NATIONAL LIBRARY
OTTAWA

CANADA

BIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR.....*ANTON J. ROZSYPAL*

TITLE OF THESIS.....*Investigation and Functional Modeling of Auditory Perception of Transient Signals*

UNIVERSITY.....*of Alberta*

DEGREE FOR WHICH THESIS WAS PRESENTED.....*Ph. D.*

YEAR THIS DEGREE GRANTED.....*Spring 1972*

Permission is hereby granted to THE NATIONAL LIBRARY

OF CANADA to microfilm this thesis and to lend or sell copies

of the film.

The author reserves other publication rights, and

neither the thesis nor extensive extracts from it may be

printed or otherwise reproduced without the author's

written permission.

(Signed)..... *A. Rozsypal*

PERMANENT ADDRESS:

 # 1016 8515 – 112 St

Edmonton

Alberta, Canada

DATED. *May 10* 19 *72*

NL-91 (10-68)

THE UNIVERSITY OF ALBERTA

INVESTIGATION AND FUNCTIONAL MODELING OF AUDITORY

PERCEPTION OF TRANSIENT SIGNALS

BY

Ⓒ   ANTON JOSEPH ROZSYPAL

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL ENGINEERING

EDMONTON, ALBERTA

SPRING, 1972

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "Investigation and Functional Modelling of Auditory Perception of Transient Signals" submitted by Anton Joseph Rozsypal in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

....
Supervisor

....

....

....
External Examiner

Date May 1, 1972

ABSTRACT

The subject of this thesis is perception of acoustic transient signals by human observers. Two experiments investigating the time-frequency resolution of hearing were carried out.

In the first experiment the difference limen for duration of tone and noise pulses was measured. The influence of signal and subjective factors on duration discrimination was investigated. The significant factors are stimulus duration and individual differences between observers. Neither the envelope curve type, nor the spectral bandwidth of the carrier signal, nor its central frequency, exhibits any significant influence on the duration discrimination.

In the second experiment, a critical duration was found below which two acoustic pulses, differing in their envelope curves, are indiscriminable. The influence of central frequency and spectral bandwidth of the carrier signal, of the combination of different envelope curves, and of subjective differences on this critical duration was investigated. All signal factors and their first order interactions tested proved to be significant.

Data obtained in these two experiments, together with data on dynamical properties of hearing taken from other investigators, are discussed. It appears that the time

evolution of critical bands is the crucial process in perception of transient sounds. A model of a neuron network is suggested which simulates the organization of critical bands using integrators in the lateral inhibitory connections. The short-term adaptation is modeled by differentiators in the direct branches. Such a model simultaneously enhances contrast both in the frequency and time domains of the observation space.

ACKNOWLEDGMENT

TABLE OF CONTENTS

Chapter I

INTRODUCTION

All information-carrying signals consist of transient components which are relevant for information transmission. In many cases of communication our aim is to attain such a degree of similarity between the original sound signal and its recorded, transmitted, and reproduced version that an average listener is unable to discriminate between them. In processing of speech, music, and other sound signals we try to achieve the high fidelity requirements by suppressing the nonlinear distortion and operating in the full frequency band and in the full dynamic range of natural sounds. But as each additional unit of channel information capacity means additional cost, it is a question not only of aesthetics but also of economy to find the tolerable limits of signal distortion for each particular case of signal transmission.

This work was started with a view to the better understanding of the perception of transient sounds, tone and noise pulses of different envelope curves. In the experiments presented here the variable signal parameter was in all cases the envelope curve, either its duration or its shape. The carrier signal within each treatment was kept constant. The envelope, determining the amplitude changes of the sound signal, the slope of its onset and offset and the duration

of sound events, is an important factor contributing to the identification of diverse sound sources such as speakers and musical instruments. On the other hand, the duration of transient sounds is of particular interest in speech perception, as the recognition of speech sounds, namely consonants, may change when their duration is varied.

In each listening situation the auditory organ, in cooperation with other sensory modalities, especially vision, performs one or several tasks simultaneously, as the case may be. Hearing involves such processes as time-frequency and amplitude analysis, loudness evaluation, and time delay evaluation. Of all of these functions performed in hearing, in this work we were interested only in monotic discrimination between two successive transient sound signals, which can be interpreted as detection of the difference between two stimuli. In particular, two cases were investigated: duration increment detection and detection of the difference between two envelope curves as influenced by individual and several stimulus factors.

The main question which this work attempted to answer was whether the envelope analysis process is independent of the carrier signal, or, if not, what is the relationship between the two processes of carrier signal perception and envelope perception. In other words, to what extent these two decoding mechanisms of amplitude and frequency demodulation can be regarded as separate and exact.

Chapter II

GENERAL CONCEPTS OF SENSORY DETECTION

This chapter presents several basic concepts from the theory of communication and psychoacoustics, as background material for the discussion of subsequent chapters. Readers familiar with these fields may find little new in this chapter.

2.1 General Communication System

The general communication system for transmitting information over space and time consists of three blocks: source of information, transmission channel, and receiver of information, see Figure 2.1. In the case of transmission of analog signals, these can be represented either as continuous waveforms or as vectors, i.e., as points in multidimensional spaces. We will use the more convenient vector representation of signals. Vector $\bar{S}$ from the signal space $\mathcal{S}$ denotes the message selected by the source of information as input to the transmission channel. On its way through the channel this message is subjected to linear and nonlinear distortions and to random disturbances, known as channel noise. These distortions decrease the possibility of recovering the full information content from the signal at the receiver end. Signal $\bar{Z}$ is the output from the transmission channel and at the same time input to the receiver. On the basis of the signal $\bar{Z}$,
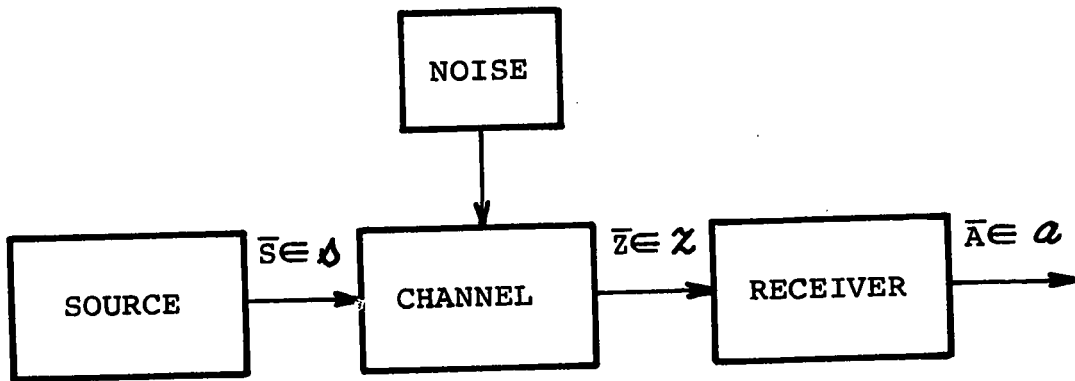
Figure 2.1 The block diagram of a general communication system.
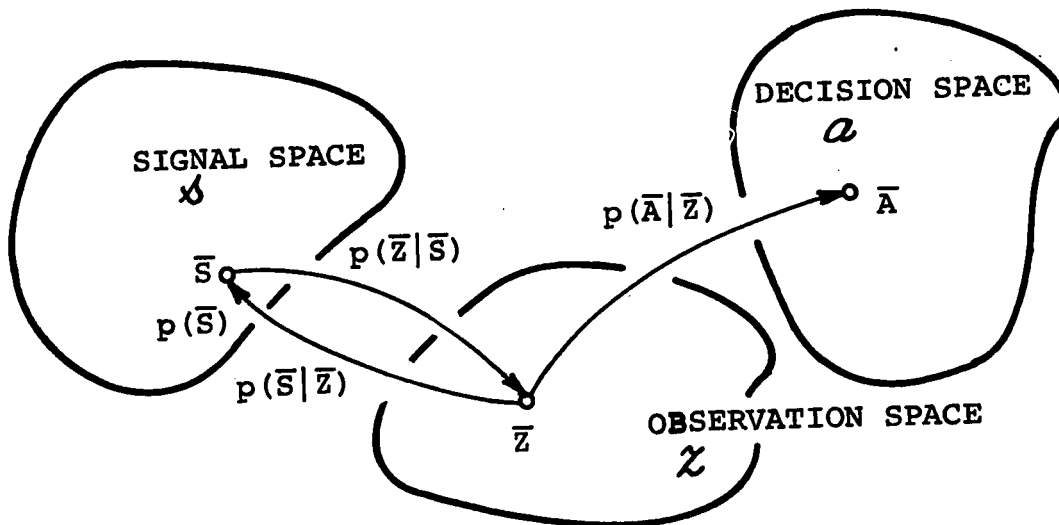
Figure 2.2 The mathematical model of a general communication system.

represented in the observation space $\mathcal{X}$, the receiver makes a decision $\bar{A}$, represented in the decision space $\mathcal{A}$. This decision is always some kind of estimation of the original signal $\bar{X}$.

The solution of communication problems calls for the application of the theory of probability. Due to the presence of the noise in the transmission channel, signal $\bar{Z}$ can be described in the observation space $\mathcal{X}$ only in terms of probability distributions. The mathematical model of the general communication system is presented in Figure 2.2. In this model $p(\bar{S})$ is the a priori probability of transmitting the message $\bar{S}$. The random properties of the transmission channel govern the probabilistic mapping of a point $\bar{S}$ from the signal space into the observation space $\mathcal{X}$. The conditional probability density distribution $p(\bar{Z}|\bar{S})$ expresses the probability of receiving the message $\bar{Z}$ provided the message $\bar{S}$ has been transmitted.

The mapping from the observation space $\mathcal{X}$ into the decision space $\mathcal{A}$ may in general be probabilistic, governed by the conditional probability density distribution $p(\bar{A}|\bar{Z})$. For most practical cases, however, this transformation is determined by some deterministic decision function.

The a posteriori probability $p(\bar{S}|\bar{Z})$ is the conditional probability that signal $\bar{S}$ was transmitted given signal $\bar{Z}$ was observed.

The interested reader may find these concepts developed in more detail in the books of H.L. van Trees (1968), or C.L. Weber (1968).

