frac{1}{RC}$).

During one complete scanning of the transient recorder the number A remains constant.

The replayed samples in the transient recorder are counted by a sample counter. The counting result is compared with a number (N_1) which is preset with the aid of 'thumb wheel' switches. When both numbers are equal, a flip-flop is set, causing an analog switch to connect the output of the transient recorder with the oscilloscope input. In addition the flip-flop enables a programmable counter which also counts the output samples. After a preset number (N_2) of samples, this counter generates a pulse which clears the flip-flop causing the oscilloscope input voltage to become zero.

The desired fundamental period, or other segment, can be selected visually therefore by adjusting N_1 and N_2. As any sample number below 1000 can be preset, the adjustment resolution is 20 µsec (the complete contents of the memory being 1000 samples of a 20 msec signal interval).

The inverted output of the flip-flop controls two analog switches simultaneously which keep the two capacitances of the gyrator filter short-circuited except during the replay of the selected segment.

The number (A) which determines the tuning of the filter is formed by a binary counter (9 bit) and is incremented by one at each pulse from a timing circuit. The same pulse serves to restart the transient recorder scanning cycle and the sample counter. The peak voltage of the gyrator filter output is obtained by a full-wave peak rectifier which is reset before the next tuning step. A sample-and-hold circuit changes the voltage directly from one measured value into the next one, thus preventing unnecessary large excursions of the writing pen of the chart recorder, which is connected via a logarithmic amplifier. The peak rectifier reset command and the sample-and-hold control are generated by the timing circuit.

This circuit ensures a time lapse between consecutive tuning steps which is sufficient for plotting the spectrum accurately in spite of the relatively low writing speed of the chart recorder.
The timing circuit is driven by the same clock frequency as the one which controls the output rate of the transient recorder. During the time lapse, the flip-flop is inhibited in order to prevent the gyrator filter from responding to improper input signals.
The replay velocity of the transient recorder is made proportional with the length of the selected segment by making the output clock frequency equal to the product of a reference frequency and the number N_2 (achieved by using the 'phase-locked-loop' principle).
The reason of this may be explained by the following: Reducing the replay velocity by a factor k causes a decrease of the signal frequency components by the same factor. As the tuning steps and the bandwidth of the gyrator filter remain the same, the result is equivalent to filtering with increased bandwidth and tuning steps both by the factor k. Hence the filter response time and frequency resolution is matched with the selected segment length. The peak voltages from the filter output therefore are independent of the segment length.
Because the timing circuit is slowed down at the same time, the tuning time is raised by the same factor. The result is that the sweep speed remains constant. As the chart speed of the recorder remains unaltered as well, the frequency scale of the obtained spectra is independent of the segment length.
Furthermore the frequency range increases with the same factor in consequence.

In order to avoid any problems caused by spurious capacitances, the frequency sweep of the gyrator filter has been shifted to a lower area together with the output clock frequency. Owing to the limited writing speed of the chart recorder the time insertion between consecutive tuning steps is large enough to allow for replaying at slower rates without the consequence of increasing the analysis time.

The actual value of k is chosen to be 5 when a 10 msec signal segment is selected (i.e. $N_2 = 500$). In that case the output clock frequency equals 10 kHz (which accounts for scanning the complete memory in 100 msec). Therefore the reference frequency equals 10 kHz/500 = 20 Hz. The resolution bandwidth and tuning step width at a 10 msec segment were made 50 Hz and $16^2/_3$ Hz respectively (the number of tuning steps within one bandwidth was chosen to be 3). That causes a 10 Hz bandwidth and a $3^1/_3$ Hz tuning step width for the gyrator filter as in this case k = 5. The chart speed being 2 mm/sec and a frequency scale of 200 Hz/cm cause the required tuning speed to be 40 Hz/sec. If the tuning step width is $16^2/_3$ Hz, the tuning speed requires $40/16^2/_3 = 2.4$ steps/sec.
Thus the output clock frequency (10 kHz) must be divided by 4167 to obtain the required step frequency.

Computation of some properties of the responses of the interrupted filter system to the simplified artificial vowel sounds as used before didn't seem a very simple matter and therefore we decided to check whether or not this system would at all function satisfactorily in practice. At a later stage the theoretical behaviour could be worked out.
The figs 25 through 29 show some outputs obtained from this interrupted filter system at different input test signals.
The outputs from a few natural vowel periods are represented by the figs 30 through 32.
It can be concluded that the 'side lobes' have completely disappeared and the formants are displayed conveniently.

