TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Development of a computational auditory model

Citation for published version (APA):
van Compernolle, D. S. J. (1991). *Development of a computational auditory model*. (IPO rapport; Vol. 784). Instituut voor Perceptie Onderzoek (IPO).

Document status and date:
Published: 13/02/1991

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 15. Sep. 2025

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

13.02.1991

Rapport no. 784

Development of a Computational Auditory Model

D.S.J. van Compernolle

Instituut voor Perceptie Onderzoek
Postbus 513
5600 MB Eindhoven

Development of a Computational Auditory Model

IPO Technical Report

Dirk Van Compernolle[1]

February 4, 1991

[1]Research Associate of the National Fund for Scientific Research of Belgium (NFWO)

Preface

This report is a summary of the work which I performed on cochlear modeling within the framework of a 2 year cooperation between ESAT-KULeuven and IPO-Eindhoven. This report gives a detailed overview of the development of a computational auditory model and of the obstacles that one can expect on the road towards it. For the casual reader some of the mathematics in it will be painful, but I thought it necessary to include as much detail as possible so that this work can serve as a good technical reference for further development. This report should be considered as a writeup on work in progress. Nevertheless the chapters on cochlear filterbanks and adaptation have reached a more or less finished form, while on the other hand the chapter on data representation and post processing leaves many questions unanswered. I hope to be able to continue work on this topic and present a more conclusive report at some point in the future.

The topic of cochlear modeling, though closely related to my previous experience, was not the core topic of my research at KULeuven during this period nor a mainstream activity at IPO. Hence this part-time cooperative research activity was an experiment and challenge both for IPO and myself. It was hard to put continuation in this "one-day-a-week" effort, often leading to frustration because of the slow progress associated with such a work schedule. Looking back on it afterwards and at this report I should conclude, however, that the time was well spent and I hope that cooperation between KULeuven and IPO will continue, be it in a more informal way. Moreover my stay at IPO had more than enough nice sides to compensate for the hard edges. As a researcher I found it refreshing and stimulating to have a "second home". So in this introductory note a warm thanks belongs to all members of the group "Horen en Spraak" for the help, talks, discussions, formal or not, which I had with them over the past two of years.

Contents

1 Introduction 5
- 1.1 Motivation . . . 5
- 1.2 Auditory Pathways . . . 5

2 A Cochlear Filterbank based on simple Multipole Filters 7
- 2.1 Preprocessing by Outer and Middle Ear . . . 7
- 2.2 The Cochlea as a Filterbank . . . 7
- 2.3 Frequency Scales . . . 8
- 2.4 Gammatone Filters . . . 9
  - 2.4.1 Impulse and Frequency Response . . . 9
  - 2.4.2 Damping Factor and Bandwidths of Gammatone Filters . . . 10
  - 2.4.3 Examples . . . 11
- 2.5 APPENDICES . . . 13
  - 2.5.1 APPENDIX I: Transfer Functions of Gammatone Filters . . . 13
  - 2.5.2 APPENDIX II: Digital Implementation of Gammatone Filters . . . 13

3 Adaptation in the Inner Hair-Cell - Auditory Nerve Synapse 15
- 3.1 Introduction . . . 15
- 3.2 Schroeder-Hall Model . . . 16
  - 3.2.1 Model Concept . . . 16
  - 3.2.2 Mathematical Description . . . 16
  - 3.2.3 Properties . . . 17
- 3.3 Meddis Model . . . 18
  - 3.3.1 Model Concept . . . 18
  - 3.3.2 Mathematical Description . . . 19
  - 3.3.3 Input Nonlinearity . . . 20
  - 3.3.4 Steady State Properties . . . 22
  - 3.3.5 Linearization of the Meddis Model . . . 24
  - 3.3.6 Dynamic Behaviour . . . 25
  - 3.3.7 Summary of Design Parameters . . . 26
  - 3.3.8 Adaptation Examples for Sinusoidal Bursts . . . 26

4 Post Processing in Auditory Models 27
- 4.1 Introduction . . . 27
- 4.2 Average Rate . . . 27
- 4.3 Synchrony Measures . . . 28
  - 4.3.1 Synchrony Measures for known Characteristic Frequencies . . . 28
  - 4.3.2 Synchronization Index . . . 29
  - 4.3.3 Predictive Synchrony Rate . . . 30
  - 4.3.4 Examples of Noise Robustness of Synchrony Measures . . . 30

4.4 Synchrony Measures with Interval Histograms 30

5 Software for Auditory Modeling **32**
5.1 Introduction 32
5.1.1 File Conventions 32
5.1.2 VAX/VMS User Interface 32
5.2 Main Programs 33
5.3 Subroutine Library 33
5.4 Code and Demos 33
5.4.1 The AMOD Directory 33
5.4.2 Filter Design Illustrations 34
5.4.3 Demo Directory 34

Chapter 1

Introduction

1.1 Motivation

Over the past decade computational models of the peripheral auditory system have gained popularity as front ends to automatic speech recognition systems [1, 2, 3] or as general analysis tools for speech research [4, 5]. These models have shown that in complex speech processing applications classical spectral analysis can be modified to one's advantage by adding principles from auditory processing. The evidence is largely empirical, however, and the precise contribution of individual blocks has not been sufficiently analyzed, nor is it clear why certain combinations of features don't work. The goal of this work is the development of a complete auditory model based on up to data physiological and psychoacoustic data with a special attention to the "why" of each processing block. Existing models have such important basic distinctions that it is obvious that in each of them a set of auditory features was selected which happened to perform well with a given application in mind. Apart from empirical evidence, the principal motivation for use of a cochlear model as a speech analysis tool has been the assumption that a better modeling of the human auditory system is by definition a good thing to do. An important caveat is required here. Physiological modeling is no guarantee for success in automatic speech recognition, and this for two obvious reasons. Mimicking what the ear and brain do might not be a good and will most likely not be an efficient way towards implementing artificial speech recognizers. Today's computers perform simple arithmetic in a manner quite different than humans do and do it much better. A second reason is that animals such as the squirrel monkey, cat and guinea pig all have peripheral auditory systems which are quite similar to the human one but their performance as a speech recognizer is poor and in several applications they will be outperformed by existing artificial systems with a poor model of the auditory periphery.

1.2 Auditory Pathways

Data flow and the corresponding signal processing role of each part in the human auditory system is schematically shown in Fig.1.1. Physiological understanding of processing in outer and middle ear is excellent, it is good as far as filtering inside the cochlea is concerned, and gradually gets worse as we move higher up the auditory chain. Models of the neural transduction process are much more speculative, though lots of data is available from single fiber recordings on the auditory nerve. And what happens beyond the first synapses of the auditory nerve is total speculation. How the brain interprets the spike trains delivered by 30.000 parallel channels in not at all known.

The model presented in this report contains three sections, two of which are physiologically well motivated and one which is required in order to make sense out of the two preceeding ones:

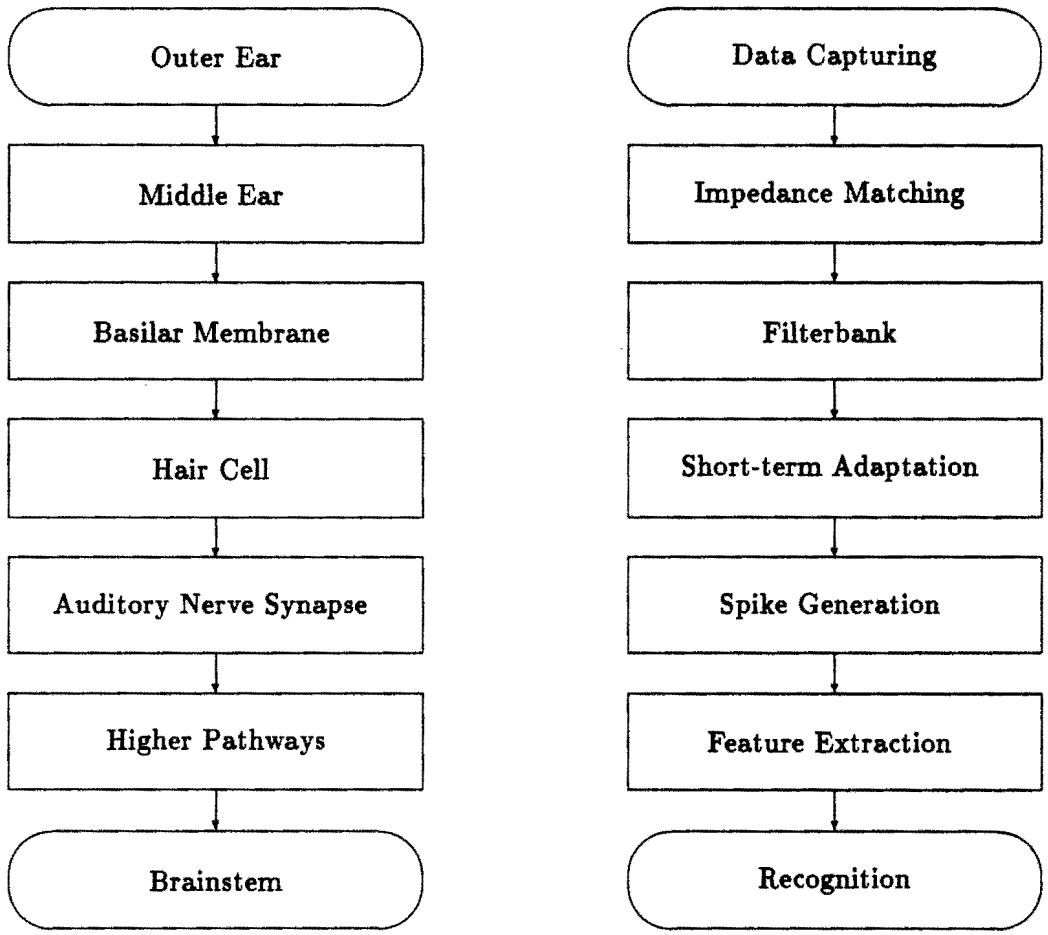


Figure 1.1: Auditory Pathways: Physiological and Functional Equivalents

1. Filterbank (middle ear + basilar membrane)
2. Adaptation (hair cell + synapse)
3. Post Processing : Data Analysis and Representation (feature extraction in higher pathways)

The output of the second section is a neural spike train which contains much detail and which is not suitable for interpretation as such. A high level of abstraction is required to reduce the data rate to a manageable level such that data interpretation or use of the data in a speech recognizer becomes possible. Controversies about "rate" or "synchrony" are at this level and can not be solved by physiological arguing until a much better understanding of high level neural processing becomes available.

Chapter 2

A Cochlear Filterbank based on simple Multipole Filters

2.1 Preprocessing by Outer and Middle Ear

The outer ear is the microphone of the auditory system, its role being the interface between the outside and inside worlds, without any signal processing role associated with it. The role of the middle ear is impedance matching between the different acoustic impedances of the air and the cochlear fluids. For very loud sounds non-linearities provide also a protective function. For common sounds signal processing is limited to bandpass filtering in the auditory range (20Hz-20kHz) with an emphasis on the most important speech range (1kHz-4kHz). The outer and middle ear will not be considered explicitly in the rest of this work, as the passive middle ear filtering can easily be included as a channel dependent gain in the cochlear filterbank.

2.2 The Cochlea as a Filterbank

The sound pressure wave induced in the cochlea by the stapes at the oval window propagates as a traveling wave on the basilar membrane and in the cochlear fluids from apex to helicotrema. The motion of the basilar membrane in turn results in the bending of hair cells which are sitting on top of it. These hair cells (there are roughly 30.000 of them) connect to the auditory nerve. The most remarkable characteristic of the traveling wave inside the cochlea is its strong frequency selectivity, and was first described by Georg von Békésy[6]. The signal processing function of the basilar membrane and the surrounding structures is to filter the incoming broadband sound into 30.000 narrowband channels.

A CONCEPTUAL COMPROMISE Current computer technology does not allow for simulation of a 30.000 channel filterbank. In practice 100 seems to be more or less an upper limit. Hence an important conceptual decision has to be made right from the start: should a single filterbank channel model *a single nerve fiber* or should it model *a local group of fibers* ? A human ear with only 100 surviving fibers can be considered as virtually deaf, hence the second option seems to be the appropriate one. Detailed modeling of the filter characteristic of a single fiber is interesting from a physiological viewpoint but currently has no place in a "global auditory model". In an auditory model, a single channel should model a local group of fibers, rather than a single one. One immediate consequence is that the incredible sharpness at the tip of a tuning curve of a single auditory nerve fiber will (and should) not be reflected in the filterbank. In this chapter a class of cost effective and easy to parametrize filters is described which are a reasonable match to the auditory filterbank.

2.3 Frequency Scales

Fourier analysis is the most widely used non parametric spectral estimation technique. A trivial interpretation is that of a narrowband filterbank analysis, with equally spaced and equally wide filters and with a single analysis window. The single channel impulse response in Fourier Analysis is the analysis window modulated by a (co)sine at channel frequency.

The clearest deviation of auditory frequency analysis from Fourier analysis is its use of a non-linear frequency axis and its use of different analysis windows for each channel. Channel spacing at low frequencies is dense and near linear while at high frequencies the auditory filters are wide and almost logarithmically spaced. Evidence of this auditory frequency scale comes from physiological as well as psychoacoustic measurements. Several scales have been proposed (mel, bark, ERB) which all are slightly different, depending on the empirical data that they were derived from. I have opted for the most recent one, i.e. the ERB (Equivalent Rectangular Bandwidth) scale, as used by B. Moore[7]. The ERB scale has a close relationship to the critical band concept, as "equivalent rectangular bandwidth" is defined as the width of a rectangular filter, which gives the same output power to a white noise input, as a cochlear filter with the same response at characteristic frequency(CF). From the consideration that a filterbank channel models a group of fibers it is plausible to have the filterbank design guided to a large extent by psychoacoustic data and not only by physiological data. Mathematically the ERB scale relates

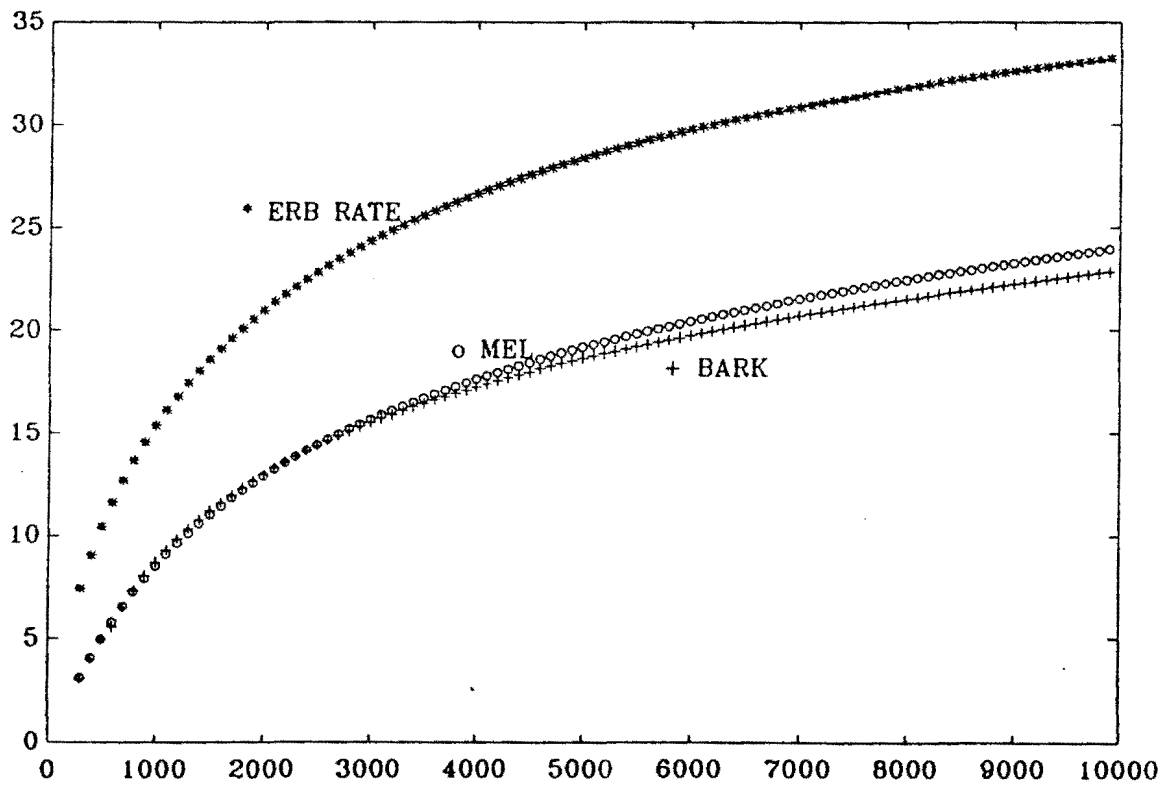


Figure 2.1: Auditory Frequency Scales

bandwidth with center frequency by following formula:

$$ERB = 6.23f^2 + 93.39f + 28.52 \quad (2.1.a)$$

$$ERBR = 11.17 \ln\left|\frac{f + 0.32}{f + 14.675}\right| + 43.0 \quad (2.1.b)$$

in which f is a channel center frequency in kHz and ERB the associated bandwidth in Hz. $ERBR$ is the 'ERB-rate', i.e. a linear scale in the warped frequency domain. The ERB frequency

Freq(Hz)	ERB(Hz)	ERBR	Mel	Bark
200	47	5.6	2.1	2.1
468	74	10.0	2.7	4.7
1000	128	15.4	8.5	8.7
1779	214	20.0	12.1	12.3
3200	391	24.8	16.0	15.9
6200	847	30.0	20.7	20.0

Table 2.1: Frequency Scales

scale is compared with two other commonly used frequency scales, the Mel and Bark scales in Fig.2.1 and Table 2.1. All scales are quite similar, however the ERB-scale suggests that considerable more channels are required at the low frequency end. The mathematical formulas describing these other scales are:

$$\begin{aligned} \text{MEL} \quad \text{m} &= 7 arcsinh\left(\tfrac{f}{.65}\right) \\ \text{BARK} \quad \text{b} &= 13 atan(0.76f) + 3.5 atan\left((f/7.5)^2\right) \\ &= 8.7 + 14.2 log(f) \qquad\qquad (f > 0.6kHz) \end{aligned}$$

2.4 Gammatone Filters

2.4.1 Impulse and Frequency Response

Filters with a so called gammatone impulse response will be used for modeling of the cochlear filterbank. These filters were first suggested on the basis of reverse correlation modeling[8]. These filters were chosen here because they allow for a simple description of a full cochlear filterbank with very few parameters. The impulse response of a gammatone filter is given by:

$$h(t) = e^{-\alpha\omega_0 t} t^{k-1} \sin\omega_0 t \tag{2.2}$$

In APPENDIX I it is shown that this is the impulse response of a multipole filter with k identical complex pole pairs $p = -\alpha\omega_0 \pm j\omega_0$ and a number of zeroes which contribute very little to the overall filter response if α is small. Omitting scaling factors and the zeroes the transfer function reduces to the following simple expression:

$$H(s) = \frac{1}{\left((s+\alpha\omega_0)^2 + \omega_0^2\right)^k} \tag{2.3}$$

and the corresponding frequency response is:

$$|H(\omega)|^2 = \frac{1}{|(j\omega+\alpha\omega_0)^2 + \omega_0^2|^{2k}} \tag{2.4.a}$$

$$20\log|H(\omega)| = -20k\log|(j\omega+\alpha\omega_0)^2 + \omega_0^2| \tag{2.4.b}$$

The above equations describe a class of bandpass filters (multipole resonators) with centerfrequency ω_0 and high frequency slopes of $12k$ dB per octave. Sharpness of the filters is largely

controlled by the choice of the damping factor α and the sharpness of filters required in cochlear modeling results in typical values for $\alpha << 1$. Truly precise cochlear modeling would require the addition of zeroes slightly above ω_0 to yield steeper high frequency slopes and thus model the asymmetry of the cochlear filters better[5]. For reasons of simplicity in implementation this option was not considered.