2.2 Hearing as a Communication Channel

Of the many possible types of information transmission,
we are concerned with the reception of transient acoustical
signals by the human auditory analyzer. For this purpose
we can consider both the stimulus generator and the trans-
mission channel outside the peripheral auditory channel, at
least in our experiments, as ideal, i.e., noiseless. Thus
the acoustic waveform at the entrance to the listener's ear
can be regarded as the output of the signal source. Corre-
sponding to Figures 2.1 and 2.2, it is the signal $\bar{S}$ from
the signal space $\mathcal{A}$. The signal transmitted through the
peripheral organ undergoes several transformations. First
come the linear and nonlinear distortions introduced by the
mechanical systems of the outer, middle, and inner ear. The
sound is diffracted by the observer's head and auricle. The
resonances of the air column in his external auditory canal
affect the transmission. The ossicular chain of the middle
ear serves as an impedance matching coupler which prevents
an excessive energy loss by reflection at the otherwise un-
matched boundary between the air and liquid media of the
outer and inner ear, respectively. The mechanical system
represented by the basilar membrane of the inner ear acts as
a frequency analyzer with limited frequency resolution. Its
deflection pattern represents the short-term Fourier spectra,
see Equation 2.1. Different frequency components are
spatially separated in the response pattern of the basilar
membrane. The time window of this short-term analysis is

given by Equation 2.2.

The total transmission characteristic of the mechanical part of the auditory organ was derived by J.J. Zwislocki (1965). This function relates the displacement amplitude pattern of the basilar membrane, which is relevant for the stimulation of the sensory cells, to the free field acoustic pressure waveform referred to the center of the listener's head. Computational models for middle ear and basilar membrane operation were presented by J.L. Flanagan (1965, p.91).

Up to this point the auditory information is encoded in mechanical motion of different parts of the peripheral auditory system. At this stage the mechano-electrical transformation is performed by the system of hair cells, arrayed on the basilar membrane. These sensory cells convert the time-space deflection patterns of the basilar membrane into patterns of neural activity. Because the neuromechanical feedback loop, adjusting the gain of the mechanical system of the middle ear, becomes effective only at high sound intensities, the mechanical and neural variables of the auditory channel can be regarded as separated. The electrical nerve signals are transmitted via the afferent auditory neural network to the auditory cortex.

The neural network is the part of the auditory transmission channel into which noise is introduced. This noise is the intrinsic noise resulting from the discrete character of neural action potentials, from the random character of the latent period, and from the neural spontaneous activity.

The ascending afferent paths have descending efferent counterparts of the same complexity. These two neural systems interact. This indicates that some preprocessing of auditory information under cortical or subcortical control takes place already at the subcortical levels.

The ultimate perception and decision process is located in the auditory cortex. This process corresponds to the receiver block of the general communication link in which the decision $\bar{A}$ is made on the basis of the observed signal $\bar{Z}$. At the present state of knowledge about the neural processing of information, it is difficult to decide at what neural level the observation space is located.

2.3 Representations of a Transient Signal in Signal Space

A nonperiodic transient waveform may be represented in the time domain by its time function $s(t)$.

In the frequency domain the most common description of a bounded energy signal is by its spectrum $S(f)$ given by the Fourier integral

$$S(f) = \mathscr{F}\{s(t)\} = \int_{-\infty}^{\infty} s(t) \, e^{-j2\pi ft} \, dt.$$

The Fourier transform is reversible

$$s(t) = \mathscr{F}^{-1}\{S(f)\} = \int_{-\infty}^{\infty} S(f) \, e^{j2\pi ft} \, dt.$$

Discarding the phase information, the signal can be

described by its energy spectral density $E(f)$

$$E(f) = |S(f)|^2 = S(f) \ S^*(f),$$

where $S^*(f)$ is the complex conjugate of $S(f)$.

Representation of the signal in time domain or in frequency domain requires only a single independent variable, time or frequency, respectively. Both these descriptions are mathematical idealizations. For auditory signals a more convenient signal description is one in which both time and frequency appear as independent variables. In that case the signal can be visualized as a surface in a three-dimensional space.

One such representation is the running, or evolution spectrum

$$S_t \ (f,t) = \int_{-\infty}^{t} s(\tau) \ e^{-j2\pi f\tau} \ d\tau$$

Another possibility is the short-term spectrum with stationary signal and sliding rectrangular temporal window of duration T

$$S_T \ (f,t) = \int_{t-T}^{t} s(\tau) \ e^{-j2\pi f\tau} \ d\tau$$

which is equivalent, except for phase factors, to short-term spectrum

$$S_T(f,t) = \int_{0}^{T} s(t-\tau) \ e^{-j2\pi f\tau} \ d\tau$$

with stationary temporal window and moving signal. In both cases, as time elapses, different portions of the signal $s(t)$ are subjected to Fourier analysis.

Instead of the rectangular temporal window, this window can be of an arbitrary form g(t). In that case the short-term spectrum takes the form

$$S_g(f,t) = \int_{-\infty}^{\infty} g(\tau)\, s(t-\tau)\, e^{-j2\pi f\tau}\, d\tau \tag{2.1}$$

$$S_g(f,t) = \int_{-\infty}^{\infty} g(t-\tau)\, s(\tau)\, e^{-j2\pi f\tau}\, d\tau$$

for the case of a stationary time window, or for the case of a stationary signal, respectively.

This generalized temporal window g(t) is also called a memory function or a weighting function, as it expresses the relative weight of past portions of the anlayzed signal s(t). In the case of the analysis of a signal s(t) by a linear time-invariant system, g(t) represents the impulse response h(t) of the analyzing system.

The short-term spectrum signal representation with the generalized temporal window is most relevant to auditory processing as the basilar membrane operates as an short-term spectrum analyzer. The basilar membrane in the inner ear can be considered as a linear time-invariant mechanical system with distributed parameters. Its temporal window at a point of maximal response to frequency f_0 was approximated by J.L. Flanagan (1965, p.126) to be

$$g(t) = (2\pi f_0\, t)^2\, e^{-\pi f_0 t} \tag{2.2}$$

The effective duration of the temporal window is inversely proportional to the frequency f_0. For the carrier frequencies used in our experiments this duration is approximately 4.8 msec for 250 Hz, 1.2 msec for 1 kHz, and 0.3 msec for 4 kHz.

Using the sampling theorem, H.S. Black (1953, page 41),
all the continuous waveforms and spectra mentioned above can,
without any loss of information, be expressed as vectors
in a multidimensional space. The vector coordinates represent
amplitude samples of a given waveform or surface sampled at
proper time and/or frequency intervals.

2.4 Signal Representation in Auditory Observation Space

It is still an open question at what stage of the audi-
tory system the observation space should be placed. Pre-
filtering and preprocessing of auditory information takes
place at all stages of the auditory neural network. If we
include preliminary information processing into the decision
process, the justifiable place for the observation space $\mathcal{Z}$
would be at the level of the hair cells and the observation
signal $\bar{Z}$ would be the space-time deflection pattern of the
basilar membrane. The other extreme would be to include all
the prefiltering and preprocessing operations into the trans-
mission channel and regard the neural activity patterns at
the input to the cortex as the observation signal $\bar{Z}$ in the
observation space $\mathcal{Z}$.

A series of three papers by A.G. Goluzina et al. (1966),
V.V. Lyublinskaya and A.G. Goluzina (1966), and M.K. Rohtla
and A. Rozsypal (1966) describes experiments investigating the
scaling of the interstimulus distance in the sensory observa-
tion space. The observers were presented with tone or narrow-
band noise pairs. The amplitude and duration of both pulses
in the pair was identical. The variable factors were the

stimulus duration, the stimulus frequency, and the frequency difference between the two stimuli in the pair. The observers were instructed to mark the perceived subjective difference between the stimuli in each pair in a scale from zero to six points.

The results were independent of the stimulus frequency when the frequency difference between the compared stimuli was expressed in mels of the pitch scale.

The results of these experiments support the assumption that the relevant coordinates of the signal representation in the observation space are the spatial and temporal samples of some excitation function of the reception system.

The pitch scale in mel units seems to be the natural scale for expressing the stimulus frequency in the observation space. It is linearly proportional to the spatial coordinate on the basilar membrane, to the jnd in frequency, and to the width of the critical band. Namely, one critical band spans about 1.3 mm of the length of the basilar membrane. This length corresponds approximately to 25 jnd in frequency and to 100 mels on the pitch scale.

2.5 Decision in Auditory Perception

In our experiments, as in most psychoacoustical experiments, the repertoire of observers' decisions was limited to two possible responses. So we will present here the basic ideas of detection theory, as they apply in the simple case of binary hypothesis testing, encountered in the experiments described in this thesis.

The observer was presented with a continuous series of stimuli, in which two pulses, differing either in duration or in envelope curve, were ordered by turns. The observer's task was to compare two successive stimuli and decide whether they were identical or different. In other words, to detect the difference between these stimuli. In this interpretation, the difference between two consecutive stimuli will be regarded as our signal $\bar{S}$ from the signal space $\mathcal{A}$. Figure 2.3 represents the mathematical model of a general communication system, Figure 2.2, altered for the case of a deterministic binary decision function. Owing to the random character of the transmission channel, represented in this case above all by the neural noise of the auditory neural network, the mapping of signal $\bar{S}$ into the observation space $\mathcal{X}$ is
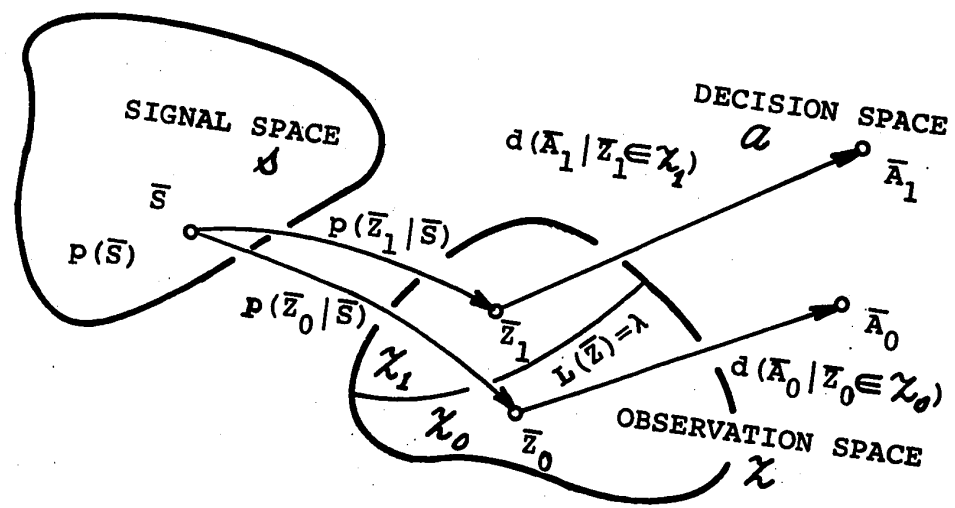
Figure 2.3 The mathematical model of a communication system with a deterministic binary decision function.

probabilistic, governed by the conditional probability
density $p(\bar{z}|\bar{s})$.