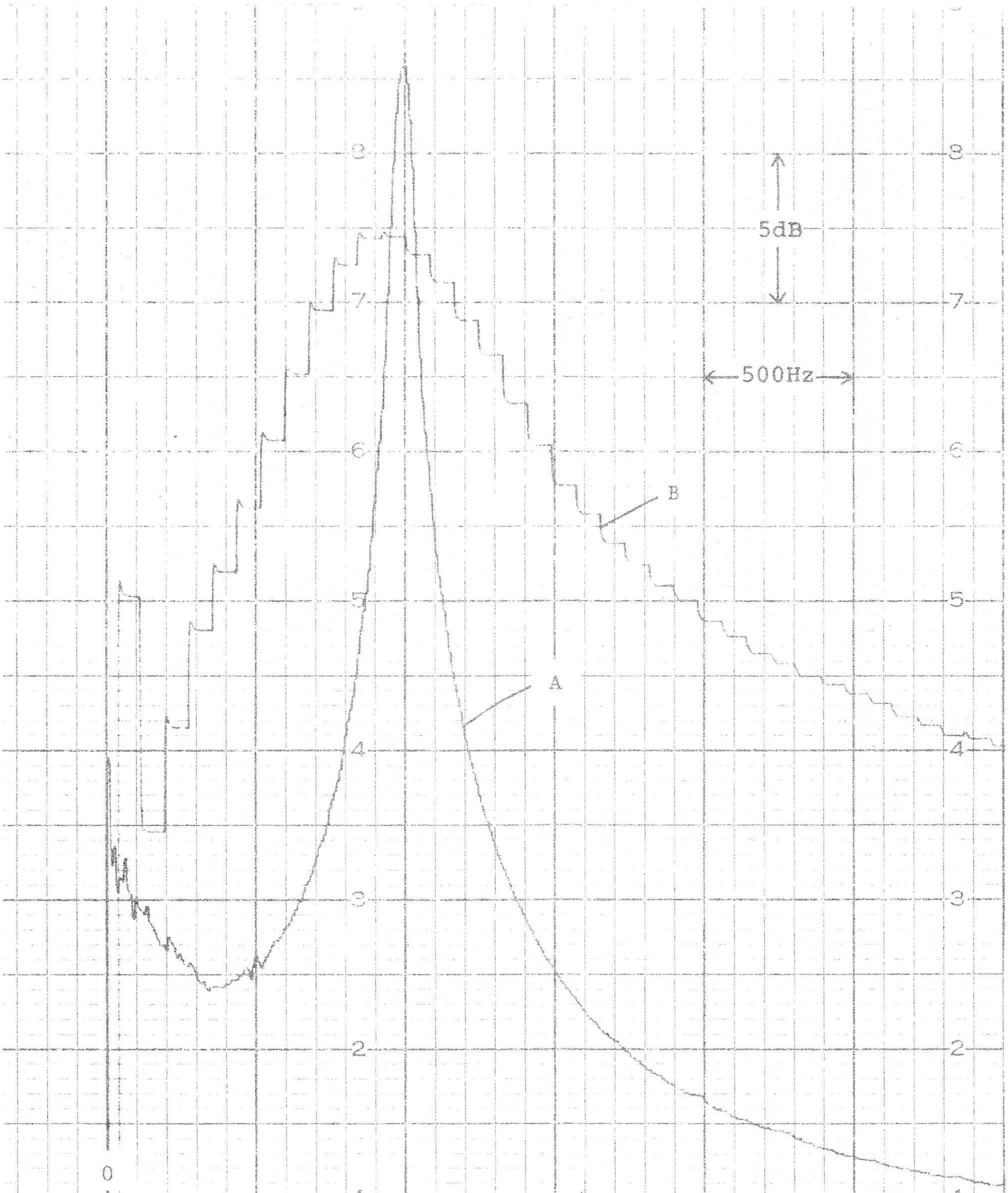


Fig.25

A: Spectrum of 20 periods of 1000Hz sinusoid (T_o=20msec)

B: Spectrum of 2 periods of 1000Hz sinusoid (T_o=2msec)