2.4.2 Damping Factor and Bandwidths of Gammatone Filters

In order to design a filterbank specified by centerfrequencies and bandwidths we must find a relationship between bandwidth (3-dB or ERB) and the damping factor α. Because of the small damping factors it is possible to simplify the frequency response from (2.4) even further. In and

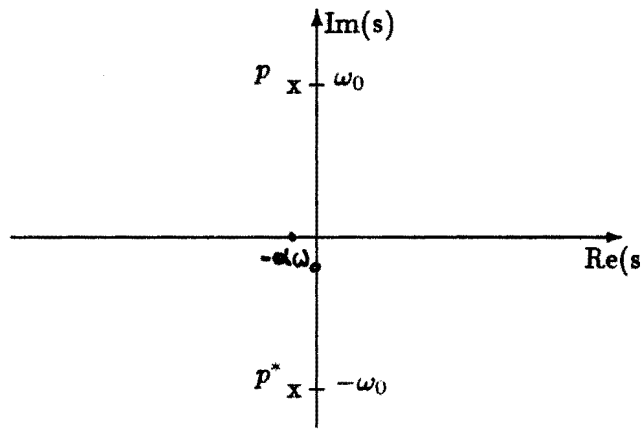


Figure 2.2: Pole Location of a Typical Gammatone Cochlear Filter ($\alpha = 0.15$)

around the passband of a filter only the contribution of one pole from each complex pair must be considered as the contribution of the other one is quite constant[1]. Hence for small α and in the neighbourhood of the complex pole $p = -\alpha\omega_0 + j\omega_0$, i.e. small $\Delta\omega = (\omega - \omega_0)$ we can rewrite (2.4) as:

$$|H(\omega)|^2 = \frac{1}{|j\omega - p|^{2k}|j\omega - p^*|^{2k}} \quad (2.5.a)$$

$$= \frac{A}{|j\omega - p|^{2k}} \quad (2.5.b)$$

$$= \frac{A}{|\alpha\omega_0 + j(\omega - \omega_0)|^{2k}} \quad (2.5.c)$$

$$|H(\omega_0 + \Delta\omega)|^2 = \frac{A}{|(\alpha\omega_0)^2 + (\Delta\omega)^2|^k} \quad (2.5.d)$$

Peak response is reached at resonance frequency ω_0:

$$|H(\omega_0)|^2 = \frac{A}{(\alpha\omega_0)^{2k}}$$

The 3-dB or half energy point is now easily found as the frequency for which the frequency response reaches half this value. Hence:

$$(\alpha^2\omega_0^2 + (\omega_{3dB} - \omega_0)^2)^k = 2(\alpha^2\omega_0^2)^k \quad (2.6.a)$$

[1] This conclusion would equally follow from looking at a pole location plot and applying the so called geometric method for the determination of the filter response.

$$(\omega_{3dB} - \omega_0)^2 = (\sqrt[k]{2} - 1)\alpha^2\omega_0^2 \quad (2.6.b)$$

$$\frac{BW_{3dB}}{\omega_0} = 2\alpha\sqrt{\sqrt[k]{2} - 1} \quad (2.6.c)$$

The ERB bandwidth of the gammatone filters is found by application of the ERB definition. Again the simplified expression (2.5) is used. This simplification is also valid for this derivation because the filter output to a white noise input is largely dominated by the contribution around the center frequency. Furthermore the power output to a white noise input is symmetric with respect to center frequency and is obtained by simple integration:

$$P_F = 2\int_0^\infty \frac{1}{(\alpha^2\omega_0^2 + (\Delta\omega)^2)^k} d(\Delta\omega) \quad (2.7.a)$$

$$= 2\alpha\omega_0 \int_0^\infty \frac{1}{(\alpha\omega_0)^{2k}} \frac{1}{(1+x^2)^k} dx \qquad x = \frac{\Delta\omega}{\alpha\omega_0} \quad (2.7.b)$$

$$= \frac{2}{(\alpha\omega_0)^{2k-1}} \int_0^\infty \frac{1}{(1+x^2)^k} dx \quad (2.7.c)$$

$$= \frac{2}{(\alpha\omega_0)^{2k-1}} \frac{2k-3}{2k-2}\cdot\frac{2k-5}{2k-4}\cdots\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{\pi}{2} \quad (2.7.d)$$

The power output of a rectangular filter with response at centerfrequency equal to $|H(\omega_0)|^2$ and bandwidth ERB is given by:

$$P_R = ERB.|H(\omega_0)|^2 = ERB.\frac{1}{(\alpha\omega_0)^{2k}} = \frac{ERB}{\alpha\omega_0(\alpha\omega_0)^{2k-1}} \quad (2.8)$$

Thus from $P_F = P_R$:

$$\frac{ERB}{\omega_0} = 2\alpha\frac{2k-3}{2k-2}\cdot\frac{2k-5}{2k-4}\cdots\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{\pi}{2} \quad (2.9)$$

For the lower orders of 'k' Table 2.2 summarizes the 3-dB and ERB bandwidths as a function of α and centerfrequency.

k	$\frac{BW_{3dB}}{\omega_0}$	$\frac{ERB}{\omega_0}$
1	2α	3.14α
2	1.28α	1.57α
3	1.02α	1.18α
4	0.86α	0.98α
5	0.77α	0.86α

Table 2.2: Relation between α, filter order and bandwidths

For orders 2-4 and given centerfrequency a number of α's for the ERB filters are computed in Table 2.3. These values clearly illustrate that the assumption "α small" is quite valid throughout.

2.4.3 Examples

In this section filterbank designs for different choices of filter orders are illustrated. The basic design is a filterbank in which the channels have minimal overlap and with both spacing and bandwidth of each filter equal to 1 ERB. Twenty channels cover the frequency range most

Freq(Hz)	ERB(Hz)	ERBR	$\alpha(k=2)$	$\alpha(k=3)$	$\alpha(k=4)$
200	47	5.6	0.150	0.199	0.239
468	74	10.0	0.115	0.154	0.185
1000	128	15.4	0.081	0.109	0.130
1779	214	20.0	0.077	0.102	0.122
3200	391	24.8	0.078	0.104	0.125
6200	847	30.0	0.087	0.116	0.139

Table 2.3: ERB Frequency Scale and Filter Design Parameters

important for speech purposes i.e. from 333 to 4181 Hz (8 to 27 ERBR). For comparison from a signal processing viewpoint a Hamming filterbank with linear and ERB spacing is also given. Also filters with characteristics used by Flanagan[9] and based on the original Békésy data are shown. These filters also have a gammatone impulse response but much wider bandwidths ($\alpha = 0.5$!!) as it is well known that the Békésy filter characteristics are much too shallow due to the extreme sound pressures used and resulting non-linearities.

The illustrated designs are:

(a) Hamming Filterbank: impulse responses are 256 pt. cosine modulated Hamming windows, yielding very sharp filters.

(b) A Békésy/Flanagan filterbank: second order gammatone filters with fixed damping factors $\alpha = 0.5$.

(c) Second order gammatone filters with variable damping: $\alpha\omega_0 = \frac{ERB}{1.57} = 0.637ERB$.

(d) Fourth order gammatone filters with variable damping: $\alpha\omega_0 = \frac{ERB}{0.98} = 1.019ERB$.

(e) This design is for comparison from a signal processing viewpoint only: a linearly spaced 20 channel Hamming filterbank spanning the 200Hz to 4000Hz range with 200Hz bandwidth for each channel.

Fig. 2.3 illustrates frequency and impulse responses for the channel with centerfrequency of 2006Hz for the designs (a-d). In Figs. 2.4(a-e) frequency responses of full filterbank designs according to the different methods are shown, while Figs. 2.5(a-e) show the corresponding impulse responses. The 4th order gammatone design was in several studies found to be the most appropriate one for cochlear modeling and will be used as the reference model throughout the rest of this report [10, 11, 12][2]. Fig. 2.4.d illustrates a nice side property of the reference filterbank design: it has excellent analysis-synthesis properties in a classical signal processing sense. The sum of the individual channel filter responses is almost unity. The ripple over most of the pass-band is less than 0.1 dB, though considerable higher around the edges (which can be reduced by including more channels) . A sum of filterbank outputs will apart from a phase shift barely differ from the original input signal.

[2]Exactly the same value 1.019 is used as damping factor in [10]; I'm not aware, however, how this value was derived.

2.5 APPENDICES

2.5.1 APPENDIX I: Transfer Functions of Gammatone Filters

The Laplace transform corresponding to an impulse response of the form

$$h(t) = e^{-\alpha\omega_0 t} t^{k-1} \sin\omega_0 t \tag{2.10}$$

can best be obtained using following differential equations:

$$\frac{d}{dt}(t^k \sin\omega t) = kt^{k-1} sin\omega t + \omega t^k \cos\omega t \tag{2.11.a}$$

$$\frac{d}{dt}(t^k \cos\omega t) = kt^{k-1} \cos\omega t - \omega t^k \sin\omega t \tag{2.11.b}$$

Taking Laplace transforms of both sides and using the recursion twice we get :

$$(s^2+\omega^2)S_k(s) = skS_{k-1}(s) + \omega kC_{k-1}(s) \tag{2.12.a}$$

$$(s^2+\omega^2)C_k(s) = skC_{k-1}(s) - \omega kS_{k-1}(s) \tag{2.12.b}$$

Starting from the known Laplace transform pairs for k equal to 0 and 1, it is possible to derive the exact Laplace transform pair for any power k.

$f(t)$	$F(s)$
$\sin\omega_0 t$	$\frac{\omega_0}{s^2+\omega_0^2}$
$\cos\omega_0 t$	$\frac{s}{s^2+\omega_0^2}$
$t\sin\omega_0 t$	$\frac{2\omega_0 s}{(s^2+\omega_0^2)^2}$
$t\cos\omega_0 t$	$\frac{s^2-\omega_0^2}{(s^2+\omega_0^2)^2}$
$t^2\sin\omega_0 t$	$\frac{2\omega_0(3s^2-\omega_0^2)}{(s^2+\omega_0^2)^3}$
$t^2\cos\omega_0 t$	$\frac{2s(s^2-3\omega_0^2)}{(s^2+\omega_0^2)^3}$
$t^3\sin\omega_0 t$	$\frac{24\omega_0 s(s^2-\omega_0^2)}{(s^2+\omega_0^2)^4}$
$t^3\cos\omega_0 t$	$\frac{6(s^4-6\omega_0^2 s^2+\omega_0^4)}{(s^2+\omega_0^2)^4}$

Table 2.4: Impulse Response and Laplace Transform Pairs for Gammatone Filters

The above table summarizes transform pairs for filters with zero damping. The influence of the damping factor α is the addition of $e^{-\alpha\omega_0 t}$ in the impulse response what corresponds to replacing s by $s+\alpha\omega_0$ in the Laplace transforms. From the above table it can be seen that an all-pole filter approximation will be excellent as long as α is small, i.e. for sharp filters, which is the case for a cochlear filterbank.

2.5.2 APPENDIX II: Digital Implementation of Gammatone Filters

A most straightforward digital implementation of the gammatone filters is to use the impulse responses directly and implement them as FIR filters. This style of implementation is numerically

very stable and precise, but computationaly expensive, especially for the low frequency channels. Alternatively, one can apply the 'impulse invariant' mapping from s-domain to z-domain. This technique is quite appropriate for the narrow bandpass filters at hand. The impulse invariant mapping technique, maps all s-plane poles and zeros to corresponding z-plane poles and zeros, using the standard formula:

$$s \rightarrow e^{sT}$$

in which T is the sampling period. For the above example, this implies a mapping of the s-domain poles to:

$$p_s = -\alpha\omega_0 \pm j\omega_0 \quad \rightarrow \quad p_z = e^{-\alpha\omega_0 T}(\cos\omega_0 T \pm j \sin\omega_0 T)$$

resulting in a second order block per complex pole pair of the form:

$$H(z) = \frac{1}{1 - 2e^{-\alpha\omega_0 T}\cos\omega_0 T z^{-1} + e^{-2\alpha\omega_0 T} z^{-2}}$$

which is implemented in the time domain as:

$$y(k) = x(k) + 2e^{-\alpha\omega_0 T}\cos\omega_0 T y(k-1) - e^{-2\alpha\omega_0 T} y(k-2)$$

Chapter 3

Adaptation in the Inner Hair-Cell - Auditory Nerve Synapse

3.1 Introduction

The mechano-electrical transduction at the inner hair-cell - auditory nerve synapse is an important element in the peripheral auditory signal processing chain as it is at this level that short term adaptation should be situated. Modern understanding of the mechano-electrical transduction is based on following principles:

- The motion of the basilar membrane is passed on to the inner hair cells, the last mechanical element in the auditory processing chain. The signal content of hair cell motion is a frequency sharpened version of the local basilar membrane motion [5]. This filtering is in principle included in the filterbank design of the previous chapter.
- The permeability of the inner hair cell membrane is a function of the bending of the hair cell. Permeability functions of most hair cells, including cochlear inner hair cells, have two common characteristics: halfwave rectification and saturation.
- Chemical transmitters are available inside the hair cell and their release from the hair cell into the synaptic cleft is controlled by the membrane permeability.
- Nerve fiber firing probability is, except for refractory properties, proportional to the amount of chemical transmitter available in the synaptic cleft.
- Chemical transmitters dissipate quickly from the synaptic cleft and find their way back into the original pool because of electrical imbalance or other mechanisms.

Modeling the mechano-electrical transduction process means deriving a mathematical relationship between hair cell motion and concentration of chemical transmitter in the synaptic cleft, or similarly nerve firing probability. One of the first and most simple models based on these principles is the widely used Schroeder-Hall model[13]. Using one non-linear differential equation and one static nonlinearity it models fairly well the adaptation behaviour of single burst onsets and offsets in silence. Since its introduction in 1974 more physiological measurements have become available which show some deficiencies in the SH-model, especially concerning the modeling of transients in the presence of a pedestal.

Many models have built on the SH model, trying to explain equally well the more recent physiological data. Some of them require the subdivision of chemical transmitter in global and many local pools with different time constants associated with them[14]. These models are highly complex and computationally very demanding. One of the simpler models, proposed by

R.Meddis[15, 16] [1], uses only 3 first order coupled non-linear differential equations to describe the inner hair-cell - auditory nerve synapse. This model was chosen as the base model in this work, because the computational load is relative small and because it seemed capable of modeling most of the described neural adaptation characteristics. It also has the basic possiblities in it to characterize different types of neurons. The model is much more complex, however, than one would expect at first glance, because of the presence of multiple non-linearities and is therefore very hard to parametrize. R. Meddis followed a trial and error design procedure in which he described system characteristics as a function of model parameters, rather than setting parameters in function of desired characteristics[17]. This way he was able to design a class of different fibers each with their own properties. However he did not show how to derive parameters from a set of specifications, nor exactly which class of neurons could be covered by the model.

In this chapter we will first review the Schroeder-Hall and Meddis models. Then we will take a constructive approach to parametrizing the Meddis model and describe how to design a "Meddis synapse" according to specs. A common nomenclature, applicable to both models, is used throughout so that names and symbols will slightly deviate from the original papers. Lower case symbols are used for system variables and upper case ones for parameters. The overstrike is used to indicate "one cycle averages" in steady state analysis.

3.2 Schroeder-Hall Model

This section is a short summary of the relevant parts of [13].

3.2.1 Model Concept

The model is defined by four rules:

- Quanta (electrochemical agents) are generated in the hair cell at a fixed average rate and stored in a temporary pool from where they are lost are can be released in to the synaptic cleft.
- Quanta move into the synaptic cleft at a rate proportional to their number and a permeabilitiy function.
- Nerve firing is proportional to the number of quanta released in to the synaptic cleft.
- Quanta disappear from the free pool at a rate proportional to their number without having any effect on the firing.

3.2.2 Mathematical Description

SH-model Variables:

- $s(t)$: input signal
- $p(t)$: permeability
- $q(t)$: free pool concentration
- $c(t)$: cleft concentration
- $f(t)$: firing rate

[1] The original paper contains a serious mathematical error. The dB scale is of by a factor of 2, hence parametrizations in it are senseless

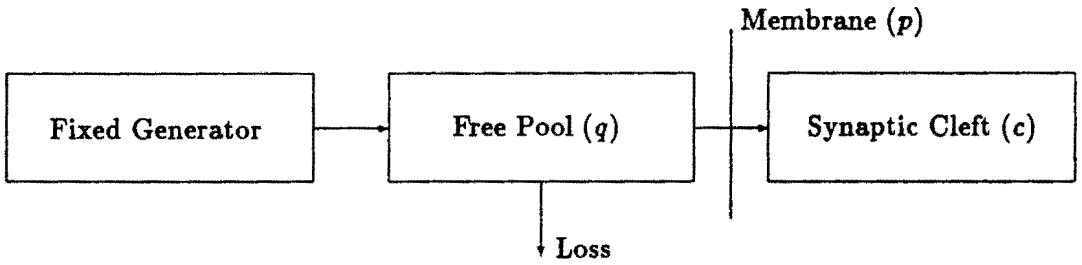


Figure 3.1: Schroeder-Hall Model

SH parameters:

- G: generator rate ($150sec^{-1}$)
- L: loss rate ($33.3sec^{-1}$)
- P_0: permeability constant ($16.7sec^{-1}$)

Model Equations:

$$p(t) = P_0\{\frac{1}{2}s(t) + \{\frac{1}{4}s^2(t) + 1\}^{\frac{1}{2}}\} \quad (3.1.a)$$

$$\dot{q}(t) = G - L.q(t) - p(t).q(t) \quad (3.1.b)$$

$$c(t) = p(t).q(t) \quad (3.1.c)$$

$$f(t) \approx c(t) \quad (3.1.d)$$

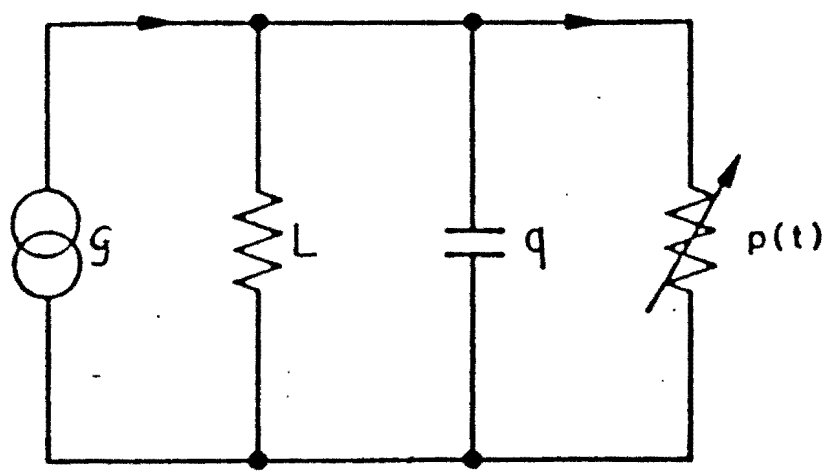


Figure 3.2: Electrical Equivalent of Schroeder-Hall Model

3.2.3 Properties

A closer look at Fig.3.1 and the corresponding equations gives us an understanding of the basic principles of this model and all its derivatives. The electrical equivalent from Fig.3.2 can also help in understanding. For a steady periodic input following behaviour will emerge:

- After an initial transient behaviour the whole system will evolve to a periodic behaviour.