In the case of the deterministic binary decision function
of the form

$$d(\bar{A}|\bar{z}) = \begin{cases} \bar{A}_0 & \text{for } z \in \mathcal{X}_0 \\ \bar{A}_1 & \text{for } z \in \mathcal{X}_1 \end{cases}$$

the decision space $\mathcal{A}$ contains only two points, $\bar{A}_0$ and $\bar{A}_1$.
These two decisions are generated using the observer's criter-
ion dividing the observation space $\mathcal{X}$ into two non-overlapping
subregions, $\mathcal{X}_0$ and $\mathcal{X}_1$. Whenever signal $\bar{s}$ is mapped into
a point $\bar{z}_0$ in the subregion $\mathcal{X}_0$, the decision is $\bar{A}_0$. In
the complementary case, when $\bar{s}$ is transformed into $\bar{z}_1$ in $\mathcal{X}_1$,
the decision $\bar{A}_1$ is made with certainty.

It remains to be outlined what criterion the receiver
uses for division of the observation space $\mathcal{X}$ into the sub-
regions $\mathcal{X}_0$ and $\mathcal{X}_1$, i.e., to draw the boundary between
the decisions $\bar{A}_0$ and $\bar{A}_1$.

Let us assume that the stimulus source generates only
two types of stimulus pairs. In one pair both stimuli are
identical, the difference between them is zero. Such a pair
will be indicated as signal $\bar{s}_0$. In the other pair, $\bar{s}_1$, the
stimuli differ by the just noticeable difference. Signals
$\bar{s}_0$ and $\bar{s}_1$ are mapped into the observation space $\mathcal{X}$ with
conditional probability densities $p(\bar{z}|\bar{s}_0)$ and $p(\bar{z}|\bar{s}_1)$,
respectively.

The likelihood ratio $L(Z)$, is defined as

$$L(\bar{z}) = \frac{p(\bar{z}|\bar{s}_1)}{p(\bar{z}|\bar{s}_0)}.$$

For an observer whose decision is biased by a priori probabilities and by the costs and rewards associated with the decision, the decision function takes the form

$$d(\bar{A}|\bar{Z}) = \begin{cases} \bar{A}_0 & \text{for} \quad L(\bar{Z}) < \lambda \\ \bar{A}_1 & \text{for} \quad L(\bar{Z}) \geq \lambda \end{cases}$$

$$\lambda = \frac{p(\bar{S}_0) \; (C_{01} - C_{00})}{p(\bar{S}_1) \; (C_{10} - C_{11})} \quad .$$

The criterion λ depends on the a priori probabilities $p(\bar{S}_0)$ and $p(\bar{S}_1)$ and on the value of the cost matrix elements C_{ij} which express the costs or rewards to the observer if his decision is $\bar{A}_j$ when signal $\bar{S}_i$ was transmitted.

The boundary $L(\bar{Z}) = \lambda$ divides the observation space $\mathcal{Z}$ into two subregions $\mathcal{Z}_0$ and $\mathcal{Z}_1$, as indicated in Figure 2.3.

An ideal observer reaches his decision solely on the basis of minimization of the probability of error. Such an observer disregards the a priori probabilities as well as the costs and rewards associated with each decision. For the ideal observer the decision function becomes

$$d(\bar{A}|\bar{Z}) = \begin{cases} \bar{A}_0 & \text{for} \quad L(\bar{Z}) < 1 \\ \bar{A}_1 & \text{for} \quad L(\bar{Z}) \geq 1. \end{cases}$$

The index of sensitivity or index of detectability, d', is a useful characteristic of the sensitivity of the receiver or of the detectability of a given signal. It is defined as the mean value $\overline{\Delta m}$ of the signal distribution in the observation space $\mathcal{Z}$, divided by the standard deviation σ of this

distribution, see D.M. Green and J.A. Swets (1966, p. 60),

$$d' = \frac{\overline{\Delta m}}{\sigma} .$$ (2.3)

For our case, $\overline{\Delta m}$ is the mean and σ the standard deviation of the distribution $p(\bar{z}|\bar{s}_1)$.

2.6 Uncertainty Principle

In general, the accuracy of a receiver's observation is limited by the uncertainty principle. This law restricts the analyzing power of linear time invariant systems in the time and frequency domains. It was originally introduced into quantum mechanics by W. Heisenberg (1927) as an inverse relationship, governed by the Fourier transform, between the variances of particle position and particle momentum. This principle was later reformulated by G.W. Steward (1931) and D. Gabor (1946), (1947) for use in acoustics and the theory of communication.

In its general form, this principle states that the accuracy of any observation is proportional to the duration of the observation. According to one interpretation of this principle, the product of the signal duration Δt and the bandwidth of its Fourier spectrum Δf is constant

$$\Delta f \, \Delta t = c .$$ (2.4)

Another interpretation similarly relates the duration Δt of the impulse response of a linear time invariant system and its frequency resolution bandwidth Δf. In both cases, the time and frequency functions are Fourier pairs.

The value of the constant c in Equation 2.4 is of the

order one. Its exact value depends on the arbitrary way
the effective bandwidth Δf and the effective duration Δt
are defined. D. Gabor (1946), for instance, used as his
measure of duration and bandwidth the second moments of the
time and frequency distributions, while A.A. Kharkevich (1960)
used the energy concentrations of appropriate functions in
time and frequency domains.

Caution is required in the direct application of the
uncertainty principle of observation, as described above,
to psychoacoustics. Some erroneous inferences could result
for two reasons. First, a nonlinear pitch scale in mel
units rather than a linear frequency scale in Hertz is in-
herent to hearing. Second, even though the mechanical part
of the auditory analyzer can, with a good approximation, be
regarded as a linear time invariant system, this is not
necessarily true about the following neural stage.

2.7 Critical Bands

The critical band concept appears to be crucial for the
time-frequency analysis carried out by hearing. In many
experiments on seemingly unrelated auditory perception
phenomena, such as loudness, masking threshold, threshold of
hearing, or detection of amplitude and frequency modulation,
the investigators arrive at strikingly consistent values of
critical bands as a function of central frequency of the band.
These values are plotted in Figure 6.11. Next we will men-
tion several of these experiments.

H. Bauch (1956), H. Niese (1960)(1961), B. Scharf (1959a)

(1959b) (1961) (1962), E. Zwicker and R. Feldtkeller (1955), and E. Zwicker, G. Flottorp, and S. Stevens (1957) investigated the dependence of the loudness of stationary stimuli composed of two or more spectral components on their spectral bandwidth. They all found that the loudness of these complex stimuli of constant intensity is independent of their spectral bandwidth as long as this bandwidth is narrower than the critical band. In that case the loudness of the complex stimulus is equal to the loudness of a pure tone of the same intensity and of frequency equal to the central frequency of the complex stimulus. The loudness of the complex stimulus of the same intensity begins to decrease when its spectrum is wider than one critical band.

Another group of experiments involved masking of a pure tone by band or bands of noise, as carried out by C.E. Bos and E. de Boer (1966), E.Zwicker and E.Feldtkeller (1967), P.M. Hamilton (1957), T.H. Schafer et al. (1950), and D.D. Greenwood (1961a)(1961b). These experiments indicate that as long as the spectral bandwidth of the noise masker is narrower than or equal to the critical band, the masking threshold of the pure tone as a function of its frequency is unimodal with a peak at the central frequency of the masker noise. Incrementing the bandwidth of the masker above the critical bandwidth resulted in a bimodal masking threshold function with peaks at the edges of the masker spectral band.

E. Zwicker (1954) masked a narrow-band noise by two pure tones with frequencies located symmetrically around the

maskee noise. The masking threshold of the noise band was constant until the frequency separation of the masking tones exceeded the value of the critical bands. D.D. Greenwood (1961a) expanded the above experiment by measuring the masking threshold as a function of the central frequency of the maskee noise band for a given frequency separation of the masker tones. For separations smaller than critical, the frequency course of the masking threshold was unimodal, with its peak between the frequencies of the two masker tones. For greater than critical frequency separation of the masker tones, the course of the masking threshold was bimodal.

G. Gässler (1954) measured the absolute threshold of multitone complex consisting from up to forty pure tones of equal amplitudes. Their frequency separation was either 10 or 20 Hz. The total intensity of the stimulus was kept constant, independently of the number of components. Upon the addition of a new component, the absolute threshold was constant until the stimulus spectral bandwidth reached the width of the critical band. Then, for each new component added, the stimulus intensity had to be incremented in order to reach the absolute threshold.

E. Zwicker (1952) studied the differences in perception of amplitude and frequency modulation. Differences in perception take place only for small modulation frequencies when all three components of the amplitude modulated signal

and all three most pronounced components of the frequency modulated signal fall into one critical band. If the separation of these components is wider, the hearing resolves them as three separate components. These experiments are described in more detail in Section 6.3. This indicates that hearing is sensitive to phase relations between stimulus components only as long as these components lie within one critical band.

All these experiments used stationary signals as masker and maskee. Later experiments revealed that the critical bands are composed as a response to the stimulus and that this process takes about 300 msec. Experiments investigating the dynamical properties of the critical band, as carried out by H. Scholl (1962a)(1962b), E. Zwicker (1965a)(1965b), are discussed in Chapter VII.

It can be said that the critical band characterizes the length of the interaction area on the basilar membrane, more specifically, the extent of the inhibition area of the response unit, as defined by G. von Békésy (1967).

According to J.J. Zwislocki (1966) one critical band, regardless of its central frequency, covers about a 1.3 mm long segment of the basilar membrane which contains approximately 1300 peripheral neurons. On the pitch scale one critical band corresponds to 100 mel steps, or to about 25 just noticeable increments in frequency. On the frequency scale, up to central frequency about 500 Hz, the width of the critical band is constant, about 100 Hz. For higher

central frequencies the relative width of the critical band

is constant and equal to about 20 per cent of the central

frequency.

Chapter III

EXPERIMENTAL METHOD

Two experiments were carried out, both of them con-
cerned with discrimination of the envelope curve of acoustic
pulses. Essentially, in the duration discrimination ex-
periment two pulses of the same envelope and of different
durations were to be discriminated. In the envelope
discrimination series two pulses of equal duration and of
different envelopes were compared. Both experiments were
carried out using the method of limits often called the
up-down procedure. By presenting the observer with the
stimuli only from the neighborhood of the measured
threshold, this adaptive method leads efficiently to the
threshold value, which was of our primary concern. The
efficiency of this psychophysical method is gained at the
expense of not yielding exact information about the shape
and spread of the psychometric function.

3.1 Duration Discrimination Testing

In the duration discrimination experiments the observers
listened to continuous succession of acoustic pulses, in
which the standard and variable pulses were ordered by turns.
The standard pulse was of the constant predetermined duration,
while the duration of the variable pulse was longer by a
duration increment. This increment was continuously changed
by the experimenter in the following way: Each test started

with easily perceptible duration difference between the two pulses. Then the duration increment was gradually diminished until the observer signalled "pulses identical". From this moment the duration of the variable pulse was gradually increased until the subject again perceived the difference and signalled "pulses different". After that the operator started to shorten the variable pulse again. The reversal points of the variable pulse duration were measured and recorded by the experimenter. For each subject and for each combination of signal parameters ten such up and down cycles were performed. The difference limen for duration was calculated as the difference between the standard pulse duration and the average value of the twenty turnover durations of the variable pulse.