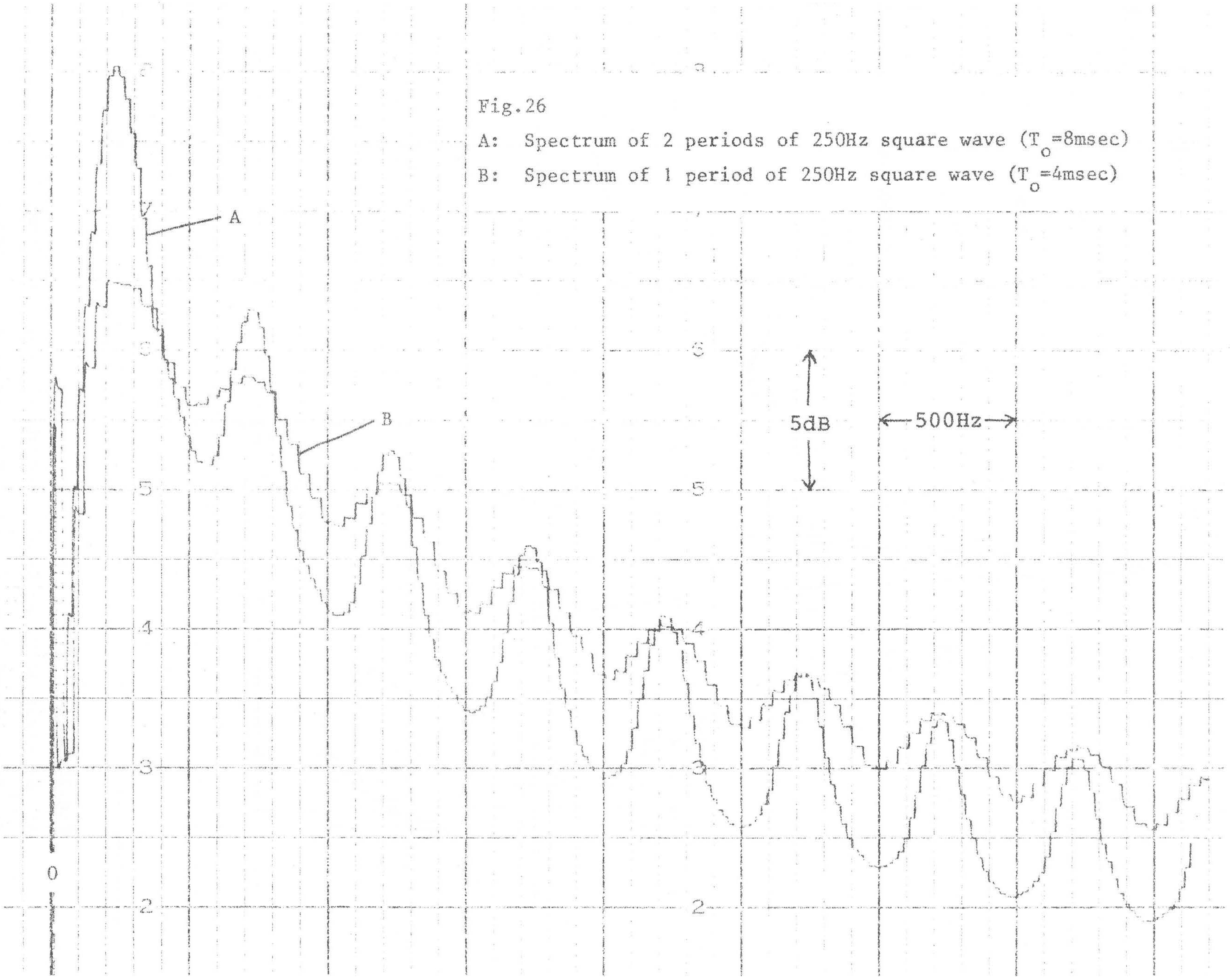


Fig.26

A: Spectrum of 2 periods of 250Hz square wave (T_o=8msec)

B: Spectrum of 1 period of 250Hz square wave (T_o=4msec)