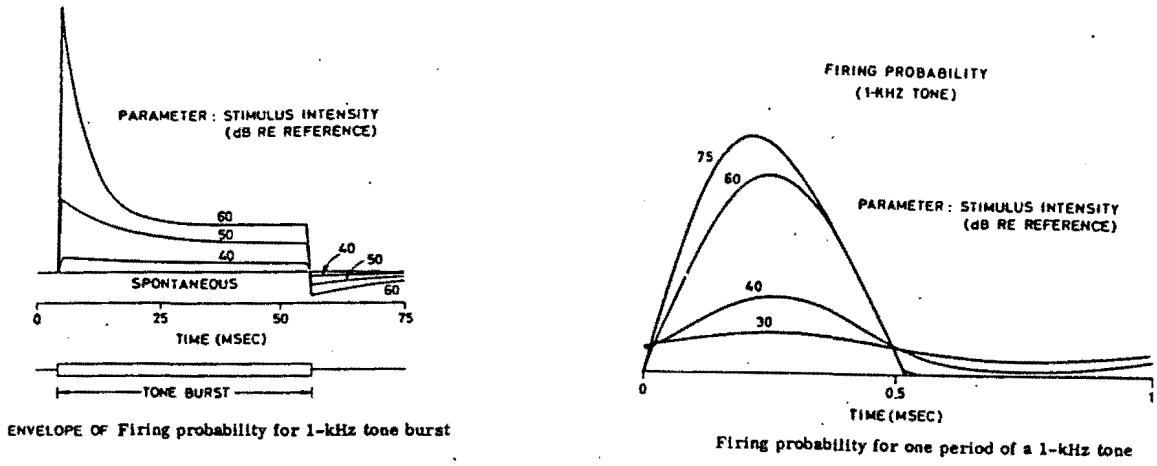


Figure 3.3: Schroeder-Hall Model Properties

- The quanta produced by the generator or either lost or dissipated in the cleft. The larger the average permeability the larger the proportion of quanta that will go into the cleft and induce nerve firing. The maximum average nerve firing is limited by the generator, the direct cause for rate saturation. Zero input will result in a non-zero spontaneous firing due to membrane leakage. Average free pool contents will be smaller with larger average firing rate.

- Onset and offset phenomena are due to the fact that the free pool needs time to settle down in its new equilibrium. With a sudden onset of stimulus a high free pool content coincides with a high membrane permeability resulting in initial firing rate overshoot, while a sudden offset will result in firing rate undershoots.

- On top of the overall long-term behaviour a "within cycle" behaviour is superimposed. The true firing probability is approximately - except for very low frequencies - the average firing rate modulated by the half-wave rectified input signal. This is the underlying cause for phase locking.

Average firing rates and within cycle firing rate probabilities are illustrated in Fig.3.3

3.3 Meddis Model

3.3.1 Model Concept

There are a few important differences between the Meddis (Model B in [15]) and Schroeder Hall models:

- Influx of quanta into the free transmitter pool from the factory is not constant but controlled by a gradient mechanism.

- Diffusion of quanta from the cleft is not immediate, but the cleft is treated as a pool with its own time constants. This results in an upper frequency limit for phase locking.

- There is immediate recuperation of quanta from the cleft into the hair cell. This phenomenon can be used to obtain a better modeling of dynamic behaviour in the presence of a pedestal.

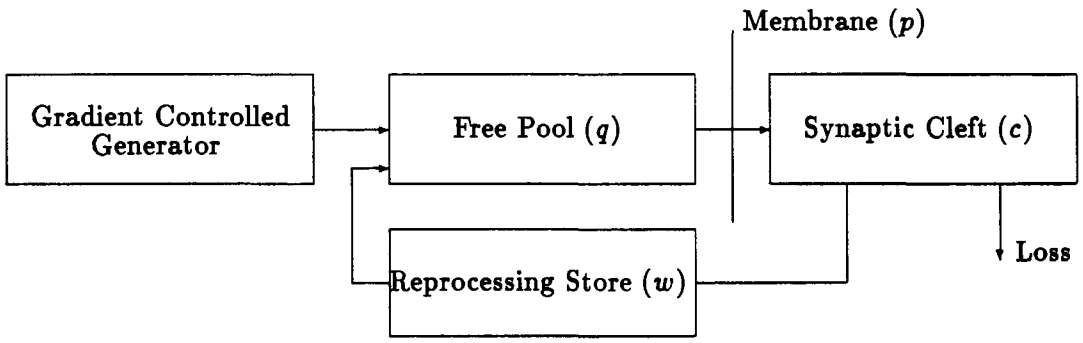


Figure 3.4: Meddis Model

3.3.2 Mathematical Description

The concentration variables are all rescaled relative to the generator and are therefore all in the range $[0, 1]$.

Variables:

- $q(t)$: free pool concentration
- $c(t)$: cleft concentration
- $w(t)$: reprocessing store concentration
- $f(t)$: firing rate
- $p(t)$: permeability
- $s(t)$: input signal

Parameters: Y, X, L, R, H, K, A, B

- K, A, B: parameters controlling the permeability function
- Y: factor controlling gradient flow from generator to free pool
- L: loss time-constant from synaptic cleft
- R: reuptake time-constant from synaptic cleft to reprocessing store
- X: reuptake time-constant from reprocessing store to free pool
- H: proportionality factor between cleft contents and firing rate

Other Symbols and Subscripts:

- M-subscript: max values
- 0-subscript: 0-input values
- α: amplitude of input sinusoid
- ΔT: sampling period in discrete implementation
- T: period of a sinusoid
- δ: firing rate dynamic range
- $\bar{}$: one period averages of a parameter

The Meddis model is described by one static input non-linearity and a set of 3 coupled (non-linear) first order differential equations. Firing rate is proportional to one of the system variables.

Continuous Time Model:

$$p(t) = K\frac{s(t)+A}{s(t)+A+B} \tag{3.2.a}$$

$$\dot{q}(t) = Y.(1-q(t)) + X.w(t) - p(t).q(t) \tag{3.2.b}$$

$$\dot{c}(t) = p(t).q(t) - L.c(t) - R.c(t) \tag{3.2.c}$$

$$\dot{w}(t) = R.c(t) - X.w(t) \tag{3.2.d}$$

$$f(t) = H.c(t) \tag{3.2.e}$$

Discrete Time Model: is derived from the continuous one with a simple forward Euler approximation. Values for q,c and w at time $t+\Delta T$ are obtained as: $q(t+\Delta T) = q(t) + \Delta q \ldots$.

$$p(t)\Delta T = (K\Delta T)\frac{s(t)+A}{s(t)+A+B} \tag{3.3.a}$$

$$\Delta q = (Y\Delta T).(1-q(t)) + (X\Delta T).w(t) - (p(t)\Delta T).q(t) \tag{3.3.b}$$

$$\Delta c = (p(t)\Delta T).q(t) - (L\Delta T).c(t) - (R\Delta T).c(t) \tag{3.3.c}$$

$$\Delta w = (R\Delta T).c(t) - (X\Delta T).w(t) \tag{3.3.d}$$

$$f(t) = H.c(t) \tag{3.3.e}$$

3.3.3 Input Nonlinearity

The nonlinearity in the permeability function

$$p(t) = K.\frac{A+s(t)}{A+B+s(t)} \tag{3.4}$$

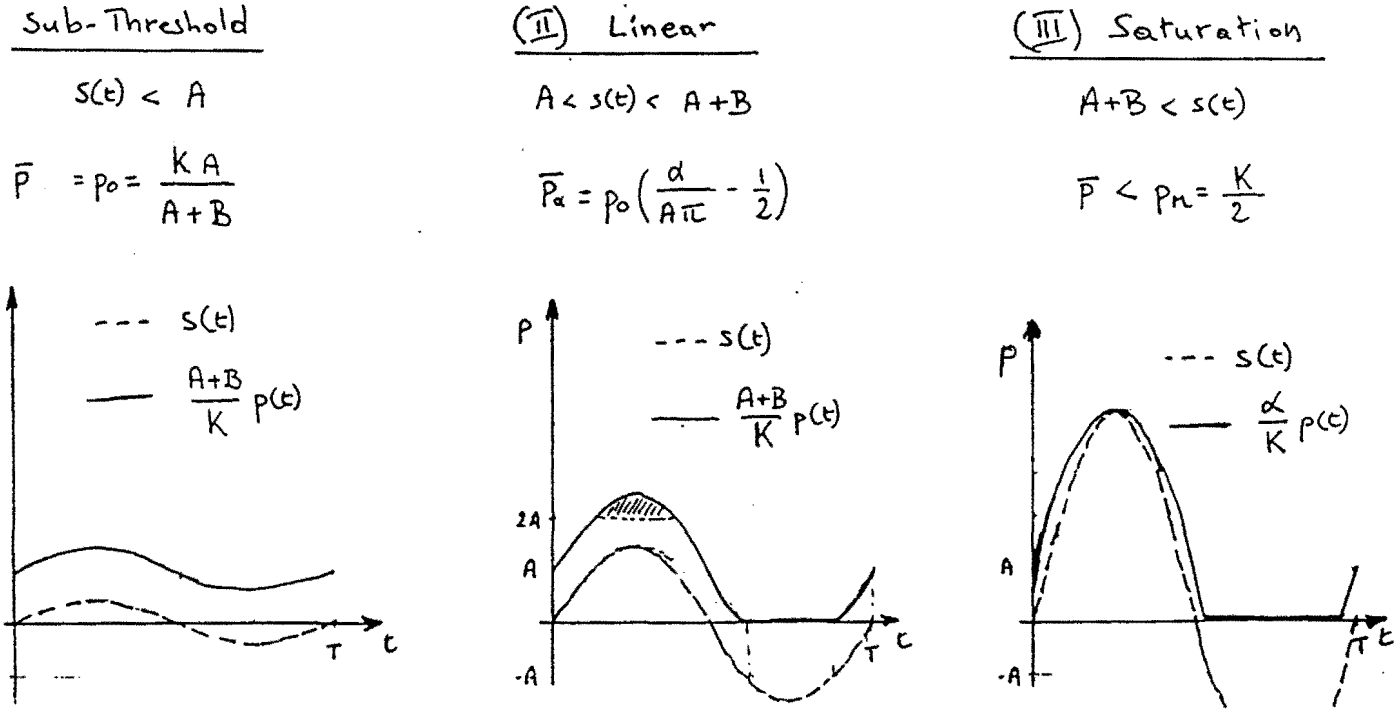
will be approximated by a 3-region piecewise linear function for further analysis. The subdivision is on the basis of the amplitude α of a sinusoidal input of any frequency and assumes, as in normal parametrizations, that $A << B$. The three conditions corresponding to each region can be described as "sub-threshold", "linear" and "saturation"(Fig.3.5). In sub-threshold and saturation regions the one period averages are easily obtained from the instantaneous values In the linear region the one period permeability average p_α is computed using following approximation:

$$p_\alpha = \frac{1}{2}\frac{K.A}{A+B} + \frac{K}{A+B}\frac{1}{2\pi}\left(\int_0^\pi \alpha \sin t dt + 2\int_0^{t_1} \alpha \sin t dt\right) \tag{3.5.a}$$

$$= \frac{K.A}{A+B}\left(\frac{1}{2} + \frac{\alpha}{\pi A} + \frac{A}{2\pi\alpha}\right) \tag{3.5.b}$$

$$\approx p_0\left(\frac{\alpha}{A\pi} + 1 - \frac{1}{\pi}\right) \tag{3.5.c}$$

The first two terms come from the positive phase of the input signal while the last (and smallest) term is a slight underestimate for the negative phase in which t_1 corresponds to the zerocrossing point on the permeability function. In the sub-threshold region average firing rate will not



Region	input	instantaneous p(t)	Avg Permeability(p_α)
Sub-Threshold	$\alpha \approx< A$	$p(t) = \frac{K}{A+B}A$	$p_0 = \frac{K.A}{A+B}$
Linear	$A << \alpha << A+B$	$p(t) = max\left(\frac{K}{A+B}(A+s(t)), 0\right)$	$p_\alpha = p_0(\frac{\alpha}{A\pi} + 1 - \frac{1}{\pi})$
Saturation	$A+B << \alpha$	$p(t) = K \quad (s(t) > 0)$	$p_{max} = \frac{K}{2}$

Figure 3.5: A 3 condition linear approximation of the input non-linearity

change with increased input amplitude but the firing will start to synchronize before threshold has been reached, what is conform physiological evidence.

For amplitudes of the order of $A + B$ nor the linear nor the saturation rates are good approximations. With some mathematical manipulation it is possible, however, to derive a single formula which is consistent with the approximations in both regions and which equals p_0 for $\alpha = A$ providing continuity with the sub-threshold region:

$$p_\alpha = \frac{K.(A\pi + \alpha)}{(A+B)\pi + 2\alpha} \tag{3.6.a}$$

$$= \frac{KA}{A+B}\left(\frac{1 + \frac{\alpha}{\pi A} - \frac{1}{\pi}}{1 + \frac{2\alpha}{\pi(A+B)}}\right) \tag{3.6.b}$$

The inverse of the above formula is given by:

$$\alpha = \pi\frac{(A+B)p_\alpha - KA(1-1/\pi)}{K - 2p_\alpha} \tag{3.7}$$

3.3.4 Steady State Properties

The steady state parameters thresholds, dynamic range, average firing rate, etc. are the easiest to analyze. Once transients have died out integration of the differential equations over a single period should equal zero. Integrating (3.2) results in a set of time invariant equations with as new variables the one period averages of the original variables, such as:

$$\bar{q} = \frac{1}{T}\int_0^T q(t)dt$$

If we further approximate:

$$\bar{pq} \approx \bar{p}.\bar{q}$$

then we find steady states estimates for free pool and cleft contents and firing rate:

$$\bar{c} = \frac{Y.\bar{p}}{L.\bar{p} + (L+R).Y} \tag{3.8.a}$$

$$\bar{q} = \frac{(L+R).Y}{(L+R).Y + L.\bar{p}} \tag{3.8.b}$$

$$\bar{w} = \frac{R}{X}\bar{c} \tag{3.8.c}$$

$$\bar{f} = H.\bar{c} \tag{3.8.d}$$

yielding following practical relationships:

$$\bar{f} = \frac{H.Y.\bar{p}}{L.\bar{p} + (L+R).Y} \tag{3.9.a}$$

$$= \frac{H.Y/L}{1 + (L+R).Y/(L\bar{p})} \tag{3.9.b}$$

$$\bar{p} = \frac{(L+R).Y}{\frac{H.Y}{\bar{f}} - L} \tag{3.9.c}$$

Spontaneous Firing Rate. The spontaneous rate, f_0, is derived by setting $s(t)$ to 0 in the permeability equation:

$$p_0 = \frac{K.A}{A+B} \tag{3.10.a}$$

$$f_0 = \frac{H.Y.p_0}{L.p_0 + (L+R).Y} \tag{3.10.b}$$

$$= \frac{H.Y/L}{1 + Y.(L+R)/(L.p_0)} \tag{3.10.c}$$

Maximum Firing Rate. In saturation $L.\bar{p}$ is much larger than $(L+R).Y$ for standard parametrizations, yielding as maximum average firing rate:

$$f_M = \frac{H.Y/L}{1 + (L+R).Y/(L.p_{max})}. \tag{3.11.a}$$

$$\approx \frac{H.Y}{L} \tag{3.11.b}$$

$$= \left(1 + \frac{L+R}{Lp_0}Y\right) f_0 \tag{3.11.c}$$

Rate Dynamic Range. The Firing Rate Dynamic range (δ) is easily derived from maximum and spontaneous rates:

$$\delta = \frac{f_M - f_0}{f_0} = \left(\frac{L+R}{L}\right)\frac{Y}{p_0} \tag{3.12}$$

The introduction of the parameters f_M and δ allows for rewriting the steady state rate equation in a more compact form:

$$\bar{f} = \frac{f_M}{1+\delta\left(\frac{p_0}{\bar{p}}\right)} \tag{3.13}$$

and taking derivatives of both sides of this equation lets us relate small changes in average input permeability to small changes in average firing rate. After some manipulation we can derive:

$$\Delta f = \frac{\delta p_0 f_M}{(p+\delta p_0)^2}\Delta p \tag{3.14.a}$$

$$\frac{\Delta p}{p_0} = \left(\frac{p}{\delta p}+1\right)^2\frac{\Delta f}{f_M} \tag{3.14.b}$$

Input Dynamic Range. The spontaneous and maximum firing rates f_0 and f_M should now be related to threshold and saturation level on the input. The response will in practice only reach f_0 and f_M at very small and very large input levels and not deviate much from them over a large range. Therefore threshold and saturation are ill defined measures. Here we will define them as the levels where a 5% deviation from the minimum and maximum firing rates is reached. For relating permeability to input amplitude the global approximation (3.7) can be used.

Threshold: From (3.14) the threshold permeability is found:

$$\frac{p_T}{p_0} = 1 + 0.05(1+1/\delta) \tag{3.15}$$

From which, by using (3.5):

$$\frac{\alpha_T}{A} \approx 1 + 0.05\pi(1+1/\delta) \tag{3.16}$$

Saturation: The linearization procedure can not be used for estimation of the saturation input level. First of all there is a small but relevant overestimate on the peak firing rate in (3.11) and the peak average permeability $\frac{K}{2}$ occurs for infinite input amplitudes. Therefore infinitesimal approximations can not be valid here. If we assume saturation to occur at a fraction $(1-\gamma)$ of the true maximum firing rate then for small γ and from (3.9):

$$1 + \frac{(L+R)Y}{Lp_S} = (1+\gamma)(1+\frac{(L+R)Y}{Lp_{max}} \tag{3.17.a}$$

$$p_S = \frac{\frac{(L+R)Y}{L}p_{max}}{\gamma p_{max} + (\gamma+1)\frac{(L+R)Y}{L}} \tag{3.17.b}$$

$$= p_{max}\frac{1}{1+\gamma(1+\frac{p_{max}L}{(L+R)Y})} \tag{3.17.c}$$

Now, for large α the input non-linearity can be approximated by:

$$p_\alpha = p_{max}\left(\frac{1}{1+\frac{\pi(A+B)}{2\alpha}}\right)$$

From combining the two previous equations we ultimately derive:

$$\frac{\alpha_S}{A} = \frac{\pi}{\gamma}\frac{(L+R)Y}{p_0 L} \tag{3.18}$$

As the threshold is very close to A the second hand side of this equation is also a very good approximation of the input dynamic range. BUT !! this latter equation has a fixed relation to the firing rate dynamic range, which means that input and firing rate dynamic range can NOT be set separately in the Meddis model. This is one of the major weaknesses which has been discovered in it.