Owing to the fact that the interstimulus interval in the pulse series was the same, both between pulses in the pair and between pulses of successive pairs, and that the rate of change of the variable pulse duration was slow, this test arrangement provided no clue as to which pulse is standard and which one is variable. So we did not regard it necessary to measure the variant of this experiment where the variable pulse is shorter than the standard one.

3.2 Envelope Discrimination Testing

In the envelope discrimination series of experiments, again, the method of limits was used. Here two pulses were compared, but in this case neither of the pulses could be denoted as standard or variable, as both were of the same

variable duration and carrier signal. They differentiated one from another only by their envelope curves. In this experiment the test started with pulse durations well above the critical duration so that the two different pulse shapes evoked markedly different sensations. Then the duration of both pulses was simultaneously shortened until the observer ceased to perceive any difference between them and signalled "pulses identical". At this point the duration of both pulses was gradually increased until the subject became aware of the difference again. Also in this experiment ten up and down cycles were performed for each observer and for each combination of signal parameters. The critical duration was thus obtained as an average value of twenty turnover points.

3.3 Interstimulus Interval

The duration of the interstimulus interval, i.e., the pause between the termination of one pulse and the onset of the following one, affects the results of all psychophysical experiments where two successive stimuli are compared. The silent interval should be chosen long enough to eliminate all carry-over effects due to mutual interference between the processes evoked by the two stimuli to be compared. Ideally, the new stimulus should arrive only after all mechanical and neural processes caused by the previous one have completely ceased. In both experiments described here the observers were subjected to incessant series of stimuli. So here additional reason applies to choose the interval

comparatively long: It gives the subject enough time to make his decision after listening to each pulse in the series. On the other hand, too long an interstimulus interval causes memory errors which can be interpreted as progressive blurring of the stimulus representation in the observation space due to imperfections of memory.

In view of the above our interstimulus interval, with no regard to the pulse duration, was in all experiments of the duration of 1.1 second. A recently published paper by W. Reichardt and H. Niese (1970) on stimulus and interstimuli interval durations in loudness evaluation tests confirmed our choice. According to the authors' recommendation, the interval of silence should in no case be shorter than 500 msec and not substantially longer than 1 second. The acceptable interstimuli interval in frequency discrimination tasks is obviously even longer. I.B. Thomas et al. (1970) measured the decline in pitch discrimination as affected by the duration of the interstimulus interval. Using an ABX procedure they stated that the percentage of correct identifications dropped abruptly from a nearly constant level only for an interstimulus interval longer than 20 seconds for the tone pulse pairs with the frequency difference 10 per cent and 5 seconds for the frequency spacing 2 per cent.

3.4 Stimulus Intensity

Loudness of the pulses was adjusted by the following procedure. Prior to each testing session, for each combination of the signal parameters and individually for each observer,

the durations of the variable pulses were set at the following value: In the duration discrimination experiment, the duration of the variable pulse was made to be equal to the standard pulse duration, while in the envelope discrimination experiment, the duration of both pulses was set to be equal to the expected critical duration as ascertained by preliminary and preceding experiments. After the threshold of hearing of this pulse series had been found, the intensity level of the stimuli was increased over this individual threshold of hearing for a particular stimulus series by 60 dB for the pulses of central carrier frequency 250 Hz and by 70 dB for the rest of the stimuli. This procedure established the loudness level of the pulses in all cases to be approximately 75 Ph which is according to I. Pollack (1952) within the region of the comfortable listening levels for monaural sound reproduction in quiet environments. Our subjects considered this loudness level convenient; they did not complain that the testing session caused fatigue or discomfort.

3.5 Group of Subjects

Observers with normal hearing sensitivity were selected on the basis of an audiometric examination. Six university students, ages 18 to 23, participated in each experiment as subjects. In the duration discrimination experiment males Va, Ja, Bl, and Be, and females Ch, and Sp took part. All these subjects were considered to be experienced listeners with previous practice as observers in experiments on auditory

discrimination of transient signals. Two persons from this group, Sp and Bl, for personal reasons, were unable to participate in the subsequent envelope discrimination experiment. They were substituted by two newly selected females Ke and Ko, who were admitted to testing after several preliminary sessions.

Previous to both experiments the subjects had been given an instruction to indicate by way of the two-position switch any perceptible difference between the pulses in the series with no regard whether they observed the difference in the duration, loudness, or quality of the sensation.

This simple criterion has been chosen after experience from our earlier experiments of a similar nature, where tone and noise pulses with exponential onset and offset were discriminated, I. Nábělek, A. Rozsypal, and V. Balko (1965); I. Nábělek (1965). The variable signal parameter was the time constant of the pulse rise and decay. Three different instructions were given to the subjects. Under one test condition they were expected to base their decision only on the loudness difference between the pulses, while under the other testing condition solely on the perceived quality of the onset and decay transients. As a criterion for the third testing condition any perceived difference between the two stimuli was selected. Data obtained under these various experimental conditions differed only nonsignificantly. This fact suggests that either all three criteria are equally powerful, or, regardless of the instruction, subjects sooner

or later tend to switch inadvertently to discrimination criteria of their own.

The subjects were not furnished any information or comment about their responses, either during the testing session or after it. They were paid on an hourly basis.

3.6 Further Testing Conditions

The rate of duration change of the variable stimulus or stimuli was in both experiments less than 1 per cent per pulse pair. The order of test treatments was randomized. All subjects were exposed eventually to all combinations of factor levels measured.

One session lasted for about 20 minutes. As the subjects were invited to testing in pairs and tested separately, each testing period was followed by a rest interval of at least the same duration. During the testing the observer was seated in an anechoic chamber with lights dimmed. Subjects spent not more than three hours daily in our laboratory, and only less than half of this time in actual testing.

The apparatus was located in a control room. The earphone and the subject's signal panel were situated in an adjoining anechoic chamber. The only means of communication between the subject and the experimenters during the test were the two-way signal lights operated by lever-key switches.

Chapter IV

INSTRUMENTATION

In both the duration discrimination and envelope discrimination experiments essentially the same apparatus was used. The block diagrams are presented in Figures 4.1 and 4.5, respectively. Technical specifications of the apparatus and complete list of instruments used in it are given in the closing paragraphs of this chapter.

Except for the part producing the control voltage for the modulator, the apparatus, its adjustment before and operation during the experiments, were identical. So following the detailed description of the setup assembled for the first experiment we will point out only the modifications needed for the second one.

4.1 Apparatus for the Duration Discrimination

Experiment

As the standard and variable pulses were of the same envelope in the duration discrimination experiment, only one envelope curve generator was required in the experimental setup, as Figure 4.1 shows. This apparatus allowed the experimenter to present to the observer a continuous series of acoustic pulses of the same amplitude, carrier signal and envelope. No provision was made for any phase lock between the carrier and envelope signals. The standard pulses were
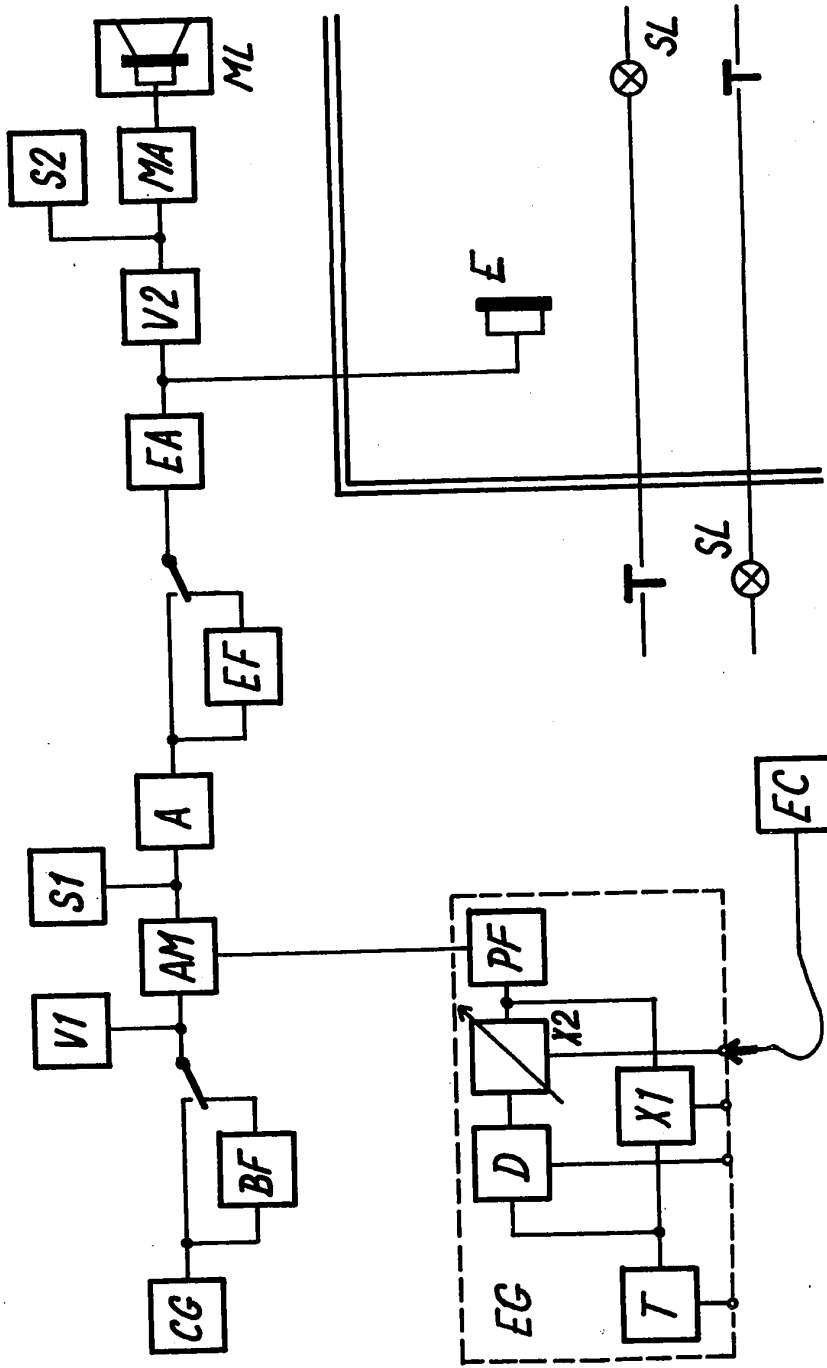
Figure 4.1   Block diagram of the apparatus used in the duration discrimination experiment.