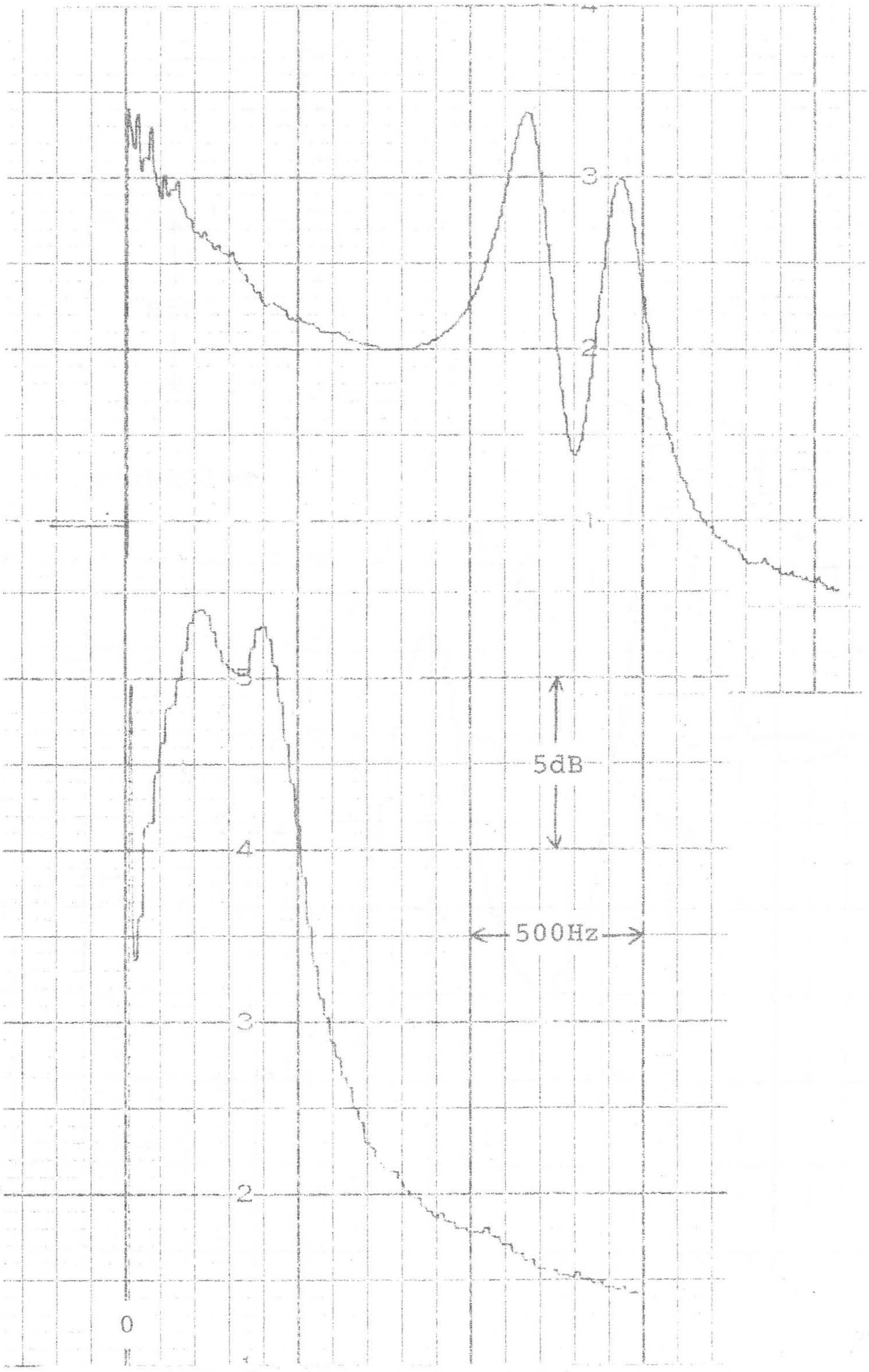


Fig.27 Spectrum of the sum of 2 damped sinusoids, F_1=1200Hz, F_2=1450Hz, T_o=17msec

Fig.28 Spectrum of the sum of 2 damped sinusoids, F_1=250Hz, F_2=400Hz, T_o=10msec