3.3.5 Linearization of the Meddis Model

Linearization. The basic strongly non-linear model can be replaced by one of two much simpler linear derivatives for most analysis purposes. As with the analysis of the Schroeder-Hall model it is convenient to distinguish two greatly different time-scales for the analysis of periodic signals.

Envelope Analysis(SLOW): The differential equations are solved for period averages. This way $p(t)$ becomes a constant for steady periodic inputs and the differential equations become linear.

- **Within Cycle(FAST):** This behaviour must be superimposed on the previous one and for a single period $q(t)$ and $w(t)$ will be treated as constants. The dynamics of $p(t)$ are therefore directly reflected in $c(t)$.

"Slow" Analysis Model On this time scale we neglect the very fast variations of all variables, and do consider their global averages. We also do so with the input $p(t)$ which is replaced by its periodic average $\bar{p}$, which is a constant during any constant amplitude period input. This way we are able to eliminate the nonlinearities in the D.E.'s.

$$\dot{q}(t) = Y.(1-q(t)) + X.w(t) - \bar{p}.q(t) \tag{3.19.a}$$

$$\dot{c}(t) = \bar{p}.q(t) - (L+R).c(t) \tag{3.19.b}$$

$$\dot{w}(t) = R.c(t) - X.w(t) \tag{3.19.c}$$

After Laplace Transformation we get:

$$s.q(s) = Y - (\bar{p}+Y)q(s) + X.w(s) \tag{3.20.a}$$

$$s.c(s) = \bar{p}.q(s) - (L+R).c(s) \tag{3.20.b}$$

$$s.w(s) = R.c(s) - X.w(s) \tag{3.20.c}$$

Yielding the closed loop system:

$$q(s) = \frac{1}{s+(Y+\bar{p})}(Y + X.w(s)) \tag{3.21.a}$$

$$c(s) = \frac{1}{s+(L+R)}\bar{p}q(s) \tag{3.21.b}$$

$$w(s) = \frac{R}{s+X}c(s) \tag{3.21.c}$$

The time constants considered in this analysis are by definition significantly greater than the inverse of the stimulus frequency. In typical parametrizations $L + R$ will be large considered to all slow time constants which leads to a further simplification and following expression for the cleft contents:

$$c(s) = \frac{(s + X)\bar{p}.Y/(L + R)}{s^2 + (Y + X + \bar{p}).s + \bar{p}.L.X/(L + R)} \tag{3.22}$$

From the latter equation the two time constants underlying rapid and short-term adaptation can be derived.

"Fast" Analysis Model For this approximation it is acceptable to consider slow moving parameters such as q (and w) as constants, which allows us to rewrite the equations as:

$$\dot{c}(t) = q.p(t) - (L + R).c(t) \tag{3.23.a}$$

$$c(s) = \frac{q}{s + (L + R)}p(s) \tag{3.23.b}$$

$$\Delta q \approx \{-(\bar{p} + Y).q + X.w + Y\}.\Delta T \tag{3.23.c}$$

Validity of the above equations requires that the integrated one-cycle depletion of the temporary pool Δq is small compared to q. Maximum depletion occurs when the system was initially at rest and a maximum stimulus is produced. Under these circumstances q was originally almost 1.0 and p equals during one half period $p_{max} = K$, hence maximum depletion is:

$$max\Delta q = (K/2 - Y).\Delta T \tag{3.24}$$

From this it is possible to check when slow or fast analysis models will be valid.

3.3.6 Dynamic Behaviour

For analysis of the dynamic behaviour (onset and offset response) we fall back on the slow and fast analysis models. Parameters to be derived are adaptation time constants and overshoots.

The different time constants in the Meddis model can be localized in the system, which was also the motivation for the model simplifications:

- **Phase Locking** is generated by the halfwave rectification in the permeability function. A requirement, however, is that the rate of dissipation in the synaptic cleft is faster than stimulus frequency.
- **Short term adaptation** is mainly influenced by the dynamics of the free pool, and to some extend the reprocessing pool. If the reservoir is well filled and the permeability suddenly puts the exit gate wide open then large instantaneous outputs can be generated. Gradually the free pool depletes and steady state behaviour is reached.
- **Rapid Adaption** is quite harder to analyse since it isn't built in in any specific way, but rather a consequence of the two previous effects combined.

Phase locking is determined from the fast model. The shortest time constant in the system is $(L + R)^{-1}$. Some phase locking (synchronization) will occur for frequencies up to $(L + R)$. Clearly observable phase locking will stop considerable earlier, with as reasonable estimate:

$$f_{SY} < \frac{L + R}{2}$$

Short-term and rapid adaptation time constants are derived from the slow model. It is a combined effect of depletion of the free transmitter pool, especially by lower stimulus amplitudes,

and by replenishment thru the reprocessing store. Rapid adaptation occurs mainly thru re(de)-plenishment of the free transmitter pool with as approximate time constant $(Y + p(t))^{-1}$. For large amplitudes the two time constants will differ considerably and for most parametrizations they can easily be found from the closed loop equation (3.22):

$$\tau_{ST} = \frac{L + R}{L.X}$$

and

$$\tau_{RA} = \frac{1}{p_{max}} = \frac{2}{K}$$

For moderate amplitudes the two time constants are closer together and their computation cannot easily be separated. The time constants should be computed as the real poles, in function of $\bar{p}$, from (3.22).

Predicting overshoots is one of the thoughest aspects in a formal mathematical analysis. No solid derivations were possible, therefore one should rely on empirical evidence.

3.3.7 Summary of Design Parameters

The usefulness of the above design formulas is illustrated at the hand of the baseline model in [17]. There is barely a significant difference between the predicted and measured values.

parameter	expression	values from [17]	predicted values
Input Dynamic Range	$20log\left(\frac{20\pi(L+R)Y}{p_0 L}\right)$	25 dB	29dB
Firing Rate Dynamic Range	$\delta = \left(\frac{L+R}{L}\right)\left(\frac{A+B}{A}\right)\frac{Y}{K}$	0.55	0.56
Maximum Firing Rate	$f_M = \frac{H.Y}{L}$	99	101
Synchrony	$f_{SY} < (L+R)/2$	-	4500
Time Constants (+20dB)	roots from quad. eq.	75 & 7.7 msec	78 & 5.1 msec
Time Constants (+50dB)	$\frac{L+R}{LX}$ and $\frac{2}{K}$	57 & 1.2 msec	55 & 1.2 msec

3.3.8 Adaptation Examples for Sinusoidal Bursts

The combined behaviour of filterbank and adaptation is illustrated for two test stimuli. Both test stimuli consist of a sequence of 9 sinusoidal bursts (1kHz) of increasing amplitudes (6dB steps), with amplitude ranges from 30dB SPL to 86 dB SPL. The onset and offset amplitude ramps are always 2msec long. The first stimulus(B1K) consists of 50 msecs bursts with 50 msecs silence between each of them, while the second one (S1K) has no silence uses 100msec bursts with no silences. Firing probability in a few channels with CF around 1kHz is shown for 1 kHz tone bursts in Fig.3.6 for a Schroeder Hall Model and in Fig.3.7 for a Meddis model with default parameters. Using the above derived mathematical properties and relationships it is possible to change one or several of the parameters in a guided way.

Chapter 4

Post Processing in Auditory Models

4.1 Introduction

The output of a physiologically based model is a spike train on the auditory nerve or firing probability. However the data rate from such a model is too large for further processing by e.g. a speech recognition system. Therefore it is necessary to make some form of abstraction of this neural spike train and the most common methods are representations of average firing rates (similar to average firing probability) or some form of synchrony measure at a low sampling rate.

It should be stressed that *physiological* understanding of what happens beyond the first synapses of the auditory nerve is limited and that none of the *post-processing* described in this chapter has a sound *physiological* motivation. Strictly speaking the auditory model stops at the nerve spike train, the algorithms developed in this chapter describe ways of *looking at the output* of it.

4.2 Average Rate

Average rate is the easiest representation of a neural spike train. In practice it isn't even necessary to compute a spike train, as average rate can be determined from firing probability. As we early on took the approach that a single channel in the model stands for a local group of fibers, statistical effects, except maybe refractory periods, are averaged out by this grouping and average rate is determined directly from firing probability. For the sake of data reduction downsampling can be used after this smoothing operation.

More formally average firing probability is computed as

$$\bar{f}(t) = w(t) * f(t) = \int_0^\infty w(\tau) f(t-\tau) d\tau \tag{4.1}$$

in which $w(t)$ is a properly chosen smoothing window. For sake of normalization we will require that $w(t)$ has following property:

$$\int_0^\infty w(\tau) d\tau = 1$$

Often used smoothing windows are first and second order leaky integrators.

$$w_1(t) = \frac{1}{T} e^{-\frac{t}{T}} \quad t > 0 \tag{4.2.a}$$

$$w_2(t) = w_1(t) * w_1(t) \tag{4.2.b}$$

The first window is a first order leaky integrator with effective window length T. The second one, which is obtained by twice applying the first one, is somewhat similar to a Hamming window

of length $4T$ [2]. Fig 3.3 show examples using a second order leaky integrator with time constant $T = 1msec$.

Recursive Implementation The exponential window can very efficiently be implemented in discrete arithmetic as a first order recursion using current estimate and the new input sample.

$$\bar{f}(i) = \frac{\Delta T}{T}\sum_{0}^{\infty} e^{\frac{-k\Delta T}{T}} f(i-k) \tag{4.3.a}$$

$$= \frac{\delta T}{T}\left(f(i) + e^{\frac{-\Delta T}{T}} f(i-1) e^{\frac{-2\Delta T}{T}} f(i-2) + \ldots\right) \tag{4.3.b}$$

$$= \frac{\delta T}{T} f(i) + e^{\frac{-\Delta T}{T}} \bar{f}(i-1) \tag{4.3.c}$$

$$\approx \frac{\Delta T}{T} f(i) + \left(1 - \frac{\Delta T}{T}\right) \bar{f}(i-1) \tag{4.3.d}$$

4.3 Synchrony Measures

Due to early rate saturation spectral resolution on average rate representations of high intensity inputs is very low. Detailed information about the stimulus signal seems to be preserved up to much higher intensities by phase locking properties (for nerve fibers with characteristic frequencies below about 2kHz). These phase locking properties have long been understood for pure tones [18]. The potential relevance to speech processing was first illustrated by the experiments of Sachs and Young[19, 20] in which they showed that the formant structure of medium to high intensity vowels is not preserved in average firing rates but in a clearer way in some form of synchrony measure applied to auditory nerve spike train. These results, and other similar ones, have convinced many researchers that the auditory system must perform some type of synchrony analysis. How the system might actually perform such an analysis has not been shown, nor is there any real evidence that the auditory system uses synchrony in one way or another.

There is also evidence that the auditory system might not need synchrony at all and that rate be a sufficient representation. Delgutte[21] showed that formant structure is well preserved in rate patterns at onsets and offsets of vowels. Hence rate would be sufficient if the higher pathways cue in on transients and pay little attention to steady state situations.

While average firing rate is a simple measure and reasonably well defined, there is no agreement within the scientific community as how synchrony could best be computed. Two approaches must be distinguished. In the first one synchrony is computed with respect to a predefined frequency. This implies that some form of physiological clock is involved in the measurement. The likelihood of some mechanism existing is up to debate, but synchronization to the characteristic frequency of a fiber is plausible because this is anyhow by for the strongest component in the typical output signal. In a second approach any fiber can synchronize to almost any frequency, hence the strength of a formant in a vowel e.g. will not only be presented by the stength in the fiber at CF but also by activity in fibers with CFs around this value.

4.3.1 Synchrony Measures for known Characteristic Frequencies

Generalized Synchrony Detector The Generalized Synchrony Detector (GSD) is determined from following equation[2] :

$$y(t) = G\frac{2}{\pi} atan\left(\frac{1}{A}\frac{< u+v > -2f_0}{< u-v > +\epsilon}\right) \tag{4.4}$$

in which

$<>$ is a smoother, using double leaky integration

$T = 1/CF$, CF is characteristic frequency of a fiber.

$u(t) = f(t)$ is the instantaneous firing probability

$v(t) = u(t - T)$: the signal, delayed by the characteristic period

f_0: the fiber's spontaneous rate.

ϵ: small number avoiding divide by 0 overflows.

A, G are scaling factors

For steady state analysis this measure reaches its minimum when correlation between $u(t)$ and $v(t)$ is zero, i.e. for a white noise input and its maximum when both are identical. It has however a most unusual behaviour around tone burst onsets. Depending on the choice of f_0 and ϵ the synchrony measure might take a deep drop. With $G = 1$, $y(t)$ lies in the range 0-1.

Modified GSD In order to alleviate the previously mentioned onset problem a slight modification to the GSD definition leads to a more sensible measure:

$$y(t) = G\frac{2}{\pi}atan\left(\frac{1}{A}\frac{< u + v > -2f_0}{< v - v_1 > +\epsilon}\right) \tag{4.5}$$

in which:

$u(t) = f(t)$

$v(t) = u(t - T)$

$v_1(t) = u(t - 2T)$

The onset problem isn't fully solved but it seems a reasonable 'hack'. One other, possibly much more important problem with this style of synchrony determination is its counteraction of auditory nerve adaptation. Synchrony during onsets will be represented as "bad" because $u(t)$ and $u(t - T)$ differ significantly. Intuitively it is hard to accept that the auditory system would first perform adaptation (to see transients more clearly ?) and in the next processing step would eliminate most of what adaptation has done !!??

Parameter Settings Both the GSD and MGSD are very sensitive to appropriate parameter settings. The a priori knowledge of the spontaneous firing rate is of key importance. An underestimate causes the output to be very smooth while an overestimate causes clearly clipping problems. A reasonable safe choice for ϵ is half the spontaneous firing rate.

4.3.2 Synchronization Index

A measure which is much less sensitive to adaptation effects is the synchronization index. In a first pass period histograms of firing (probabilities) are computed for the *known* frequency, i.e. most often the CF of a fiber. For a completely synchronized fiber all firing occurs during one half phase and none during the opposite phase, for a non synchronized fiber a period histogram is flat. Synchronization Index is a measure for the strength of synchronization going from 50% (not synchronized) till 100% (fully synchronized) [18]. With $P(t)$ representing the period histogram and T the histogram period the synchronization index is computed as:

$$SI = 100\frac{\int_0^{T/2} P(t)dt}{\int_0^T P(t)dt} \tag{4.6}$$

In the examples here a directly related measure is used spanning the range 0-100:

$$SI^* = 2\left(100\frac{\int_0^{T/2} P(t)dt}{\int_0^T P(t)dt} - 50\right) \tag{4.7}$$

Synchronization Index has most often been used in the analysis of tone burst, where a single measure is obtained over a full signal. In an auditory model the period histograms need to be computed with a forgetting factor in order to get a sequence of snapshot pictures. A rather small window can be used for this analysis leading to some, but much less obvious and non destructive counteracting of adaptation.

4.3.3 Predictive Synchrony Rate

Predictive synchrony rate is a novel hybrid rate/synchrony measure. It attempts to combine the advantages of synchrony measures and short term rate effects such that both steady state and onsets are clearly captured in a single measure. Predictive synchrony rate is defined as the average firing rate in which the weighting function is a one period normalized period histogram:

$$PSR = \int_0^T w(t)f(t)dt \tag{4.8}$$

$f(t)$ is the instantaneous firing rate and $w(t)$ the normalized period histogram. Noise robustness in synchrony measures relies on the usage of several periods which can easily be included in the smoothing of the period histogram, while instantaneous and fast adaption properties is maintained by keeping the integration time in (4.8) small, i.e. one period.

4.3.4 Examples of Noise Robustness of Synchrony Measures

At the hand of a set of tone bursts imbedded in a 50dB additive white noise disturbance some noise analysis of the different schemes is possible. Average rate, generalized synchrony detector and predictive synchrony rate are compared in Fig.4.4. Following observations are possible:

- Due to sharpness of the filters pure tone onsets and offsets are visible over a wide frequency range. This behaviour is not suppressed by synchrony processing.
- Due to smoothing average rate is unable to maintain proper rapid adaptation properties.
- The GSD suppresses the adaptation considerably.
- Predictive Synchrony Rate has the best overall characteristics.

4.4 Synchrony Measures with Interval Histograms

The principal method in defeating a priori knowledge of the synchronization frequency is the use of interval histograms. In this approach any channel can represent any frequency, though in practice contribution will obviously be most significant in the region around CF. One particular way of implementing such a synchrony measures are the level crossing histograms used by O. Ghitza[1] . Not neural spike trains but local firing probability is used as input signal. Events are then created whenever the firing probability in a channel crosses a number of levels covering the whole probability range. For each level interval histograms of events are computed and finally the different histograms are summed together.

Chapter 5

Software for Auditory Modeling

5.1 Introduction

This chapter gives a global introduction to the software which was developed at IPO that implements the algorithms and ideas described in the previous chapters. Apart from the strictly algorithmic programs a number of utilities had to be developed which allow for easy manipulation of multichannel files, these are also described here. The package as a whole is referenced as 'AMOD' (Auditory MODeling).

5.1.1 File Conventions

File Formats. All data files are conform the LVS/ILS data format. Multichannel files (ILS multiplexed files) are used for filterbank outputs and further processing.

Header Info. Header information is extensively used throughout the package. Using exclusively the package doesn't require understanding of single header entries. Detailed info can be obtained thru the help facility.

Energy Levels. Because neural adaptation is a non-linear process, one must define absolute energy levels. Both Schroeder-Hall and Meddis use the convention that a sinewave at 1kHz with rms=1 corresponds to a signal of 30 dB SPL. Within this software all amplitudes are multiplied by 4 with respect to this definition for optimization with respect of quantization errors and use of dynamic range.

5.1.2 VAX/VMS User Interface

Setup. The user should include following line in his LOGIN.COM file for easy access to programs and subroutine libraries.

```
$ @akofondisk:[compi.amod]init.com
```

Command Interface. By means of previous setup procedure it is possible to access the whole package by using a command format which is similar to the generic VAX/VMS system commands: i.e.

```
% CMD/q1=xxx/q2=yyy p1 p2 ..
```

Global Variables. One global variable is used in this package: DEBUG. DEBUG takes integer values and is by default equal to 0 (NO DEBUGGING). Progressively higher values will print out more detailed intermediate data, the highest value used is 32.

Help Facility. Full help facilities are available. Help files are available on all main programs by program name, a general introduction and overview of the programs is given in the AMOD entry, and the subroutine library is described in DSPLIB

5.2 Main Programs

The AMOD package contains of following main programs, a full help is available on line.