of preset constant duration. The experimenter had control over the duration of the variable pulse and changed it manually in a continuous way according to subject's responses. In the diagram the single blocks represent the following units and instruments: The carrier signal, either pure tone, white noise, or narrow-band noise, was generated by a sine-random generator CG. In the cases where band-limited noise was required as carrier signal, the white noise produced by CG was filtered by an adequate band-pass filter BF. The normalized frequency characteristics of the different filters used for this purpose are given in Figure 4.2. The level of the carrier signal was kept at a constant value checked by means of the electronic voltmeter V1. This continuous carrier signal was fed into the controlled input of the amplitude modulator AM, based on the original design of S.F.Vaytulevich et al.(1962). The control input of this modulator was driven by a control voltage generated in the envelope curve generator EG. This analog device was based on the photoelectric waveform generator of D.E.Sunstein (1949) and described by the author of this thesis in a separate technical report (1966). The phantastron timing unit T generated primary timing pulses at adjustable repetition rates. These pulses triggered both the first deflection unit X1 and the time delay unit D. The output pulses from D, delayed in reference to the primary timing pulses, actuated the second deflection unit X2. The time delay unit consisted of a monostable multivibrator. The basic element of both
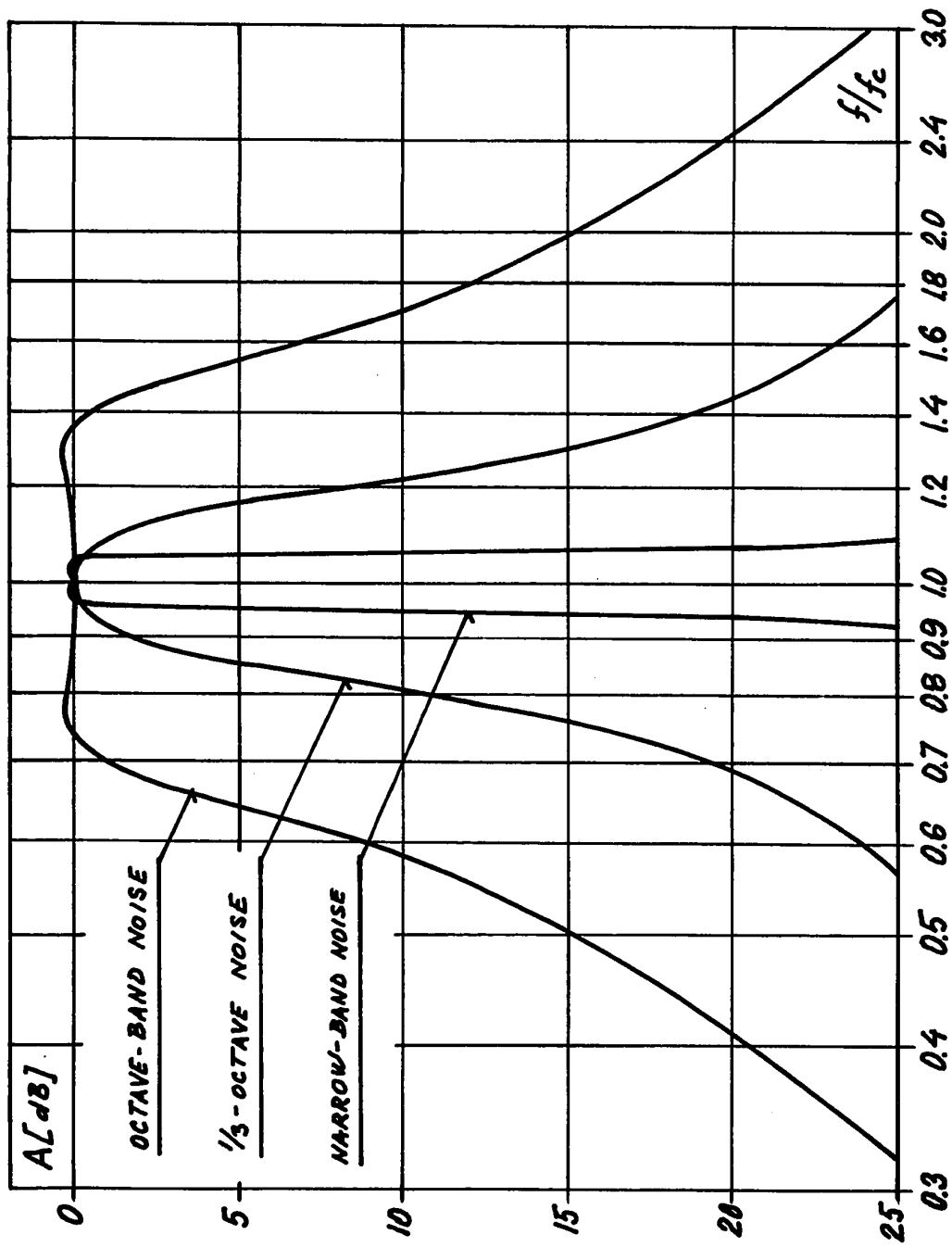
Figure 4.2   Normalized frequency responses of the band-pass filters
used to generate the band-limited noise carrier signals.

identical deflection units was a phantastron circuit, which, once triggered, produced one linear sweep. In all of these four units which formed the time base of the envelope curve generator the generated time intervals were continuously adjustable in several partly overlapping ranges. These units, if adequately set, determined the required repetition rate of the pulse pair (T), the time delay between the onsets of the standard and variable pulse in the pulse pair (D), as well as the duration of the standard (X1) and the variable (X2) pulse. The photoformer unit PF, driven by turns by the ramps from the two deflection units, produced the control voltage which corresponded in shape to the contour of the opaque mask attached to the screen of the CR tube linked-up, together with a photomultiplier and a vertical deflection amplifier, in a combined electrical-optical negative feedback loop. The function of this feedback circuit was to keep the light spot of the CR tube at the upper edge of this mask. The vertical deflection voltage was used to control the gain of the modulator. In order to preserve the DC component of this control voltage, the coupling between the photoformer and the modulator was direct. The duration of this modulator control voltage pulse was determined by the duration of the sweep pulse from deflection units, brought to the horizontal deflection amplifier of the photoformer. In this way the voltage controlled amplitude modulator yielded at its output the product of two signals, the carrier signal and the envelope signal, in the form of tone or noise pulses of the

required shape and duration. The amplitude of these pulses was adjustable in 0.1 dB steps by means of an attenuator A. The frequency response of the earphone was compensated in an equalizer filter EF, described by E.Zwicker and D.Maiwald (1963). Its output was connected to the earphone amplifier EA which drove the dynamic earphone E mounted in a foam rubber cushion and wired for monotic stimulus presentation. The experimenter and observer communicated in both directions by way of signal lights SL.

Figure 4.3 shows the pulses of different envelopes as monitored at the output of the amplitude modulator. The waveforms were photographed from the oscilloscope face. In order to permit better evaluation of the details of the envelope curve and of its noisiness, the pure tone carrier signal is of the frequency 15 kHz. The distortion of the acoustic waveform by the transients introduced by the earphone transducer, is illustrated in Figure 4.4. The traces recorded at the output of an artifical ear represent an acoustic version of a divergent triangle stimulus of an effective duration 1.7 msec and carrier frequency 1 kHz and 4 kHz, respectively.

The electronic voltmeters V1 and V2, as well as the oscilloscopes S1 and S2, and monitor loudspeaker system ML powered by the monitor power amplifier MA, all served to adjust the apparatus before starting the testing session and to check the pulses both visually and aurally during the testing. To protect the feedback circuit components against
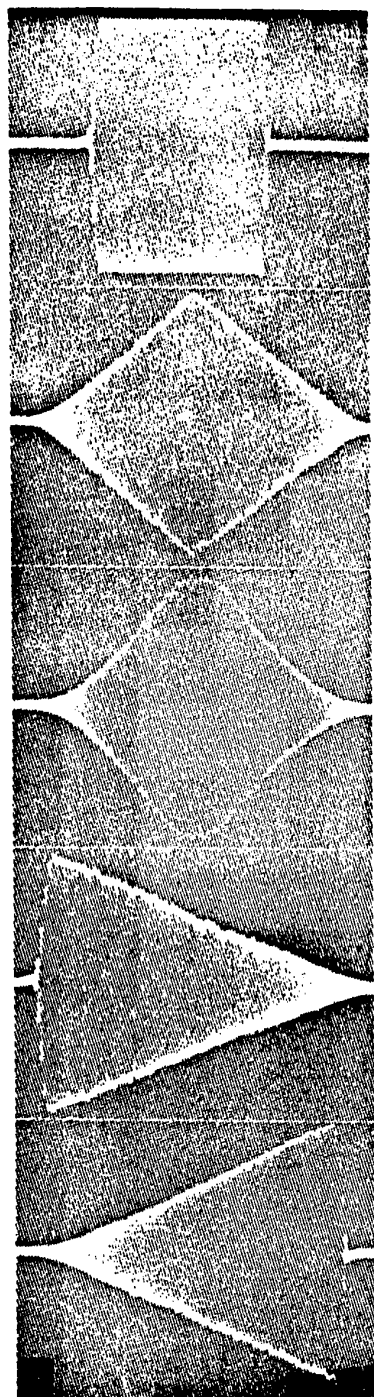
Rectangular envelope

Isosceles triangle envelope

Gaussian envelope

Convergent triangle envelope

Divergent triangle envelope

Figure 4.3   Output of the amplitude modulator representing
the stimuli of five different types of envelope
curves. Carrier tone frequency 15 kHz.
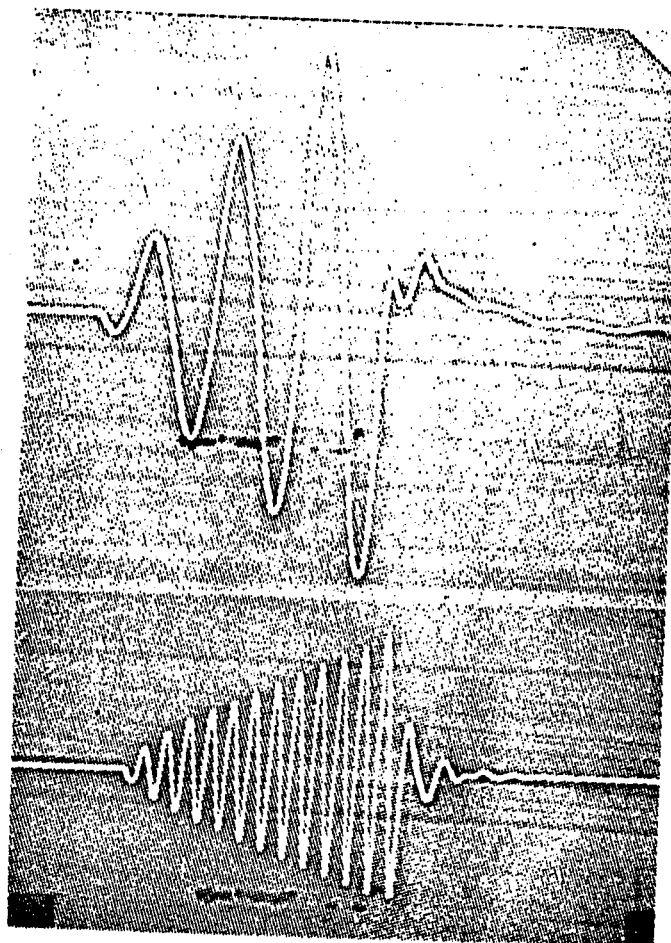
Figure 4.4   Transient distortion introduced by the earphone

transducer, illustrated by the output from the

artifical ear. Divergent triangle envelope,

effective duration 1.7 msec.