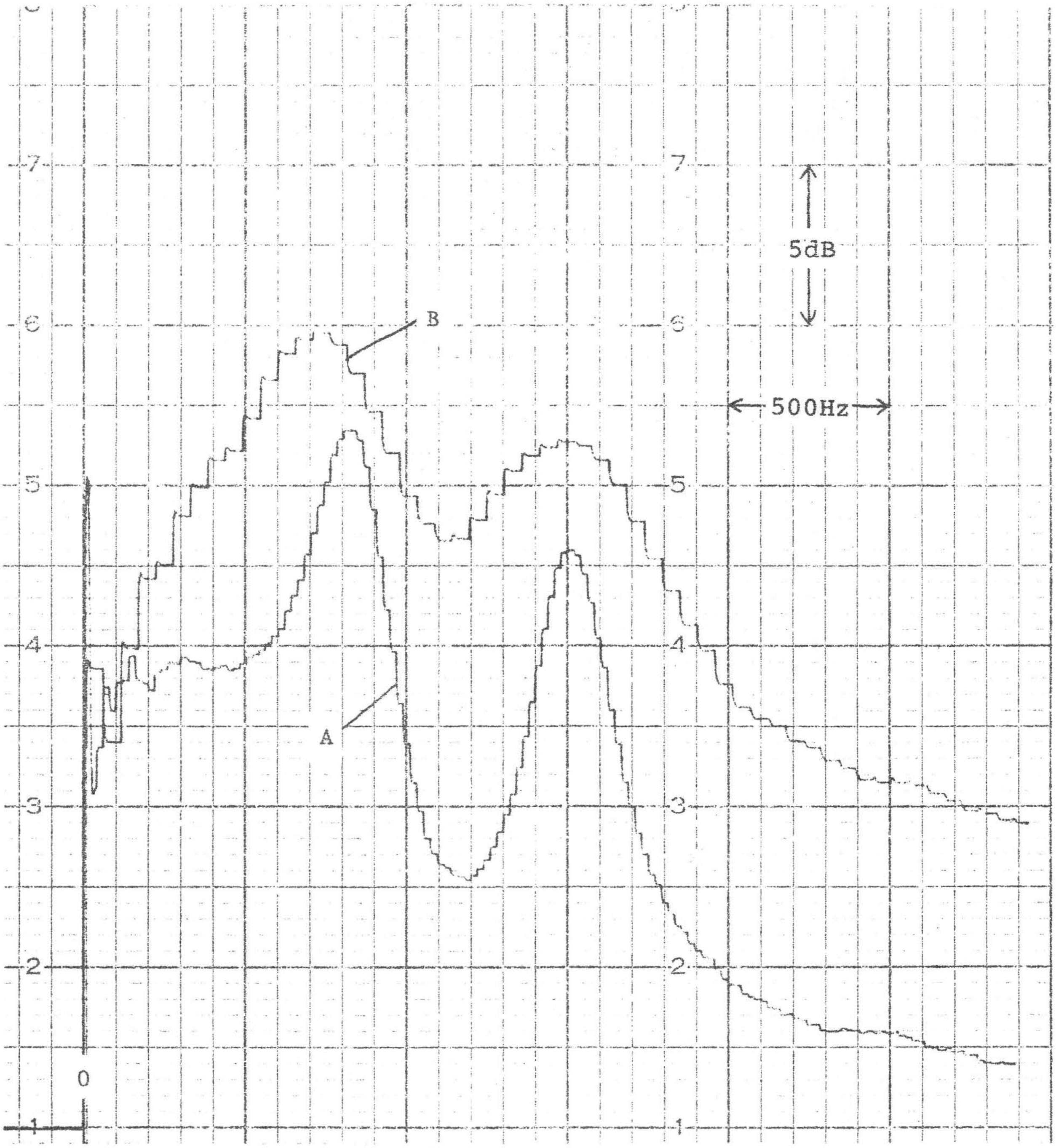


Fig.29

A: Spectrum of the sum of 2 damped sinusoids, F_1=850Hz, F_2=1500Hz, T_o=8msec

B: As A except T_o=3msec

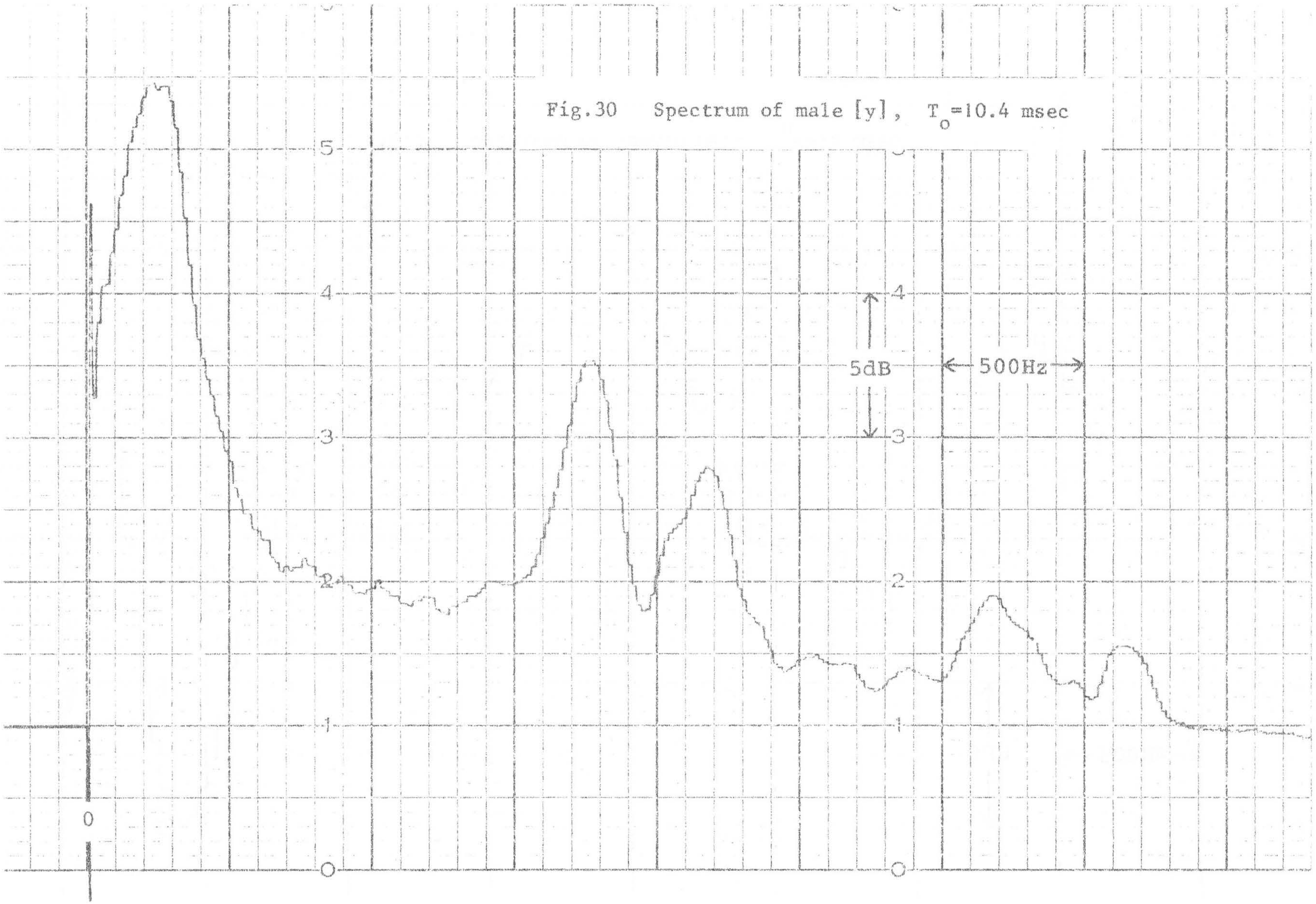


Fig.30 Spectrum of male [y], T_o=10.4 msec

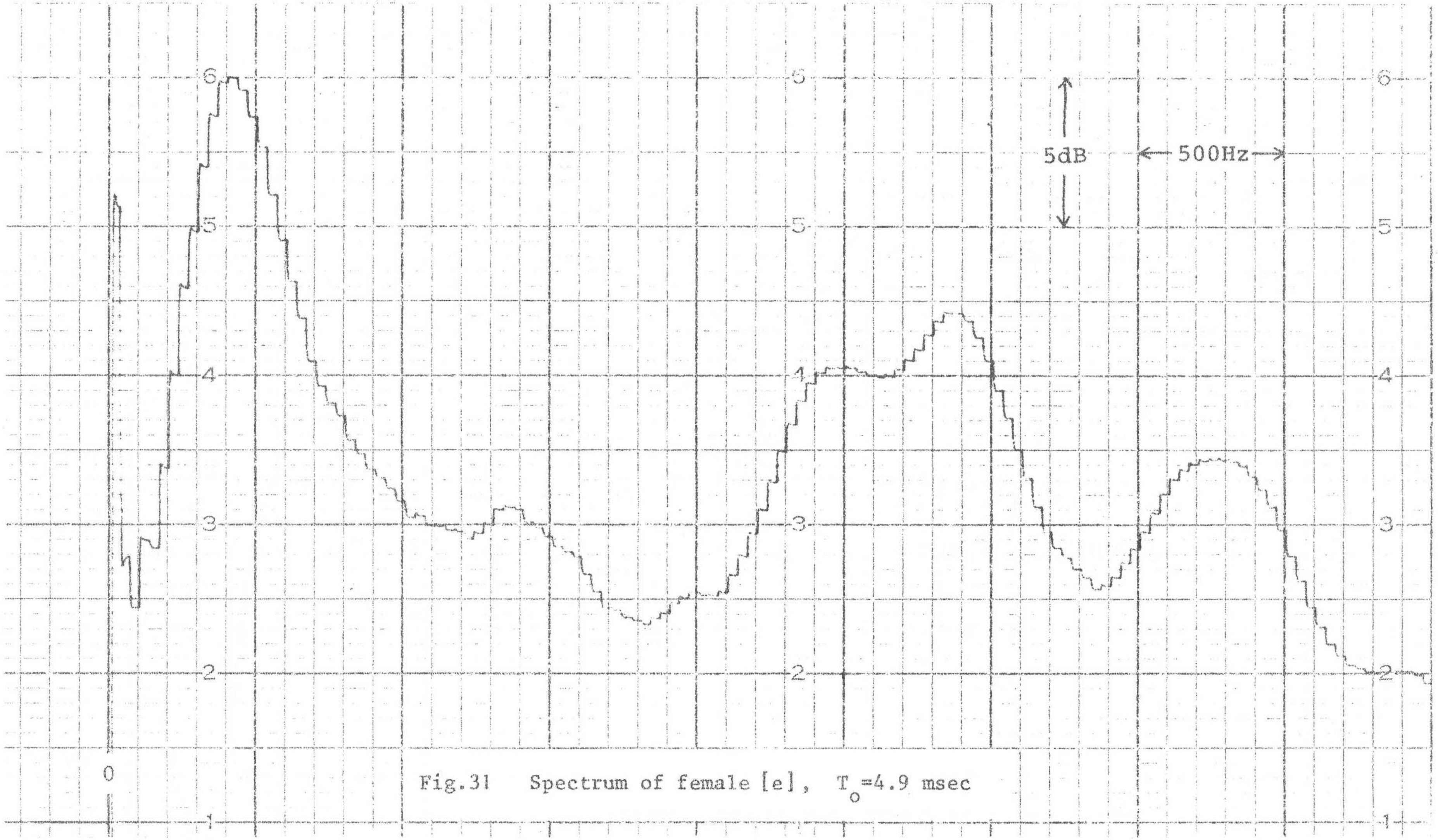


Fig.31 Spectrum of female [e], T_o=4.9 msec

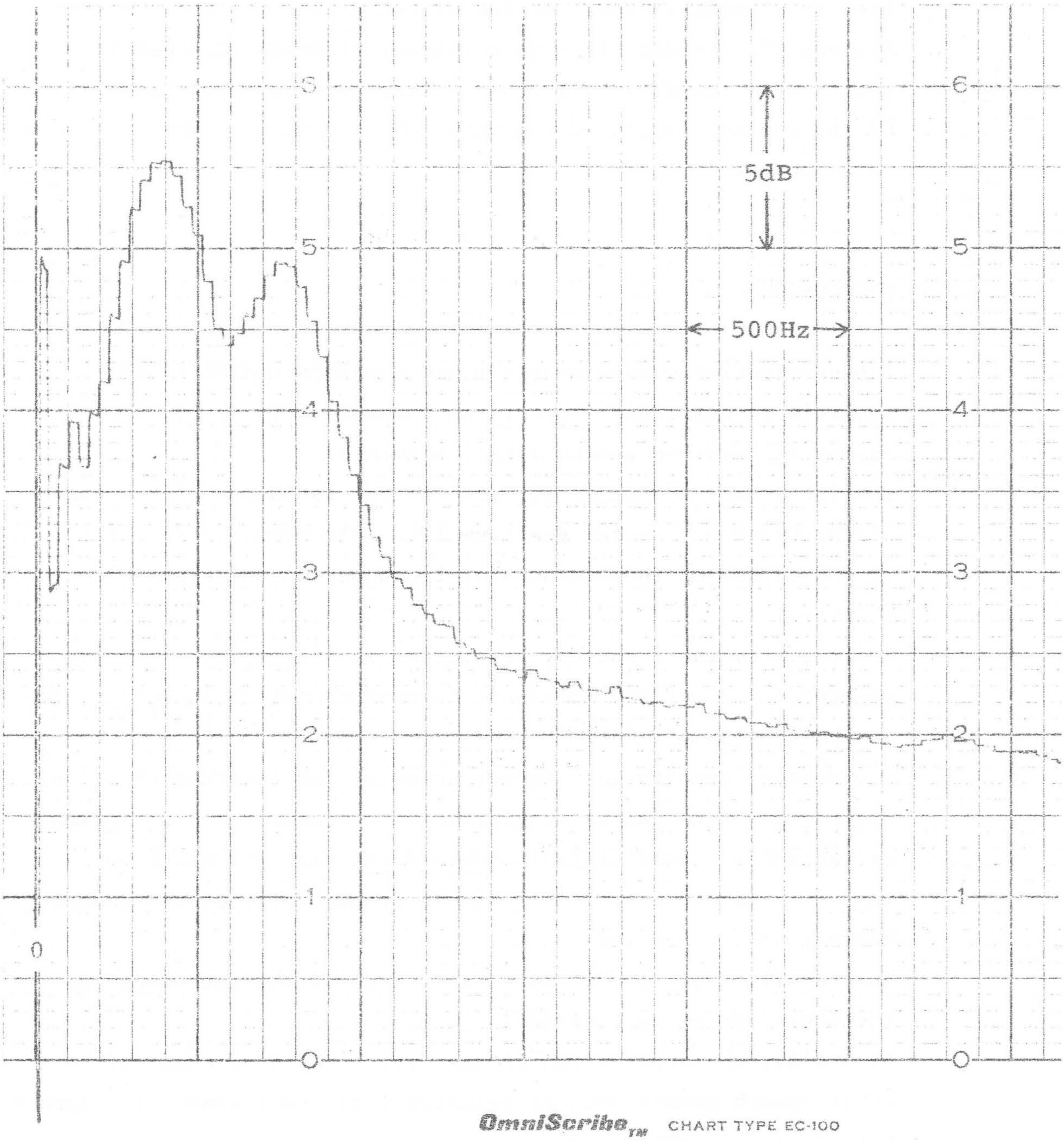


Fig.32 Spectrum of female [o], T_o=5.3 msec

Even at a segment length of one or two periods of the 'formant frequency', its detection is possible, although its spectral peak shifts somewhat to a lower value. As this shift is a function of the segment length, the measure of a correction might be taken into account.

5. CONCLUSION

The exponential window analysis and the interrupted filter analysis as described prove to be valuable methods for the analysis of voiced speech signal segments.