ADDNO: add noise

ADAPT: Schroeder Hall and Meddis Adaptation

DSAMPLE: Downsampling without filtering of multichannel file

ERBTST: prints out values from the ERB-scale

FDES: FIR Filter design program for auditory filterbanks (Gammatone, Flanagan, Hamming)

FILE: print out general file info on multichannel files or multichannel FIR filters

FIR: Multichannel FIR Filter

LCH: Level Crossing Histograms

MKNOIS: Make noise signal

PLOTMC: Multichannel Plot (sampled data and waterfall)

SELCHAN: Select one channel from a multichannel file

SMOOTH: Smoothing and downsampling

SUMCHAN: Sum all channels from a multichannel file together

SYNCD: Synchrony Detector (GSD,MGSD,PSR,SI)

TESTSIG: Create a testsignal (multilevel tone bursts)

5.3 Subroutine Library

A subroutine library with frequently used DSP routines is used and can also be used by new program developers. HELP DSPLIB gives more information on the available routines. A logical variable lnk$library is defined in the login script and points to the amod subroutine library, hence you must not explicitly mention this library when linking.

5.4 Code and Demos

5.4.1 The AMOD Directory

All code for the AMOD package, standard filterbank designs and scripts to generate the demos in this report are in subdirectories of akofondisk:[compi.amod]:

CLD.DIR: command language interface files

DEMO.DIR: Demonstration Files for combined auditory model

FLT.DIR: Filterbank designs and design scripts

HLP.DIR: help directory

INIT.COM: initialization file

LIB.DIR: subroutine library source and object code

MAIN.DIR: source and executables of main programs

5.4.2 Filter Design Illustrations

Following files extensions are used in the FLT.DIR filter design directory:

.FLT: filterbank parameters

.IR: multichannel impulse response

.CIR: combined impulse response (sum of channels in previous file)

.IRxx: single channel impulse response for channel xx

.SIG: testsignal, typically sampled at 20kHz

A number of command files are available in this directory that might be useful to illustrate other filterbank designs:

FRPLOT20	filtname	plot 20 channel frequency response
IRPLOT20	filtname	plot 20 channel frequency response
IR1	filtname	
FBKDES		Design script for all filterbanks used in this report

5.4.3 Demo Directory

Testsignals:

B1K.SIG: 50 msec 1kHz bursts with amplitudes from 30-86dB SPL with 50msec silences

S1K.SIG: 100 msec 1kHz bursts with amplitudes from 30-86dB SPL with no silences

Processed data is available in files with following extensions:

.BM: filterbank output (basilar membrane motion)

.SH: instantaneous firing probability according to Schroeder Hall model

.M2: instantaneous firing probability according to default Meddis model (NEURON=2)

.SHS,.M2S: smoothed and downsampled (4kHz) versions of the above, using two 1msec leaky integrators

.GSD: modified synchrony detector on the basis of .M2 file

.SI: synchrony index on basis of .M2 file

.PSR: predictive synchrony rate on basis of .M2 file

.LCH: level crossing histogram output on the basis of .M2 file

Bibliography

[1] **O. Ghitza.** Auditory nerve representation as a front-end for speech recognition in a noisy environment. *Computer Speech and Language*, 1(2):109–130, 1986.

[2] **S. Seneff .** *Pitch and Spectral Analysis of Speech Based on an Auditory Synchrony Model.* PhD thesis, M.I.T. Technical Report No. 504, 1985.

[3] **J. Cohen.** Application of an auditory model to speech recognition. In *Proc. 1986 DSP Workshop, Chatham*, 1986.

[4] **R.F. Lyon.** A Computational Model of Filtering, Detection and Compression in the Cochlea. In *Int. Conf. Acoust. Speech & Signal Processing*, 1982.

[5] **J.B. Allen.** Cochlear Modeling. *IEEE ASSP Magazine*, Jan. 1985.

[6] **G. von Békésy.** *Experiments in Hearing.* Wiley, New York, 1960.

[7] **B.C.J. Moore and B.R. Glasberg .** Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.*, 74:750–753, 1983.

[8] **P.I.M. Johannesma .** *The pre-response stimulus ensemble of neurons in the cochlear nucleus.* in *Hearing Theory*, IPO, Eindhoven, 1972.

[9] **J.L. Flanagan.** *Speech Analysis Synthesis and Perception.* Springer-Verlag, 1972.

[10] **B.C.J. Moore and B.R. Glasberg .** Unknown. *J. Acoust. Soc. Am.*, 0:0, 1989.

[11] **Patterson et al. .** *An Introduction to Auditory Sensation Processing.* Technical Report, Medical Research Council, 1990.

[12] **E. de Boer and C. Kruidenier .** On ringing limits on the auditory periphery. *Biol. Cybern.*, 63:433–442, 1990.

[13] **M. Schroeder and J. Hall .** Model for mechanical to neural transduction in the auditory receptor. *J. Acoust. Soc. Am.*, 55(5):1050–1060, 1974.

[14] **M.L. Brachman.** *Dynamic Response characteristics of Single Auditory-nerve Fibers.* PhD thesis, Institute for Sensory Research, Syracuse University, 1980. Special Report ISR-S-19.

[15] **R. Meddis .** Simulation of mechanical to neural transduction in the auditory receptor. *J. Acoust. Soc. Am.*, 79(3):702–711, 1986.

[16] **R. Meddis .** Simulation of auditory-neural transduction: further studies. *J. Acoust. Soc. Am.*, 83(3):1056–1063, 1988.

[17] **R. Meddis, et al. .** Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse. *J. Acoust. Soc. Am.*, 87(4):1813–1817, 1990.

[18] **J.E. Rose et. al.** . Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiology*, 30:767–793, 1967.

[19] **M.B. Sachs and E.D. Young.** Encoding of steady state vowels in the auditory nerve: representation in terms of discharge rate. *J. Acoust. Soc. Am.*, 66:470–479, 1979.

[20] **E.D. Young and M.B. Sachs.** Representation of steady state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. *J. Acoust. Soc. Am.*, 66:1381–1403, 1979.

[21] **B. Delgutte** . Representation of Speech-like sounds in the discharge patterns of auditory nerve fibers . *J. Acoust. Soc. Am.*, 68(3):842–857, 1980.

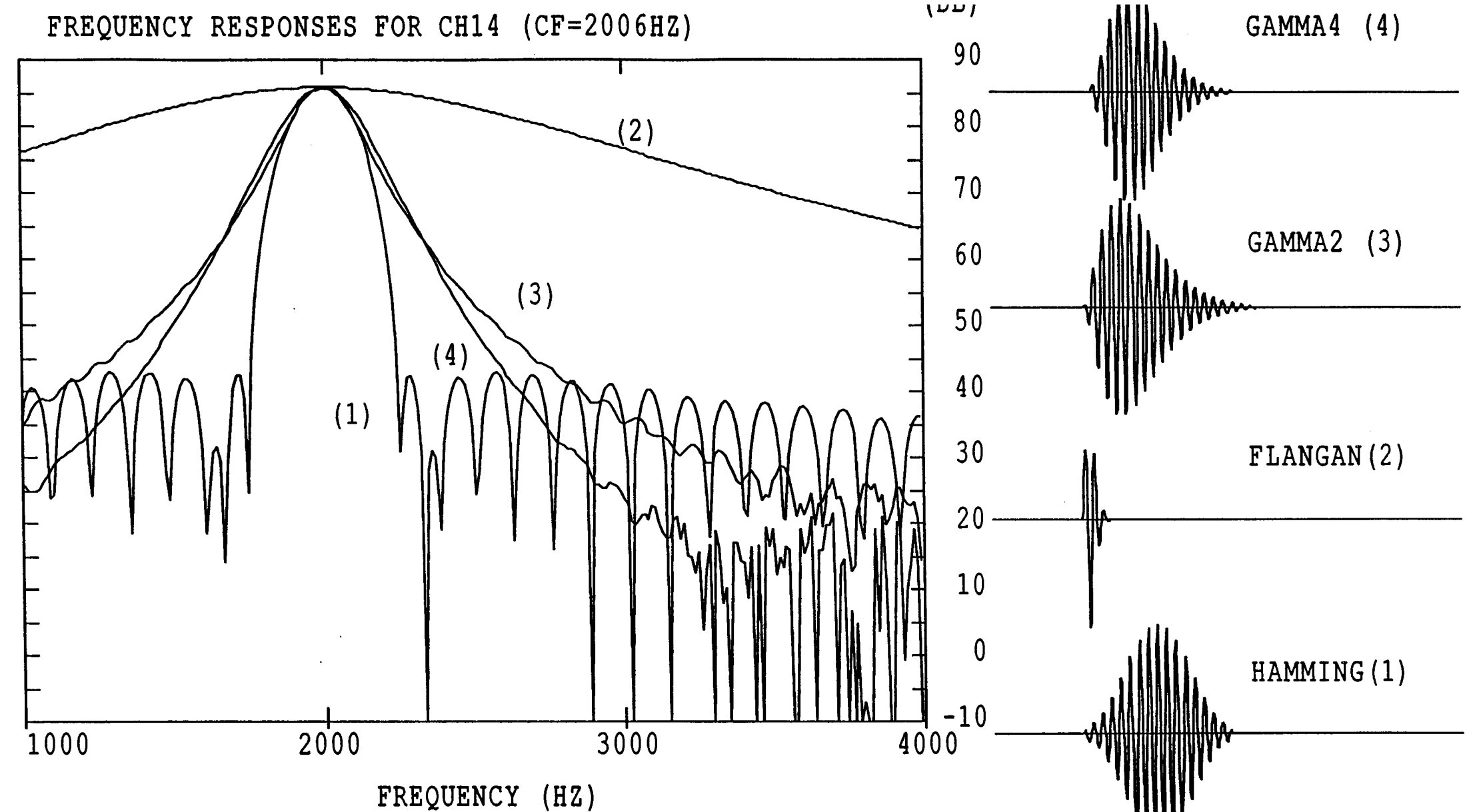

FREQUENCY RESPONSES FOR CH14 (CF=2006HZ)
(DB)
90
80
70
60
50
40
30
20
10
0
-10
(1)
(2)
(3)
(4)
1000
2000
3000
4000
FREQUENCY (HZ)
GAMMA4 (4)
GAMMA2 (3)
FLANGAN(2)
HAMMING(1)