Upper trace: carrier tone 1 kHz.

Lower trace: carrier tone 4 kHz.

the stray light, the feedback CR tube and the photomulti-
plier were placed in a lightproof enclosure, thus inacessible
to direct observation. The proper function of the photoformer
could be watched on the monitor CR tube wired in parallel
with the feedback CR tube. The setting of all timing units
was checked by an electronic chronometer EC. During the
testing this chronometer was connected to the deflection unit
determining the duration of the variable pulse to allow the
experimenter to read the duration of the variable pulse.

4.2 Apparatus for the Envelope Discrimination

Experiment

In contradistinction to the apparatus just described,
the experimental setup for the envelope discrimination
experiment featured two envelope curve generators, EG1 and
EG2, as indicated in the block diagram in Figure 4.5. Since
two stimuli of different envelopes were compared in these
tests, these two photoformers, actuated alternately by their
own respective deflection units, produced two envelope
control signals of different forms . The control input stage
of the modulator AM had been doubled for this purpose. To
guarantee that the duration of both pulses was equal, the
vernier sweep duration control potentiometers of the deflec-
tion units X1 and X2 were mechanically coupled. Care was
excercised in selection of two deflection units with identi-
cal sweep duration scale division. Otherwise the apparatus
was identical with that one used in the duration discrimina-
tion experiment.
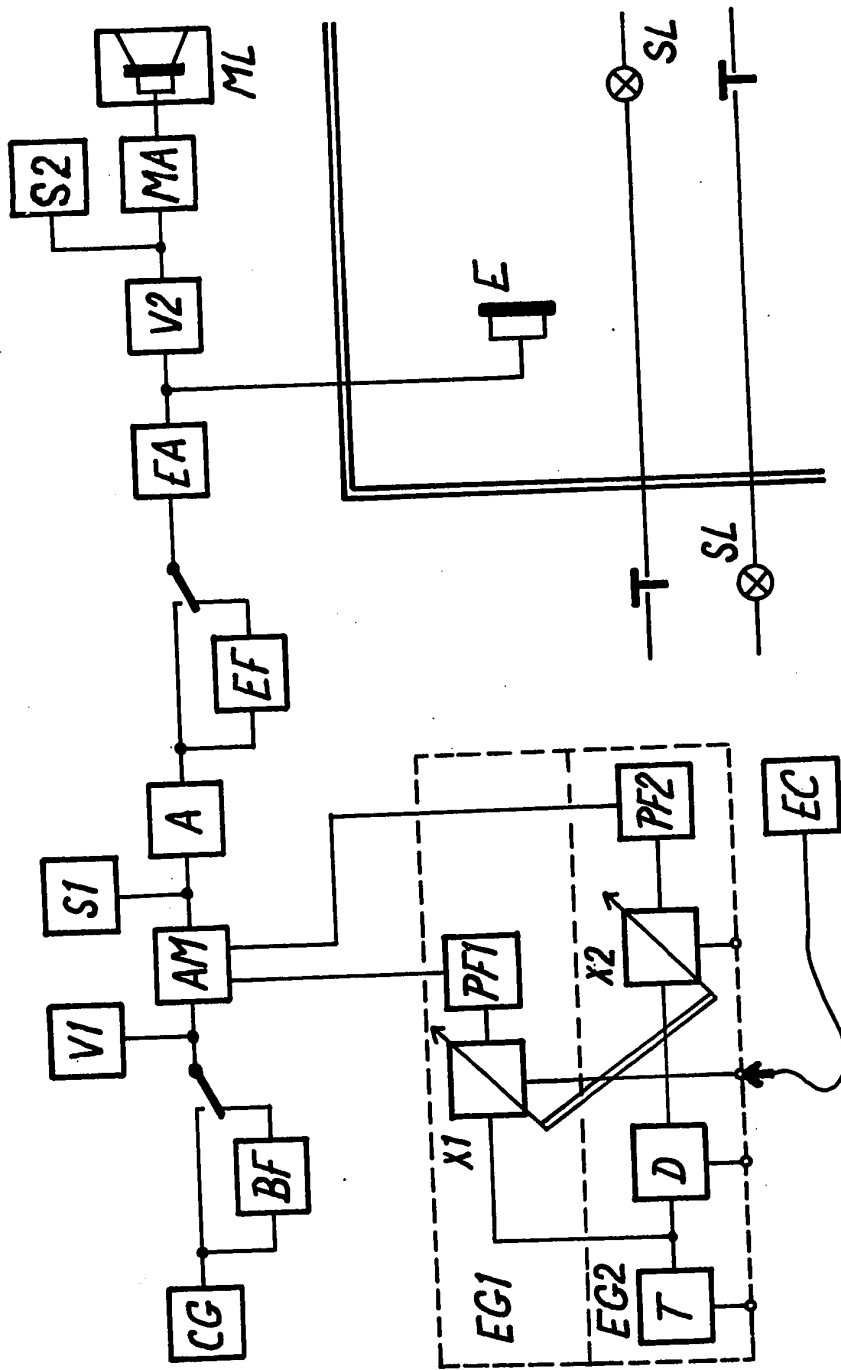
Figure 4.5   Block diagram of the apparatus used in the envelope discrimination experiment.

4.3 Technical Specifications of the Apparatus

The carrier signal channel with the equalizer filter bypassed, as measured at the input of the earphone, the amplitude modulator being opened to its nominal level:

Frequency response   50 Hz to 15 kHz ± 1.5 dB

Nonlinear distortion   less than 0.5 %

Background noise   less than -75 dB

The envelope curve channel, measured at the control input to the amplitude modulator:

Frequency response   DC to 6.8 kHz ± 1.5 dB

Linearity error referred to max. amplitude   less than 2 %

Noise level referred to the peak value

    of the control voltage   less than -38 dB

Spurious switching transients, peak to peak, observed at the output of the amplitude modulator, its carrier signal input being shorted and the control input driven with a square wave, referred to the nominal output level:

Switching transients   less than -45 dB

Frequency response of the earphone including the equalizer filter   50 Hz to 12 kHz ± 3 dB

4.4 List of Instruments Used in the Apparatus

Stimulus presentation, Figure 4.1 and 4.5:

CG   Sine Random Generator type 1024, Brüel and Kjaer

BF   Band-Pass Filters, used types and ranges:
    Third-Octave Filter: TZF 1/1, 228-284 Hz; TZF 1/2,
    900-1140 Hz; TZF 1/3, 3.6-4.5 kHz; WF VEB für
    Fernmeldewesen, Berlin

Octave Filter BP1, used ranges: 200-400 Hz, 800-1600 Hz, 3.2-6.4 kHz; RFT VEB Werk für Fernmeldewesen, Berlin

AM Amplitude Modulator developed in the Institute of Physics of the SAS in Bratislava, Czechoslovakia

EG Envelope Curve Generator developed in the Institute of Physics of the SAS in Bratislava, Czechoslovakia

EC Electronic Chronometer type MSM 1a, Radiometer, Denmark

A Decibel Attenuator type 992 E, Tesla, Czechoslovakia

EF Equalizer Filter designed in the Institute of Physics of the SAS in Bratislava, Czechoslovakia

EA Modulation Line Amplifier type ZU 281, Tesla, Czechoslovakia, modified to attain output impedance 0.1 Ohm

E Dynamic Earphone type DT 48, Beyer, W.Germany

V1 Low Frequency Electronic Millivoltmeter type BM 310, Tesla, Czechoslovakia

V2 Electronic Voltmeter type 2409, Brüel and Kjaer

S1 Oscilloscope type T 565, Křižík, Czechoslovakia

S2 Long Persistence Oscilloscope type OPD 250, Tesla, Czechoslovakia

MA Power Amplifier 10 W developed in the Institute of Electroacoustics in Prague, Czechoslovakia

ML Monitor Loudspeaker System developed in the Institute of Physics of the SAS in Bratislava, Czechoslovakia

Earphone calibration:

Sine Random Generator type 1024, Artifical Ear type 4153 with a 1/2" Microphone Cartridge type 4134, Cathode Follower type 2615, Level Recorder type 2304, Pistonphone type 4220, all Brüel and Kjaer.

Chapter V

DURATION DISCRIMINATION EXPERIMENT

The objective of the duration discrimination experiment was to investigate how the just noticeable increment in duration of acoustic pulses ΔT is affected by such stimulus parameters as the type of carrier signal, i.e., its bandwidth and center frequency, further by the form of the envelope curve, and by the duration of the pulse. Also the individual differences in duration discrimination between subjects were investigated.

5.1 Variable Factors

Next we will describe in more detail the stimulus factors and the levels selected for each factor in this experiment.

The first stimulus factor T was the effective duration T of the stimulus. This duration was defined for different envelope functions e(t) as the duration of the equivalent rectangular envelope with identical peak amplitude A

$$T = \frac{1}{A} \int_{-\infty}^{\infty} e(t) \, dt \quad . \tag{5.1}$$

For the rectangular envelope the effective duration is equal directly to its actual duration. The effective duration of the triangular envelope equals half its actual duration. Using Equation 5.1 the effective duration of the time

unlimited Gaussian envelope can be expressed in terms of its standard deviation σ and is equal to 2.51σ. The effect of the pulse duration factor T was investigated at these four levels: 8, 24, 80, and 240 msec.

The second stimulus factor E was the shape of the envelope curve of the acoustic pulses. It was investigated at three levels, as three types of envelope were used in this experiment: rectangular, isosceles triangle, and Gaussian pulse. These functions were selected in order to employ both smooth envelope and envelopes with different numbers of abrupt amplitude transitions. The envelope functions e(t) and the corresponding energy spectral densities E(ω) of these envelopes are given by the following formulae:

Rectangular envelope curve

$$e_R(t) = \begin{cases} 1 & \text{for} \quad 0 < t < T \\ 0 & \text{elsewhere} \end{cases}$$

$$E_R(\omega) = T^2 \, Sa^2(\omega \, \tfrac{T}{2}) \quad,$$

where the function Sa(x), known as the sampling function, stands for $Sa(x) = (\sin x)/x$.