The latter method completely cancels the influence of the fundamental frequency on the spectrum (the well-known 'side lobes') if one period of the fundamental frequency of the signal is selected.
Additionally the frequency resolution is limited by either the segment length or the intrinsic formant bandwidth and <u>not</u> by the analysis method.
The segment spectrograph device based on the interrupted filter method therefore can be considered as a useful tool for formant extraction of speech signals <u>per period</u>, high-pitched signals included. Thus for example reliable formant trajectory measurements could be carried out.
The requirements as mentioned at the end of the Introduction therefore can be considered as fulfilled.
Naturally, the application of the device could be extended to other speech sounds such as fricatives and furthermore to signals outside the speech field.

Acknowledgments.

I'd like to thank all people who, in one way or another, helped to carry out this project and in particular Drs.A.C.M.Rietveld for his stimulating co-operation in an early stage of the experiments and Mr.A.W.van Maanen for his important role in the construction of the segment spectrograph.

BIBLIOGRAPHY

1. Broch, J.T. and Olesen, H.P.
On the frequency analysis of mechanical shocks and single impulses
Brüel & Kjær Technical Review no 3 1970

2. Engelson, M.
Spectrum Analyzer Measurements
Tektronix Inc. 1971

3. Lynn, P.A.
The Analysis and Processing of Signals
Macmillan 1973

4. Markel, J.D.
Digital inverse filtering, a new tool for formant trajectory estimation
IEEE Transactions on Audio and Electroacoustics 1972, AU-20, 129-137

5. Nierop, D.J.P.J.van, Pols, L.C.W. and Plomp, R.
Frequency analysis of Dutch vowels from 25 female speakers
Acustica Vol.29 Heft 2 1973

6. Papoulis, A.
The Fourier Integral and its Applications
Mc Graw-Hill 1962

7. Potter, R.K. and Steinberg, J.C.
Toward the specification of speech
J.A.S.A. vol.22 no 6 p. 807-820 1950

8. Randall, R.B.
Frequency Analysis
Brüel & Kjær 1977

9. Randall, R.B.
High speed narrow band analysis using the digital event recorder type 7502
Brüel & Kjær Technical Review no 2 1973

10. Tellegen, B.D.H.
The gyrator, a new electric network element
Philips Research Reports no 3 1948 p. 81

11. Wempe, A.G.
A system for gating segments of taped speech with great precision
Proceedings 4, Institute of Phonetic Sciences Amsterdam, 1976