Fig. 2.3

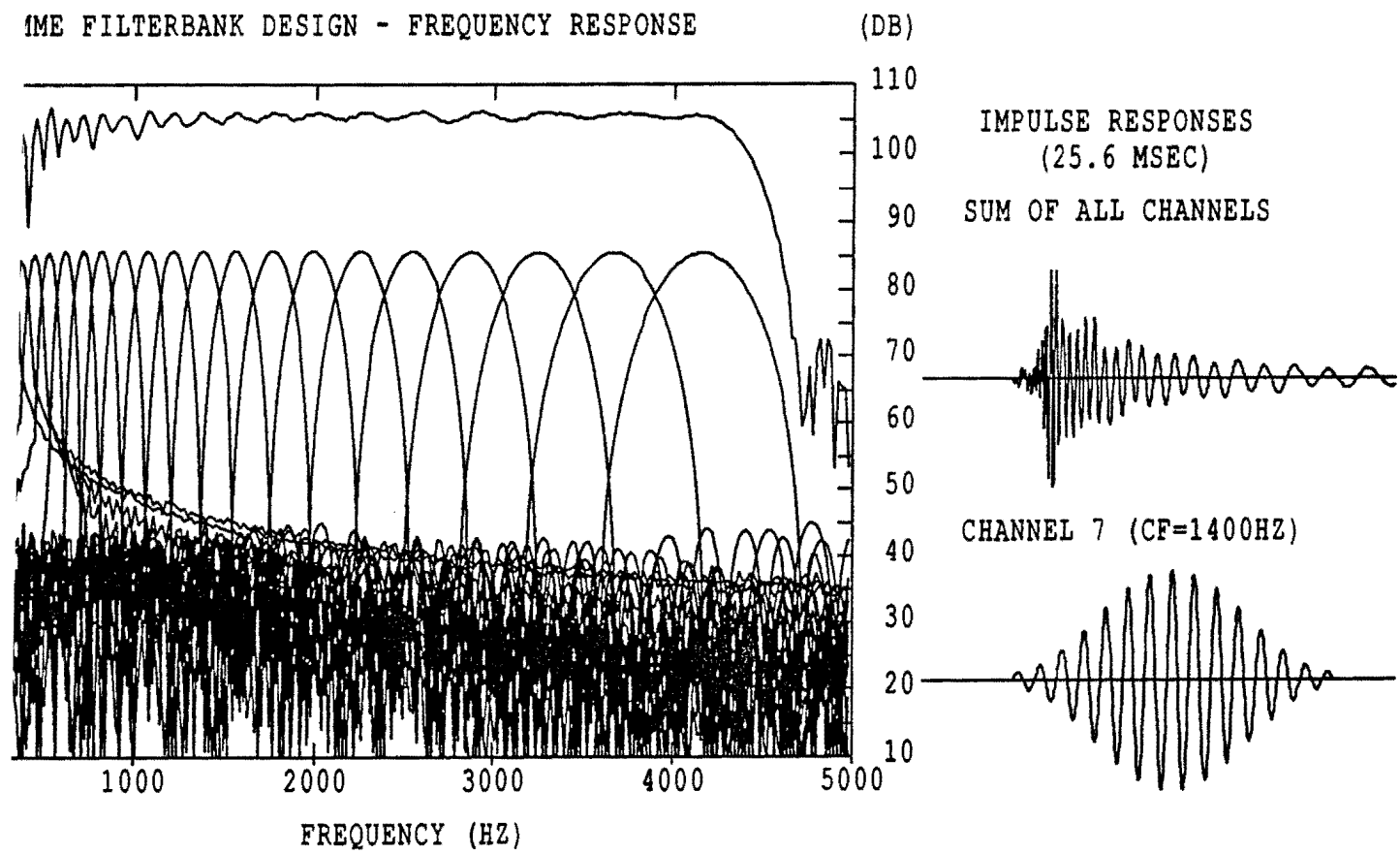


Fig. 2.4a

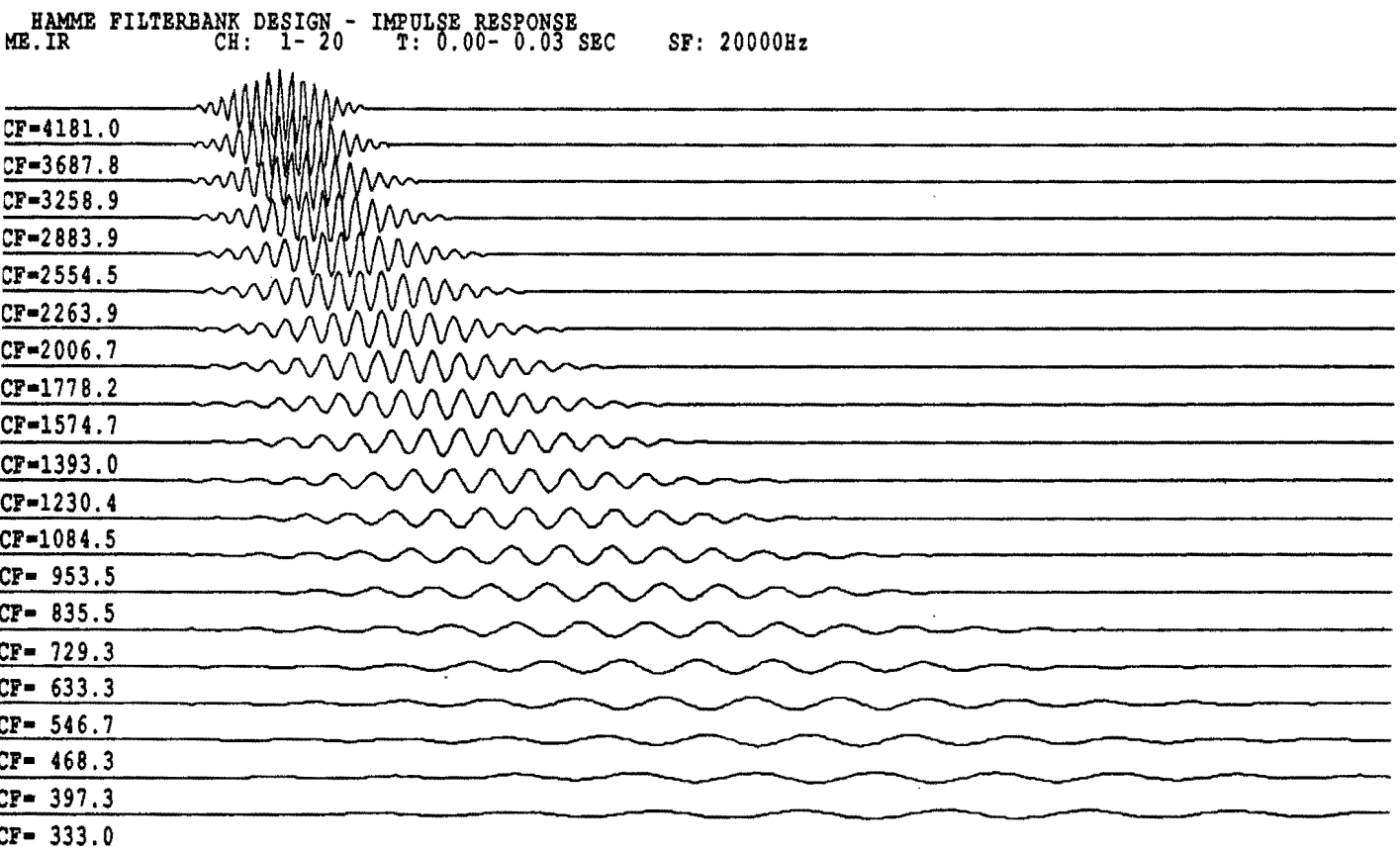


Fig. 2.5a

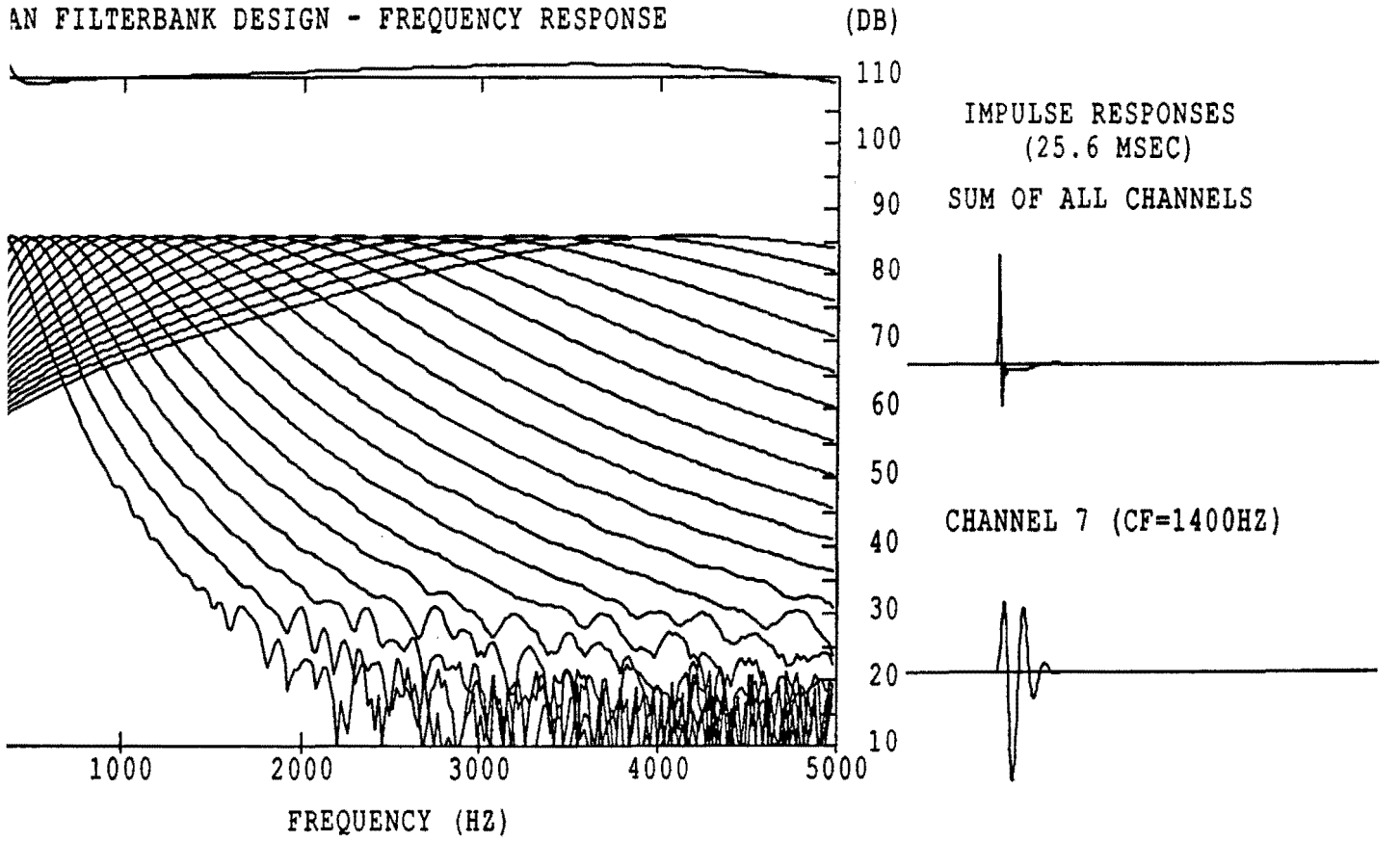


Fig 2.4b

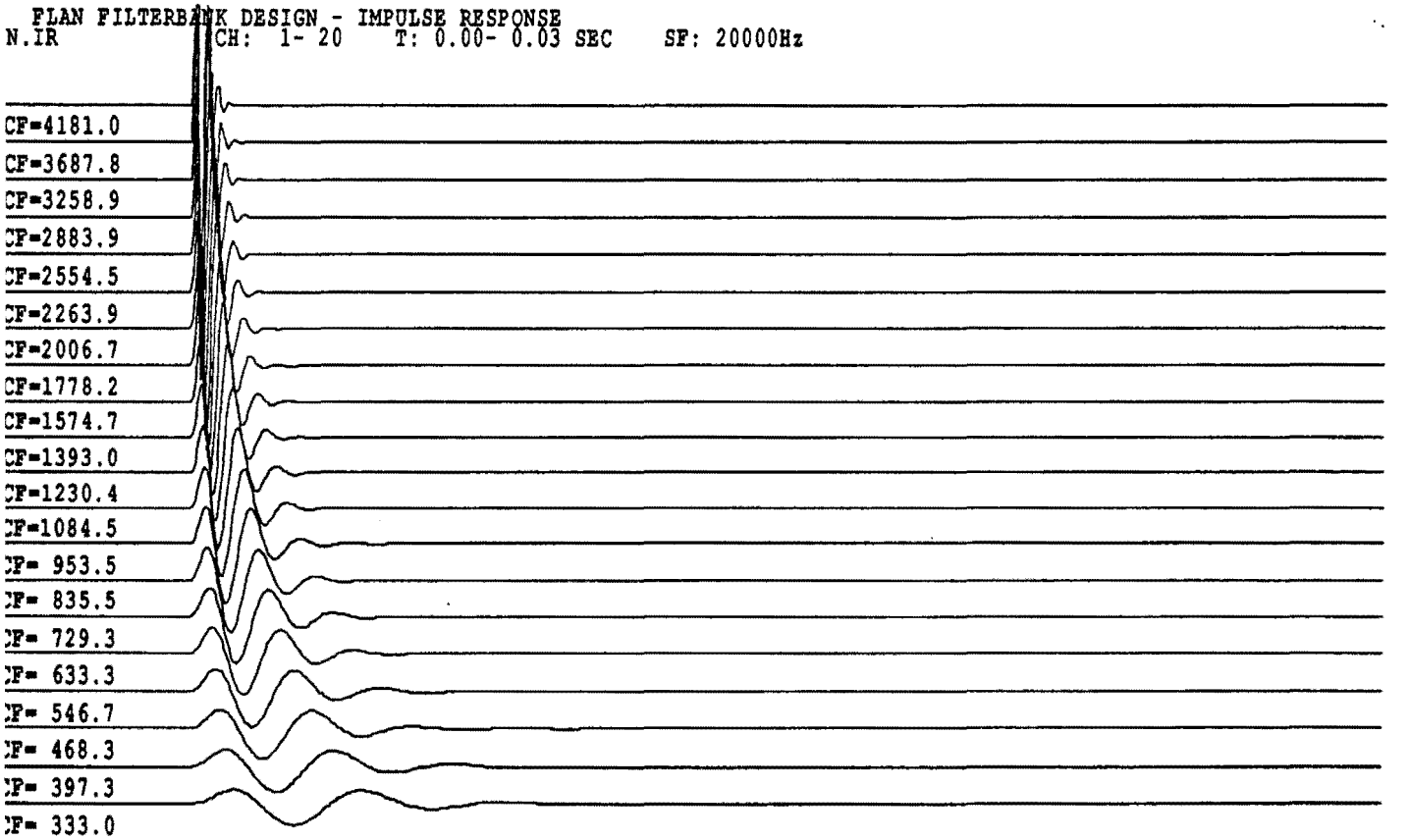


Fig 2.5b

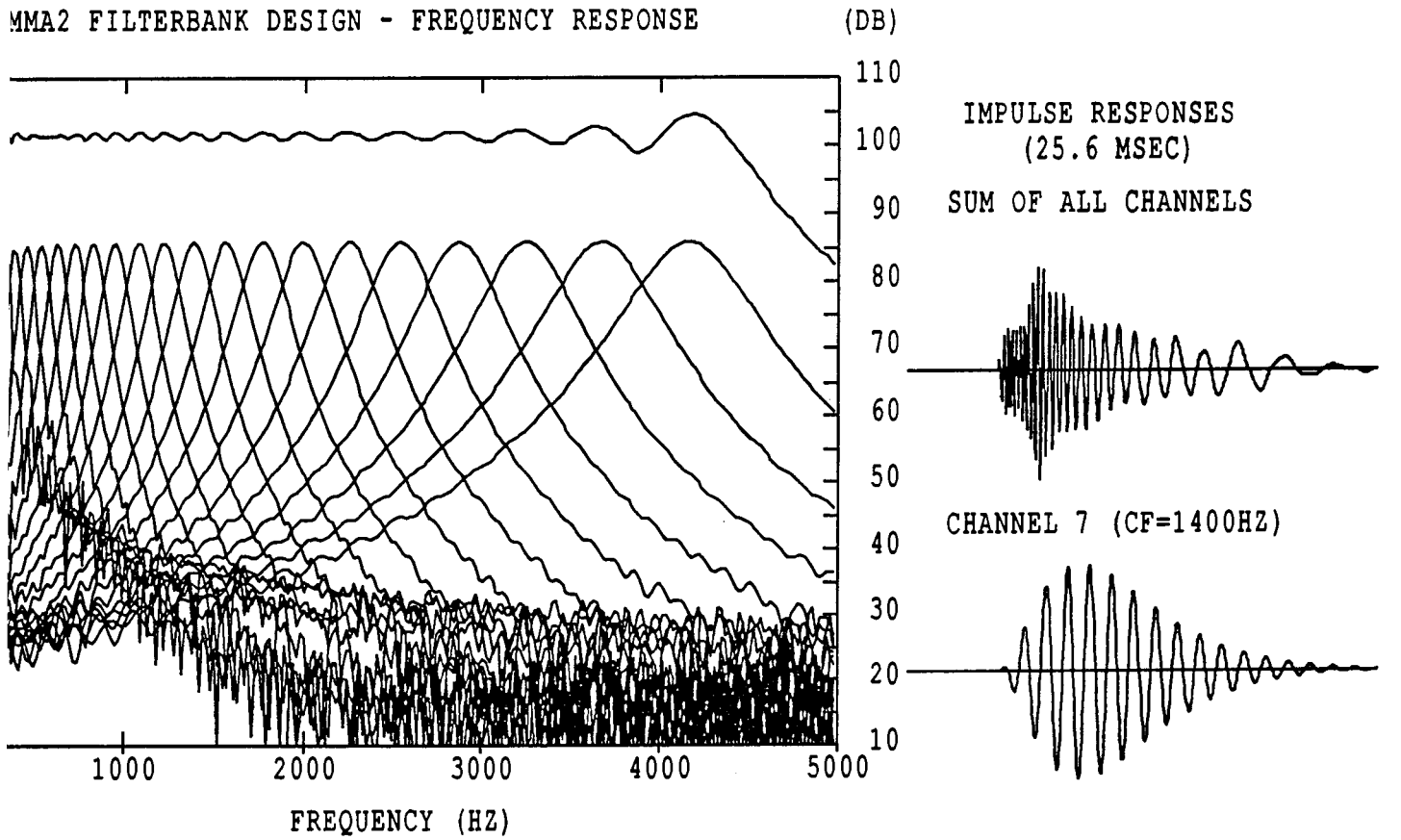


Fig. 2.4c

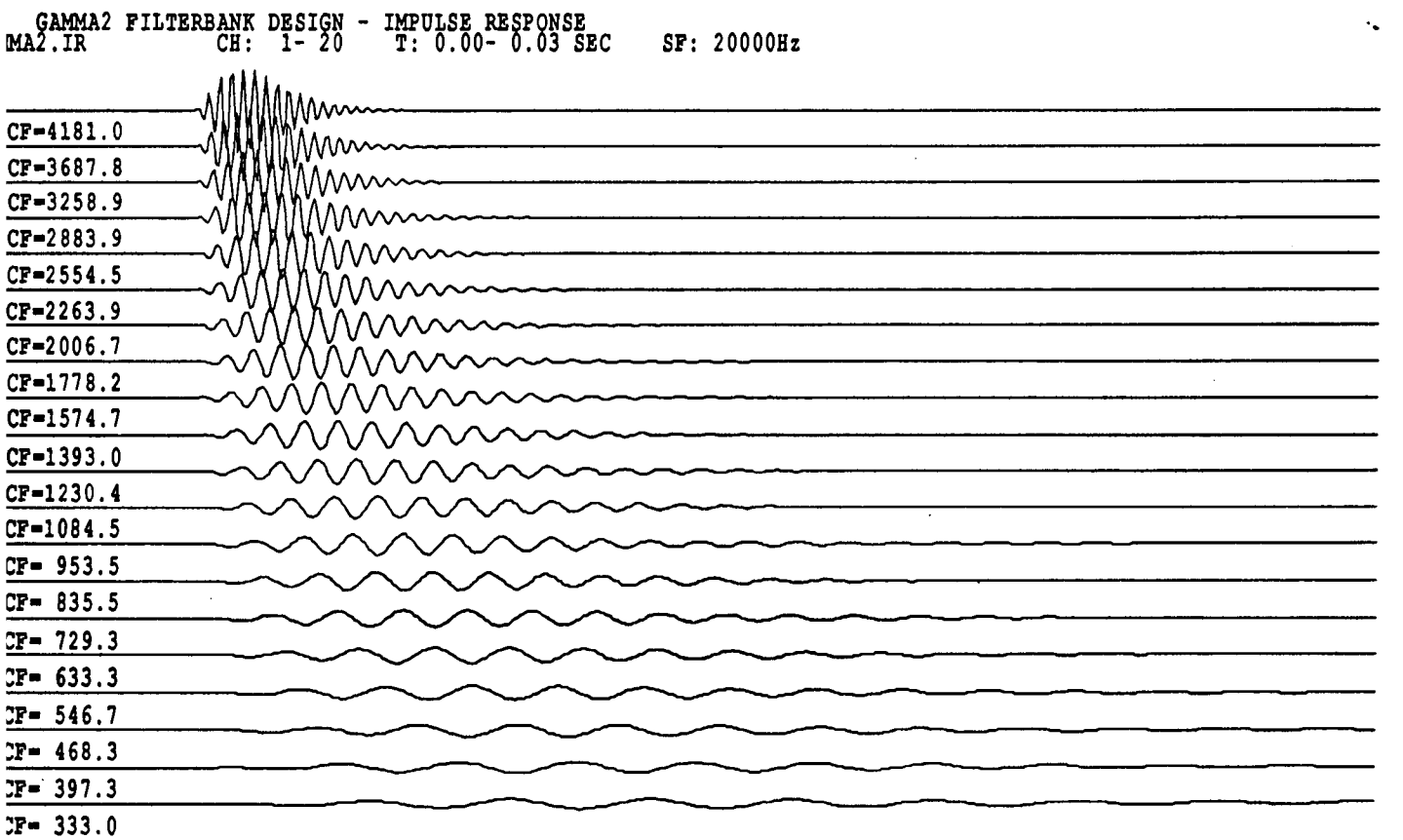


Fig 2.5c

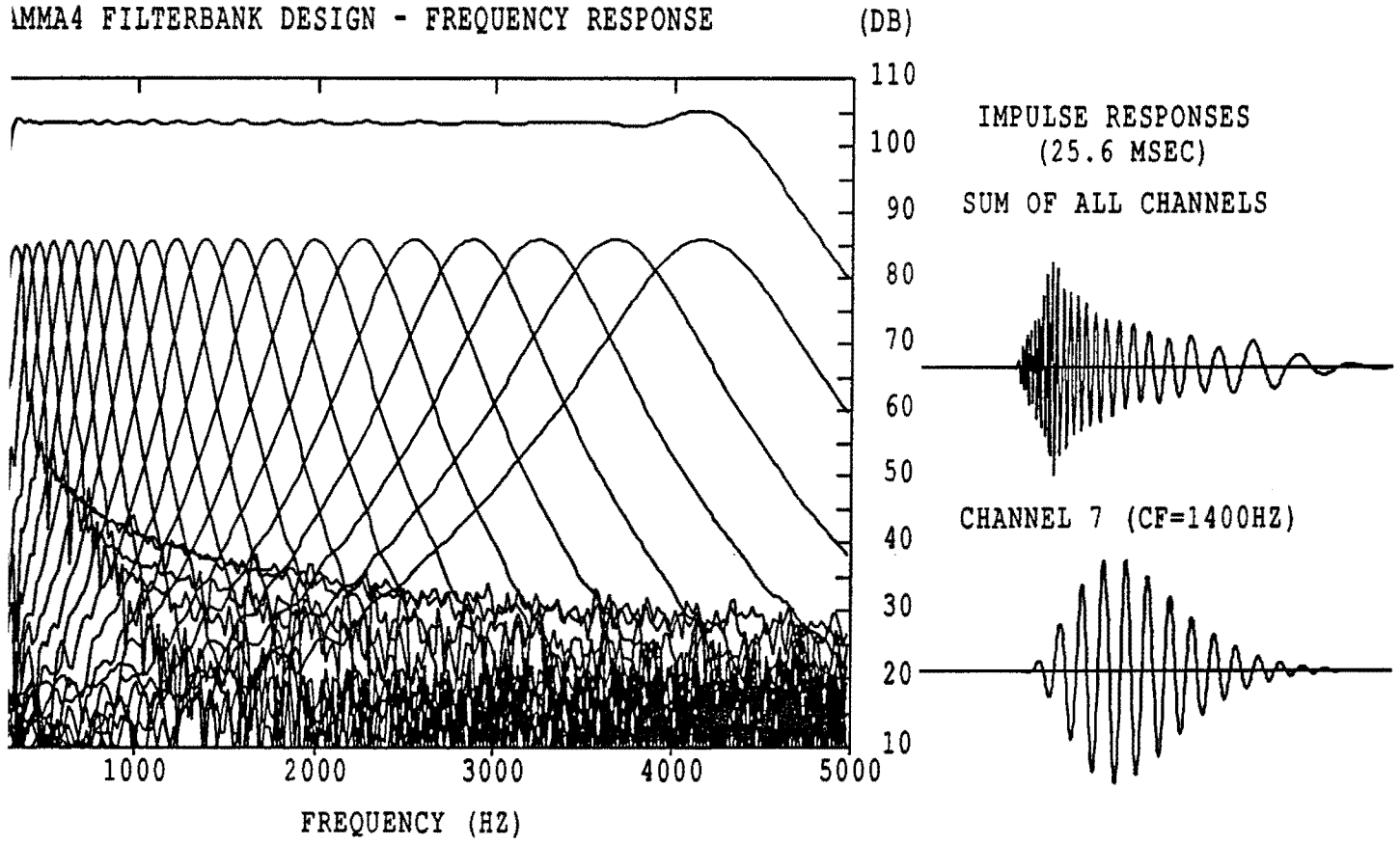


Fig 2.4d

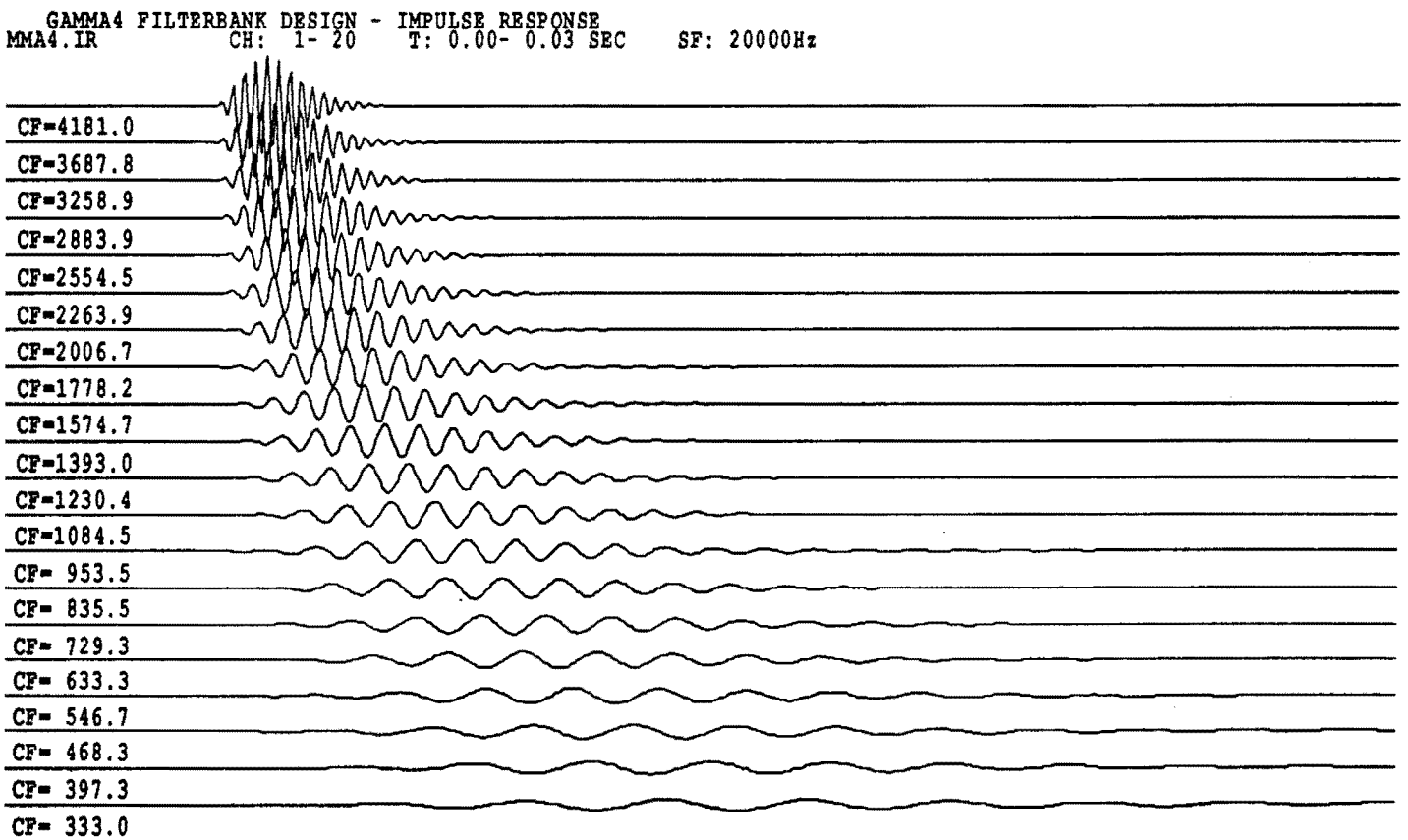


Fig. 2.5d

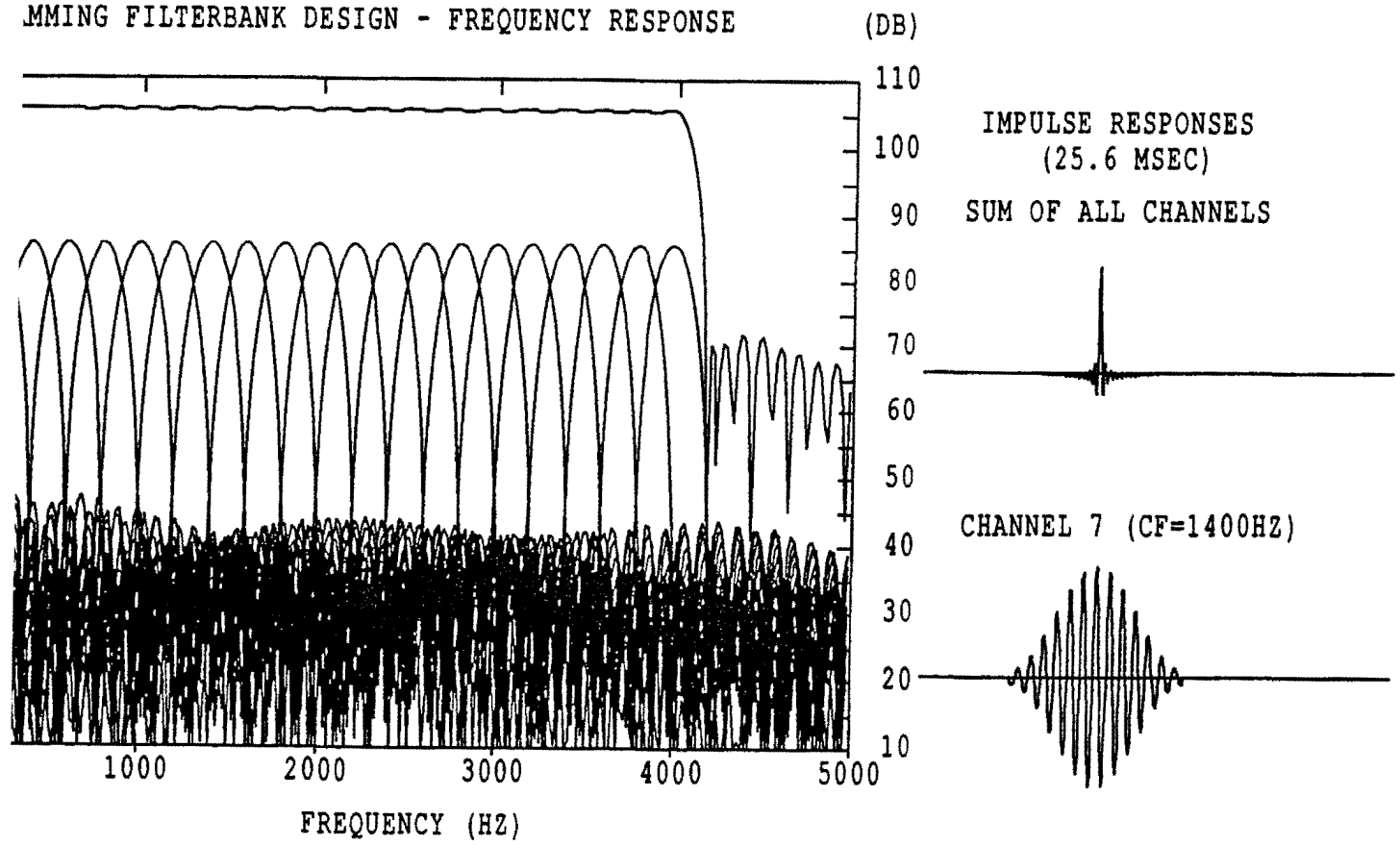


Fig 2.4e

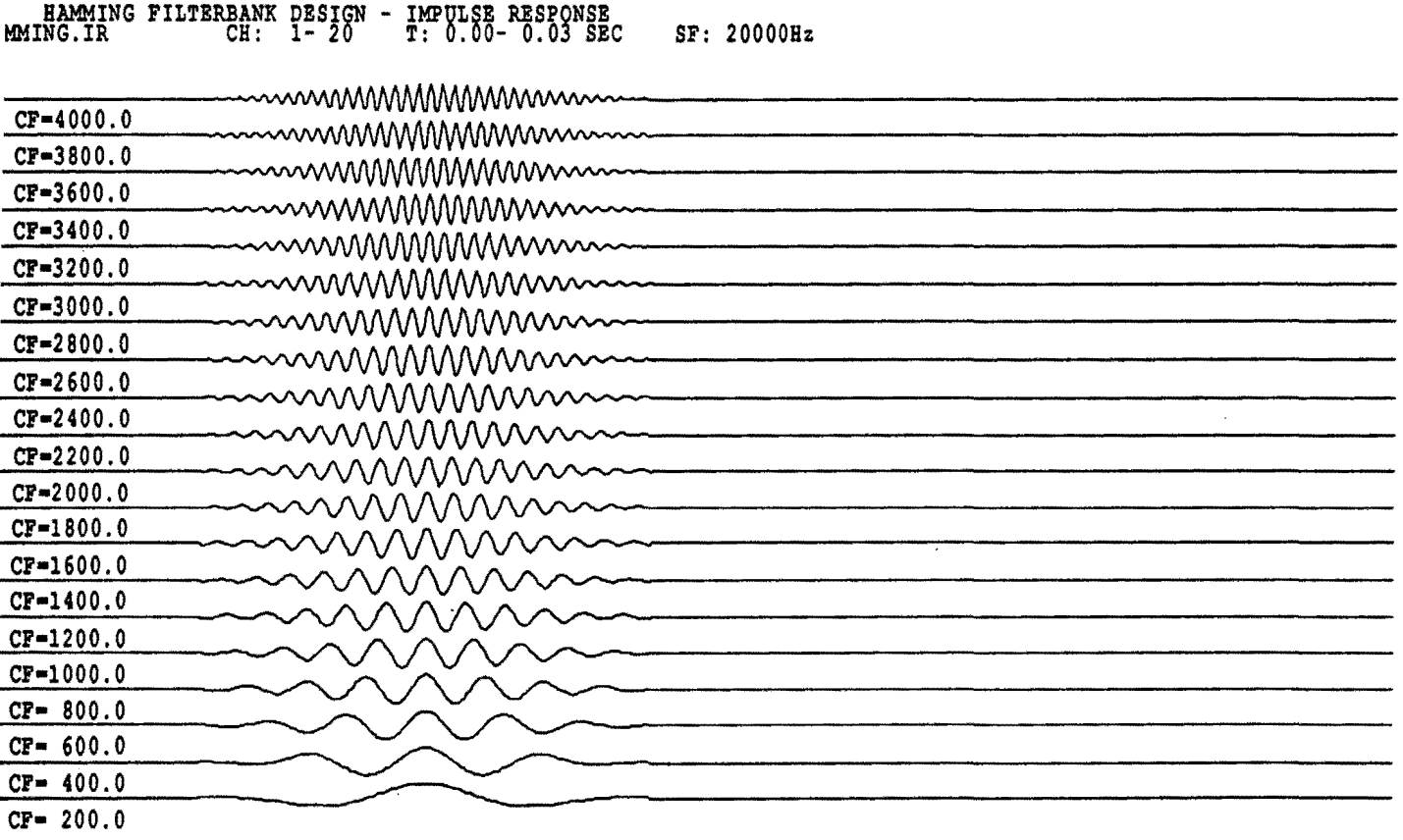