Isosceles triangle envelope

$$e_I(t) = \begin{cases} 1 - \dfrac{|t|}{T} & \text{for } -T < t < T \\ 0 & \text{elsewhere} \end{cases}$$

$$E_I(\omega) = T^2 \, Sa^4(\omega \, \tfrac{T}{2}) \quad.$$

Gaussian envelope

$$e_G(t) = \begin{cases} e^{-\frac{1}{2}\left(\frac{t}{0.399\ T}\right)^2} & \text{for} \quad -T < t < T \\ 0 & \text{elsewhere} \end{cases}$$

$$E_G(\omega) = T^2 e^{-0.159\ T^2 \omega^2} \quad .$$

Parameter T in these expressions represents the effective stimulus duration. The envelope peak amplitudes are normalized to one. The spectrum of the Gaussian envelope $E_G(\omega)$ does not include the effect of truncation of the time function $e_G(t)$ at time instants $-T$ and $+T$. The graphical representations of the time and spectral functions for all three envelopes are given in Figures 5.1, 5.2, and 5.3. In order to normalize these spectra with respect to the stimulus duration, the frequency f, as the independent variable, is expressed in these figures in terms of the reciprocal value of the effective stimulus duration T, defined by Equation 5.1,

$$f = \frac{\omega}{2\pi} = \frac{1}{T} \quad .$$

The spectral bandwidth of the carrier signal was regarded as the third stimulus factor B. It was investigated at five levels. Pure tone, white noise, octave band noise, third octave band noise, and narrow band noise were used as carrier signals. The spectral bandwidth of the narrow-band noise was approximately equal to one tenth of the central band frequency. The frequency characteristics of the filters used to generate these band-noise carrier signals by filtering the white noise are given in Figure 4.2.
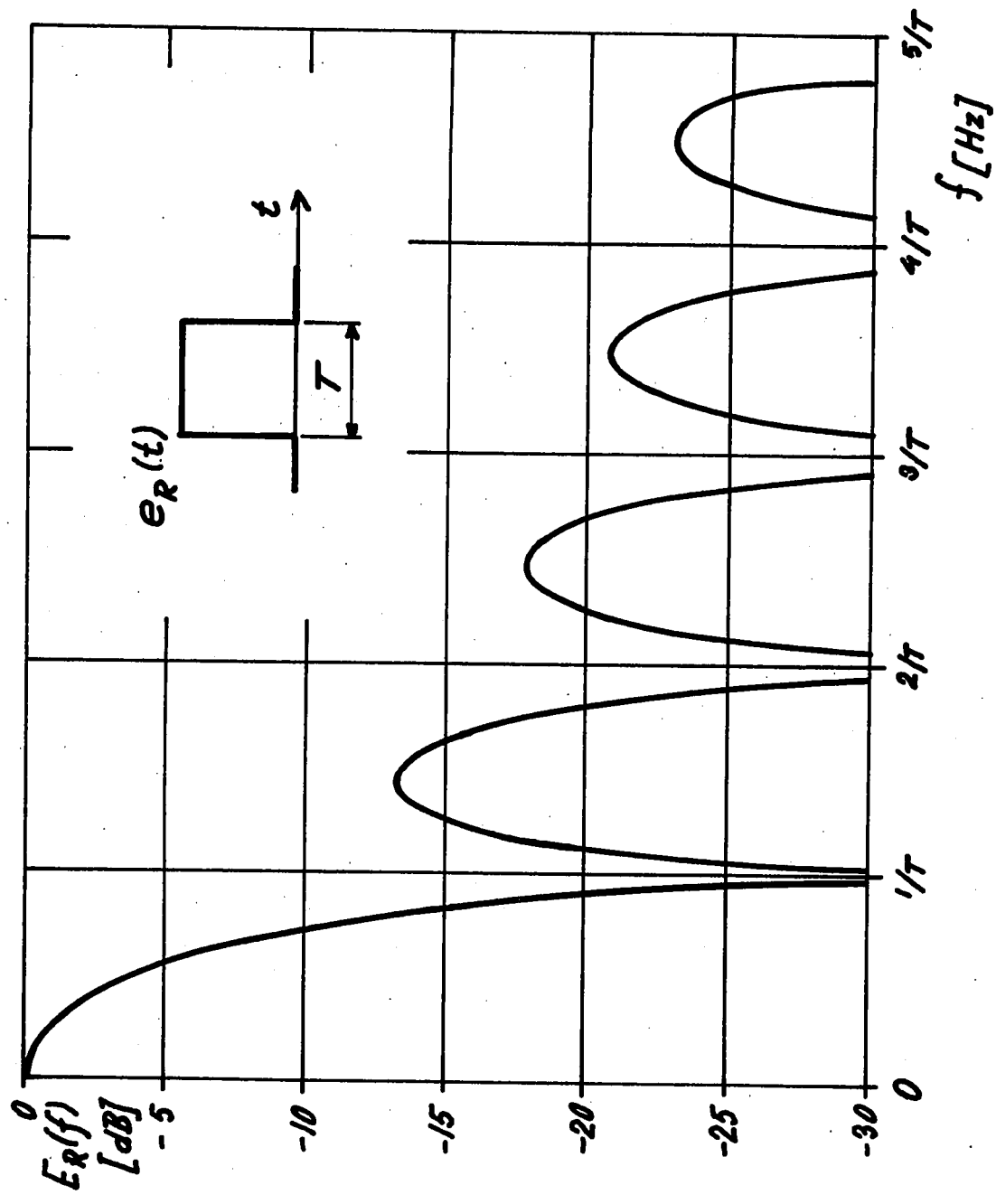
Figure 5.1 The rectangular envelope $e_R(t)$ and its energy spectral density $E_R(f)$.
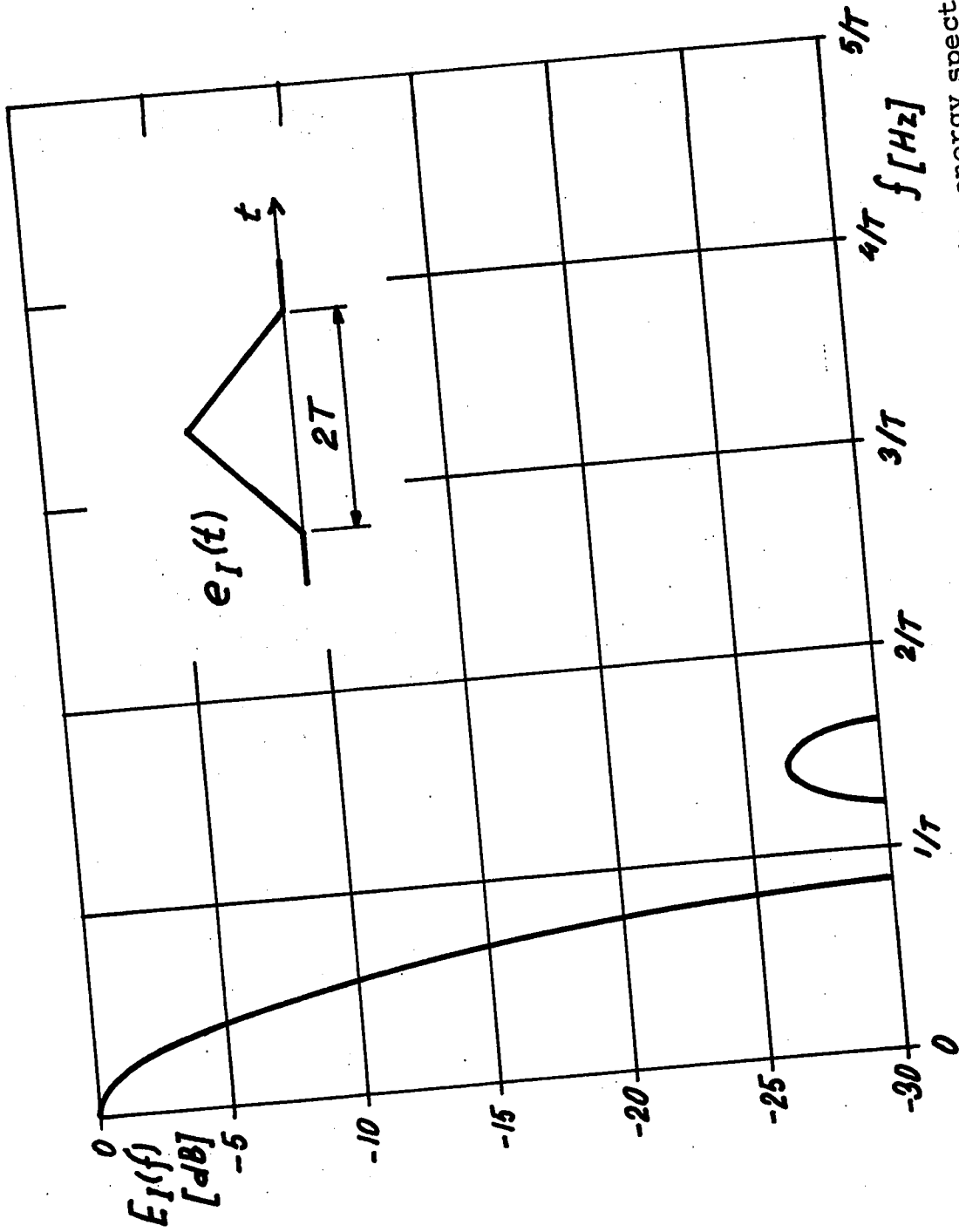
Figure 5.2   The isosceles triangle envelope $e_I(t)$ and its energy spectral density $E_I(f)$.
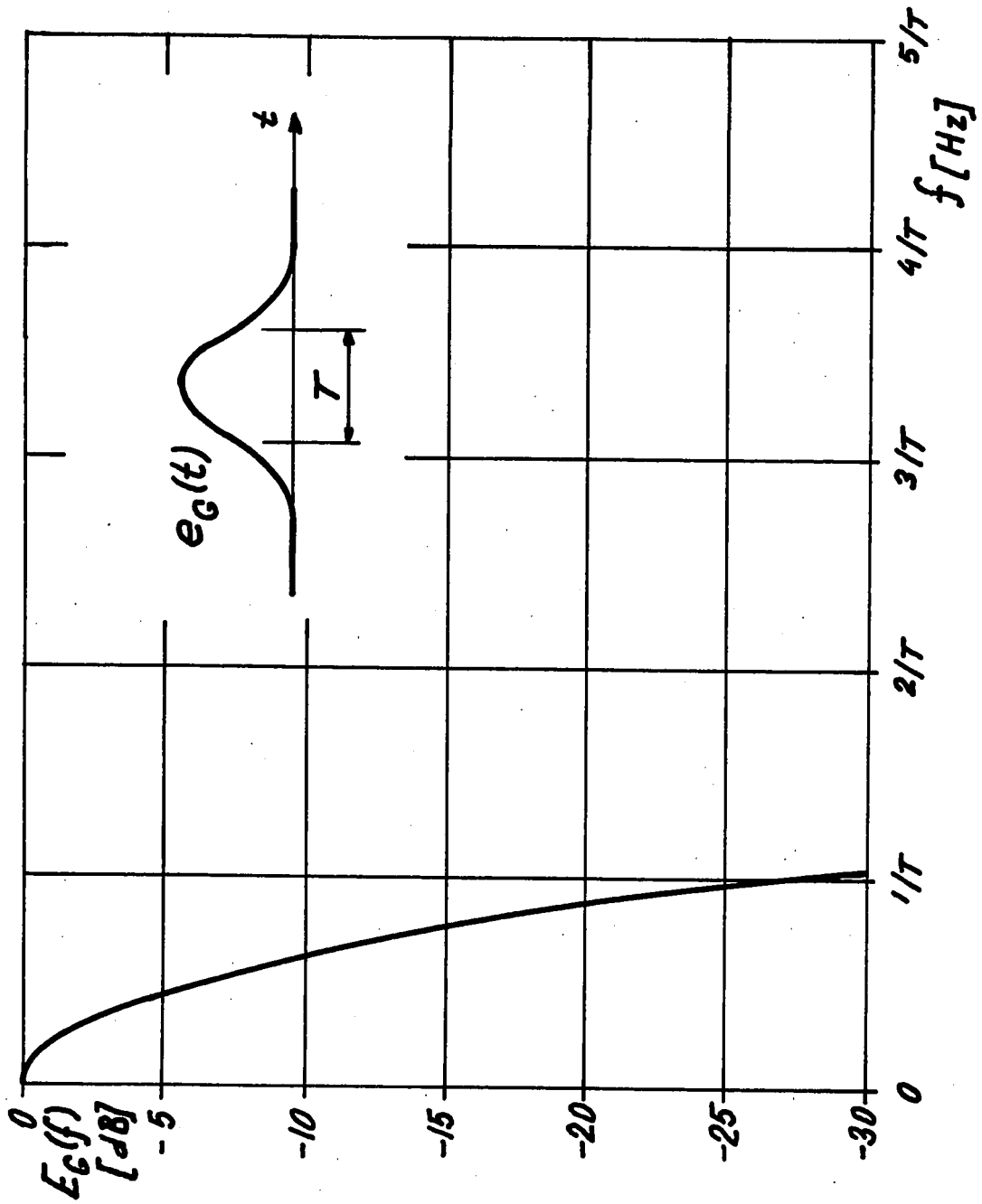
Figure 5.3 The Gaussian envelope $e_G(t)$ and its energy spectral density $E_G(f)$.