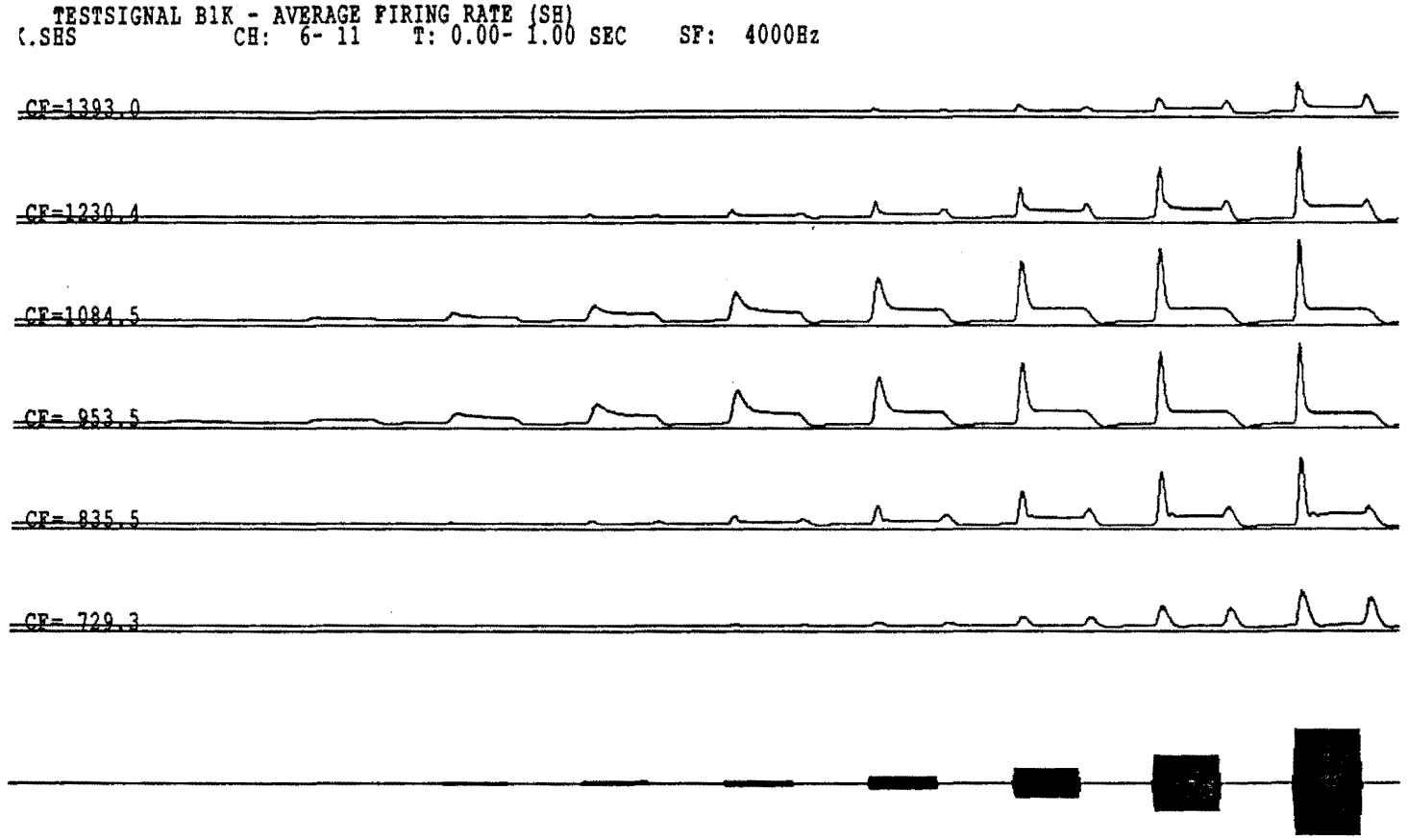


Fig. 2.5e


TESTSIGNAL B1K - AVERAGE FIRING RATE (SH)
(.SHS CH: 6- 11 T: 0.00- 1.00 SEC SF: 4000Hz
CF=1393.0
CF=1230.4
CF=1084.5
CF= 953.5
CF= 835.5
CF= 729.3

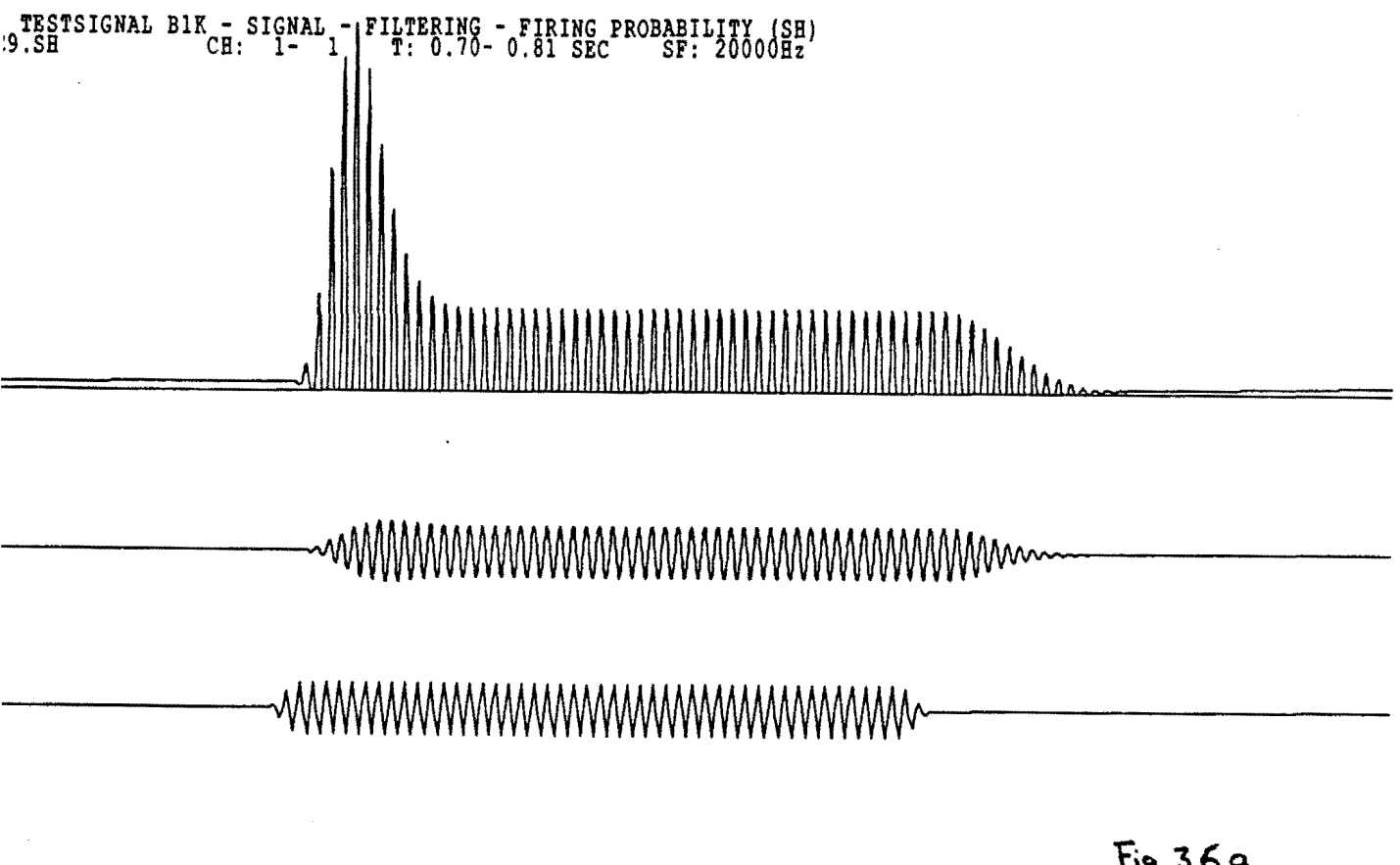

TESTSIGNAL B1K - SIGNAL - FILTERING - FIRING PROBABILITY (SH)
9.SH CH: 1- 1 T: 0.70- 0.81 SEC SF: 20000Hz

Fig 3.6a

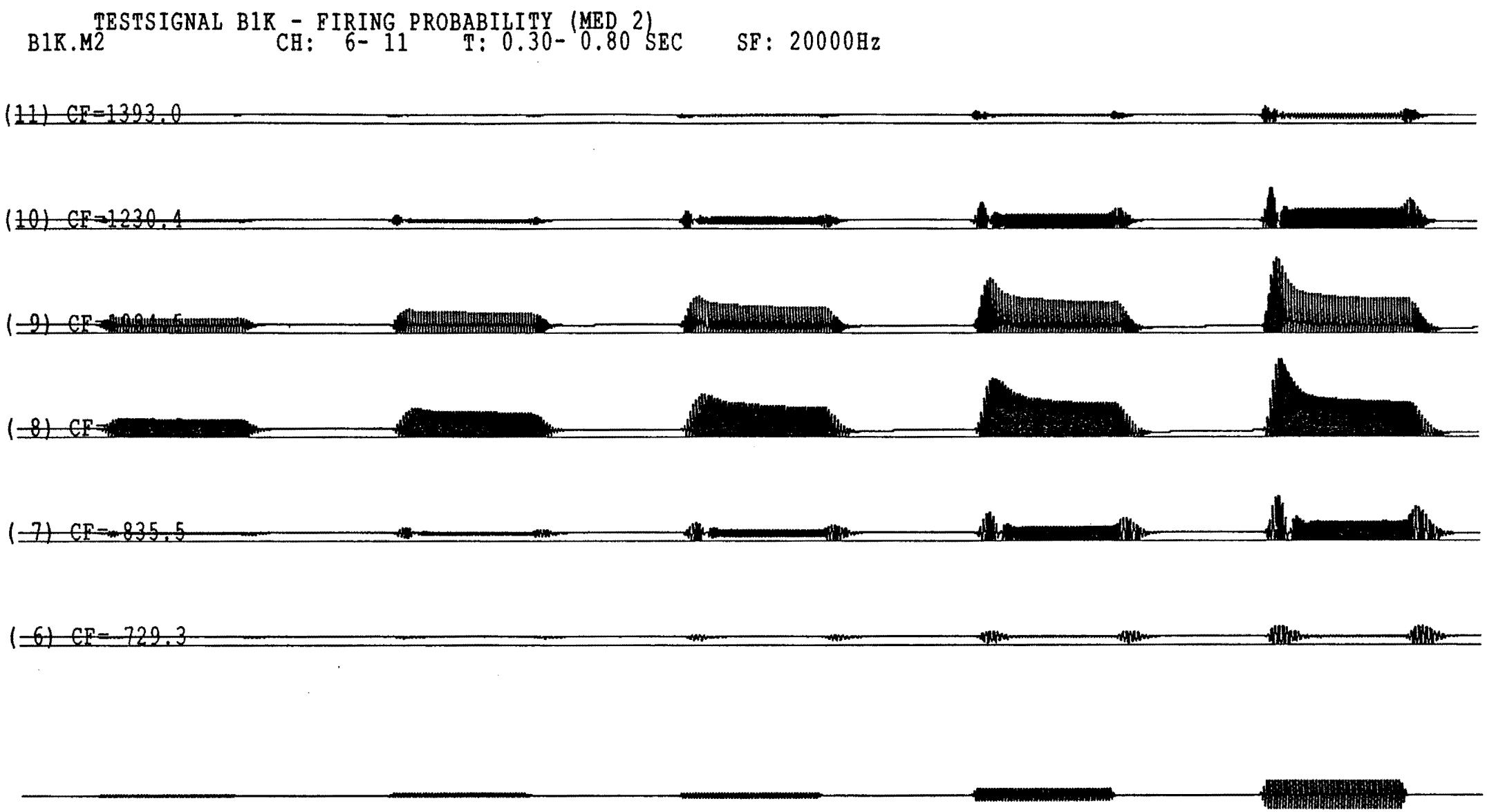

TESTSIGNAL B1K - FIRING PROBABILITY (MED 2)
B1K.M2 CH: 6- 11 T: 0.30- 0.80 SEC SF: 20000Hz
(11) CF=1393.0
(10) CF=1230.4
(9) CF=
(8) CF=
(7) CF= 835.5
(6) CF= 729.3

Fig 3.6b

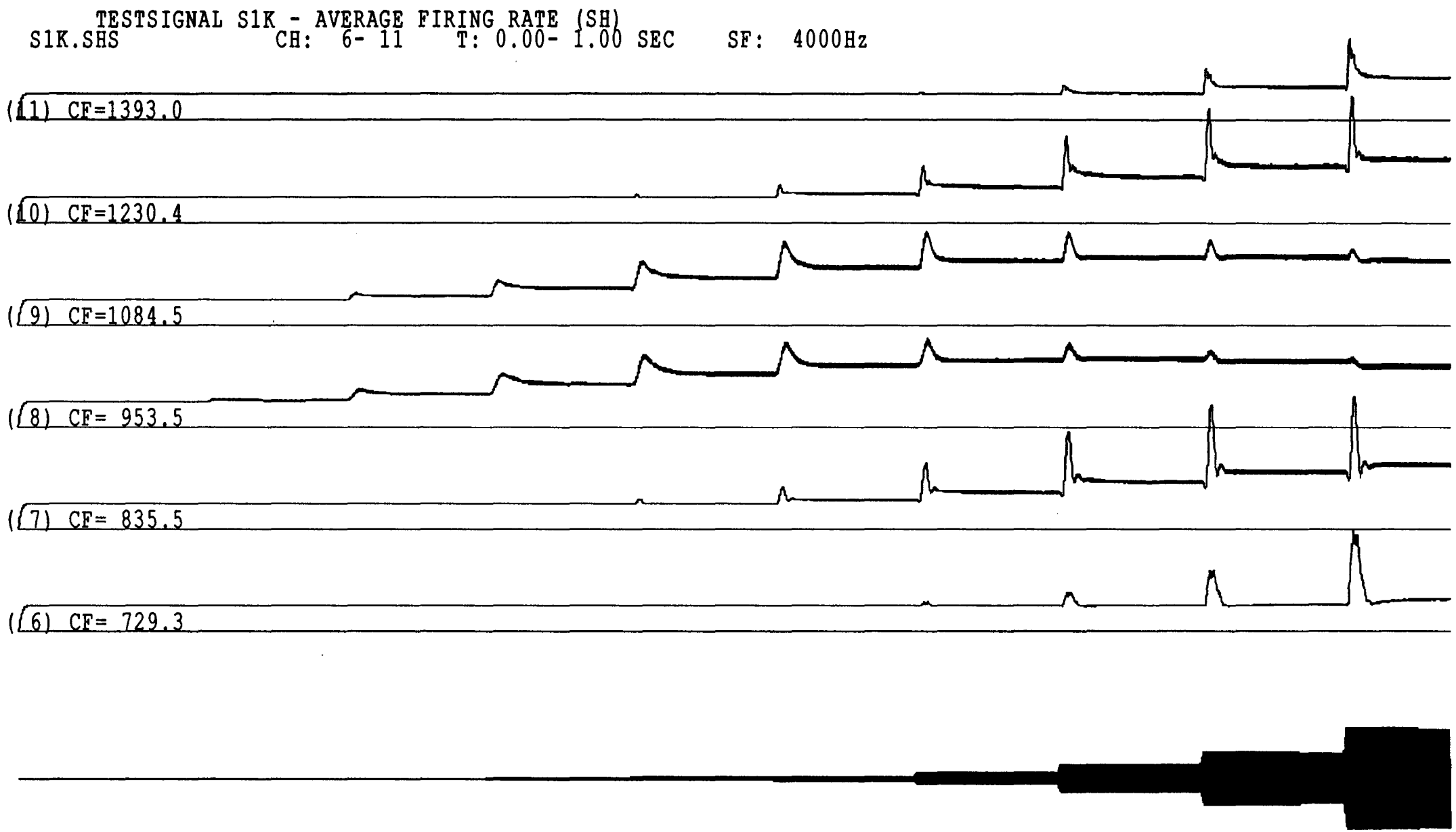

TESTSIGNAL S1K - AVERAGE FIRING RATE (SH)
S1K.SHS CH: 6- 11 T: 0.00- 1.00 SEC SF: 4000Hz
(11) CF=1393.0
(10) CF=1230.4
(9) CF=1084.5
(8) CF= 953.5
(7) CF= 835.5
(6) CF= 729.3

Fig 3.7a

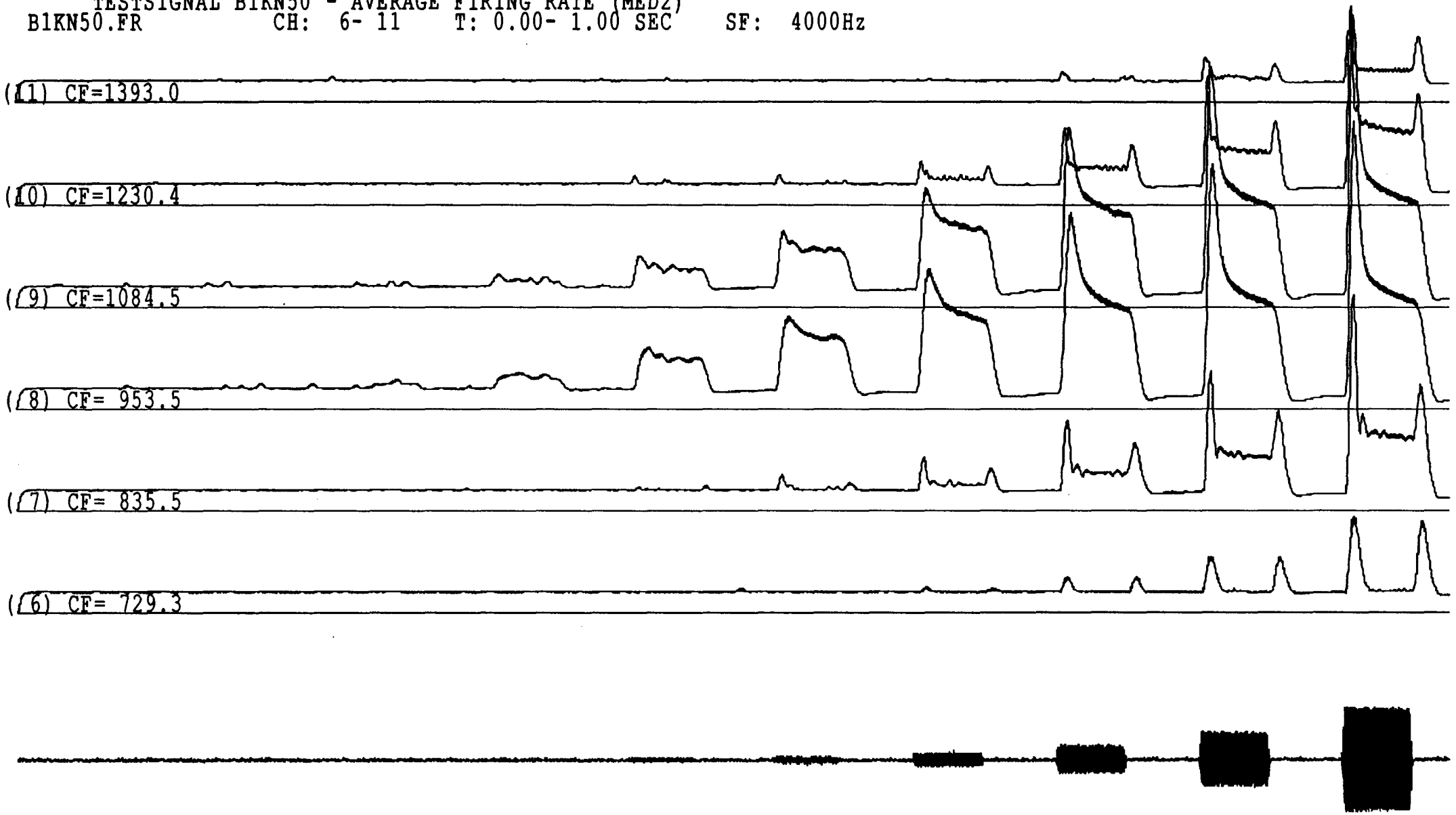

TESTSIGNAL B1KN50 - AVERAGE FIRING RATE (MED2)
B1KN50.FR CH: 6- 11 T: 0.00- 1.00 SEC SF: 4000Hz
(11) CF=1393.0
(10) CF=1230.4
(9) CF=1084.5
(8) CF= 953.5
(7) CF= 835.5
(6) CF= 729.3

Fig 4.1a

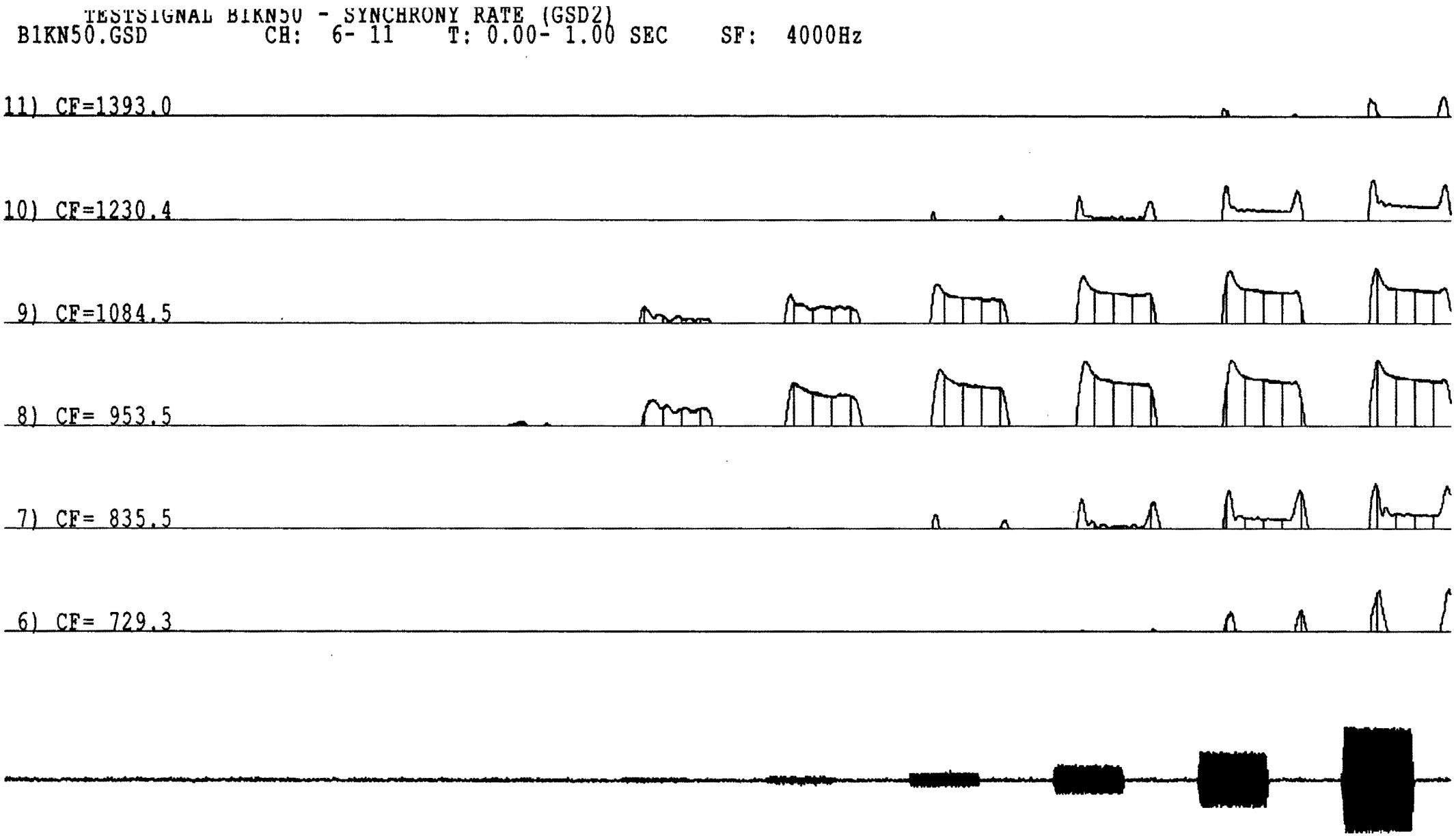

TESTSIGNAL B1KN50 - SYNCHRONY RATE (GSD2)
B1KN50.GSD CH: 6- 11 T: 0.00- 1.00 SEC SF: 4000Hz
(11) CF=1393.0
(10) CF=1230.4
(9) CF=1084.5
(8) CF= 953.5
(7) CF= 835.5
(6) CF= 729.3

Fig. 4.1b

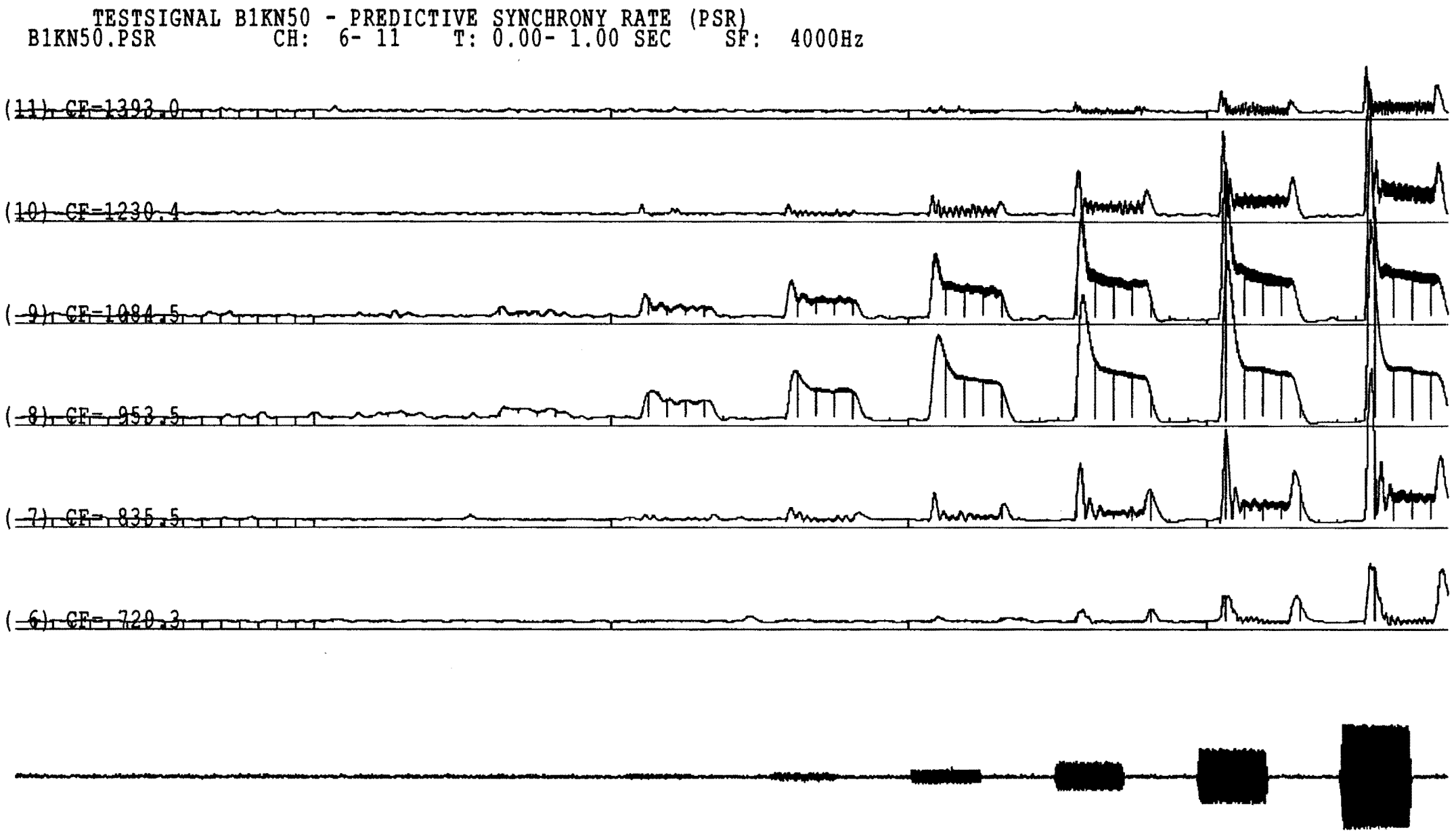

TESTSIGNAL B1KN50 - PREDICTIVE SYNCHRONY RATE (PSR)
B1KN50.PSR CH: 6- 11 T: 0.00- 1.00 SEC SF: 4000Hz
(11) CF=1393.0
(10) CF=1230.4
(9) CF=1084.5
(8) CF= 953.5
(7) CF= 835.5
(6) CF= 720.3

Fig 4.1.c