The fourth and last stimulus factor F in the duration discrimination experiment was the central frequency of the carrier signal. Its influence was analyzed at three levels. To cover the region of the audio range most relevant to speech and music perception, the central frequencies 250, 1000, and 4000 were used.

In addition to these four stimulus or signal factors, the individual differences between the six observers within the test group were also taken into consideration.

5.2 Data Processing

The data were processed using the analysis of variance procedure. This statistical technique is based on testing the null hypothesis in which it is presumed that the influence of a given factor or combination of factors on the result of the experiment falls below some predetermined level of significanse α. The "F-Ratio" in Tables 5.1, 5.2, and 5.3 represents the ratio of two variance estimates, the so-called "between-group" mean square, divided by the "within-group" mean square. These mean squares are given in columns "MS" in our tables. Depending on the chosen value of α, any systematic factor contribution may cause the value of the "F-Ratio" to exceed the acceptance limit value F_α. In such a case the null hypothesis is rejected and the factor or factor combination is considered as significant. The value of α represents the probability of false rejection of the null hypothesis. The values of F_α may be found in tables of F-distribution. For instance, in an experiment

investigating the influence of three factors B, F, and T, our model for the outcome x_{jkm} of a single trial for a given treatment, i.e., for a selected combination of factor levels j, k, and m, takes the form

$$x_{jkm} = m_{BFT} \, b_j \, f_k \, t_m \, bf_{jk} \, bt_{jm} \, ft_{km} \, e_{BFT} \, . \qquad (5.2)$$

The term m_{BFT} is the estimate of the overall true mean value, averaged over all levels of all factors. The coefficient b_j is due to the single effect of the factor B at its certain level j. Similarly f_k and t_m represent the contribution of factors F and T, respectively. The coefficients bf_{jk}, bt_{jm}, and ft_{km} denote the combined influence, called the first order interaction, of two factors. These terms arise only if the joint effect of the two particular factors acting simultaneously differs from the product of their effects taken separately. The last, error term e_{BFT}, is the only random term in this equation. It represents the effects of factors not included in the factorial design. Its mean value is one. For all levels j of an insignificant factor, for instance B, the corresponding terms are equal to one

$$b_j = 1 \qquad (5.3)$$

and can be omitted in Equation 5.2. The product of terms denoting the contribution of a single factor, for instance F, over all of its levels k is equal to one

$$\prod_k f_k = 1 \, . \qquad (5.4)$$

This form of mathematical model, Equation 5.2, was preferred to the conventional form in analysis of variance

with factor terms in sum. We believe that the model in which the factor influences appear as coefficients of the overall mean is more proper for description of perception phenomena. For instance, for the preceding three factorial design the analysis of variance uses the following model

$$\xi_{jkm} = \mu_{BFT} + \beta_j + \phi_k + \tau_m + \beta\phi_{jk} + \beta\tau_{jm} + \phi\tau_{km} + \varepsilon_{BFT}. \quad (5.5)$$

To account for the transformation between Equations 5.2 and 5.5 we have used natural logarithm of our data in the analysis of variance. The corresponding terms in these two equations have the same meaning. The error term ε_{BFT} is assumed to be normally distributed with zero mean and standard deviation σ_{BFT}. In our three factor illustration, Equation 5.5, six null hypotheses are tested: $\beta_j = 0$, $\phi_k = 0$, $\tau_m = 0$, $\beta\phi_{jk} = 0$, $\beta\tau_{jm} = 0$, and $\phi\tau_{km} = 0$.

Our complete duration discrimination experiment, incorporating all five factors E, B, F, T, and S, with each factor at all levels mentioned above, would result in a 3 x 5 x 3 x 4 x 6 factorial design, consisting of 1080 treatment combinations. Bearing in mind that for each treatment twenty observations had been planned, we arrived at an excessively large total number of observations. For these reasons we decided to split the duration discrimination experiment into three partial experiments. In each of these partial experiments only four factors appeared at variable levels: EFTS, EBTS, and BFTS. The remaining fifth factor was kept at a constant level. Factors S and T were employed in all three cases, as the same group of subjects was tested

in all partial experiments at all four stimulus durations.

While the signal factors can be regarded as fixed-effect factors, the subject factor is a random-effect factor. So that a mixed model of analysis of variance applies for all our experiments. Usually in such a design the significance of the random-effect factor is not tested, as there is no appropriate error term available. To compare the observer's performance in the duration discrimination experiments, we have tested the random-effect subjective factor S against an error term in which all the interactions of the subject factor were pooled.

Unless otherwise stated, in our further discussions we will regard as significant the 1 per cent level.

In our equations the contributions of factors E, B, F, T, and S at levels i, j, k, l, and m, are denoted by letters η, β, ϕ, τ, and ω in the analysis of variance models (Equation 5.5) and by letters v, b, f, t, and s in the psychophysical models (Equation 5.2), respectively.

5.3 Results

The table and graph presentations of results will be described in more detail for the first partial experiment. The same data arrangement applies also for the results of the partial experiments EFTS and EBTS. In order to simplify the graphical representation of the results, the data presented in all diagrams were averaged along the group of observers, despite the fact that the subjective factor S proved to be significant in all three partial experiments.

To minimize the possible trends resulting from learning, the succession of the treatment combinations from each partial experiment and the partial experiments themselves were randomized. In those cases where some treatment combination was common to two or three partial experiments, it was tested only once with each subject.

5.3.a Partial Experiment EFTS

In the first partial experiment the spectral bandwidth of the carrier signal, i.e. factor B, was excluded from investigation, as here only pure tones were used as stimuli. Factors E, F, T, and S were variable. Table 5.1 was obtained as a result of subjecting the data, both absolute and relative, to the analysis of variance. In the column "Source" of this table the factors and factor interactions are listed. The column "d.f." shows the number of degrees of freedom for particular factor or factor interaction. The third and fifth columns "MS" show the mean squares for the absolute and relative data, respectively. By the relative data is understood the relative difference limen for duration, i.e. the Weber fraction of the absolute difference limen ΔT, divided by the stimulus duration T. Columns four and six present the "F-Ratio" mentioned above. In these columns significance at the 1 per cent level is indicated by two asterisks, and at the 5 per cent level by one asterisk.

According to Table 5.1 only single factors T and S produce significant effect on the duration discrimination. Thus the detection of the increment in the duration of

E F T S

Source	d.f.	Absolute Data		Relative Data	
		MS	F-Ratio	MS	F-Ratio
S	5	0.600	15.752**	0.600	14.999**
E	2	0.007	0.101	0.007	0.103
F	2	0.316	3.254	0.316	3.254
T	3	120.527	3653.455**	0.835	25.479**
EF	4	0.057	0.832	0.057	0.832
ET	6	0.023	0.749	0.023	0.739
FT	6	0.069	2.299	0.069	2.286
EFT	12	0.040	1.912	0.040	1.900
SE	10	0.072		0.072	
SF	10	0.097		0.097	
ST	15	0.033		0.033	
SEF	20	0.069		0.069	
SET	30	0.031		0.031	
SFT	30	0.030		0.030	
SEFT	60	0.021		0.021	
S pooled	175	0.038		0.038	

**$p < 0.01$

*$p < 0.05$

Table 5.1   Analysis of variance for the partial experiment EFTS.

sinusoidal stimuli with different envelope curves can be modeled for absolute jnd in correspondence with Equation 5.2 by the formula

$$x_{EFTS} = m_{EFTS} \quad t_m \quad s_n \quad e_{EFTS}.$$

The data were averaged over the group of subjects and, applying the results of the analysis of variance, also over the insignificant factors E and F and plotted in Figures 5.4 and 5.5 for absolute and relative jnd in duration, respectively.

5.3.b   Partial Experiment EBTS

In the second partial experiment the levels of factors E, B, T, and S were variable. The level of factor F was constant, as here the central frequency of the carrier signal for all stimuli was 1000 Hz.

The results of analysis of variance listed in Table 5.2 proved that, from all the variable factors employed in this partial experiment, only the stimulus duration factor T and the individual factor S are significant. The model for this partial experiment takes the form

$$x_{EBTS} = m_{EBTS} \quad t_m \quad s_n \quad e_{EBTS}.$$

Figures 5.6 and 5.7 display the data averaged over the group of observers and over the nonsignificant factors.

In this partial experiment, interactions ET and EBT are significant at the 5 per cent level.

5.3.c   Partial Experiment BFTS

In the third and last partial duration discrimination experiment the envelope curve factor E was eliminated by
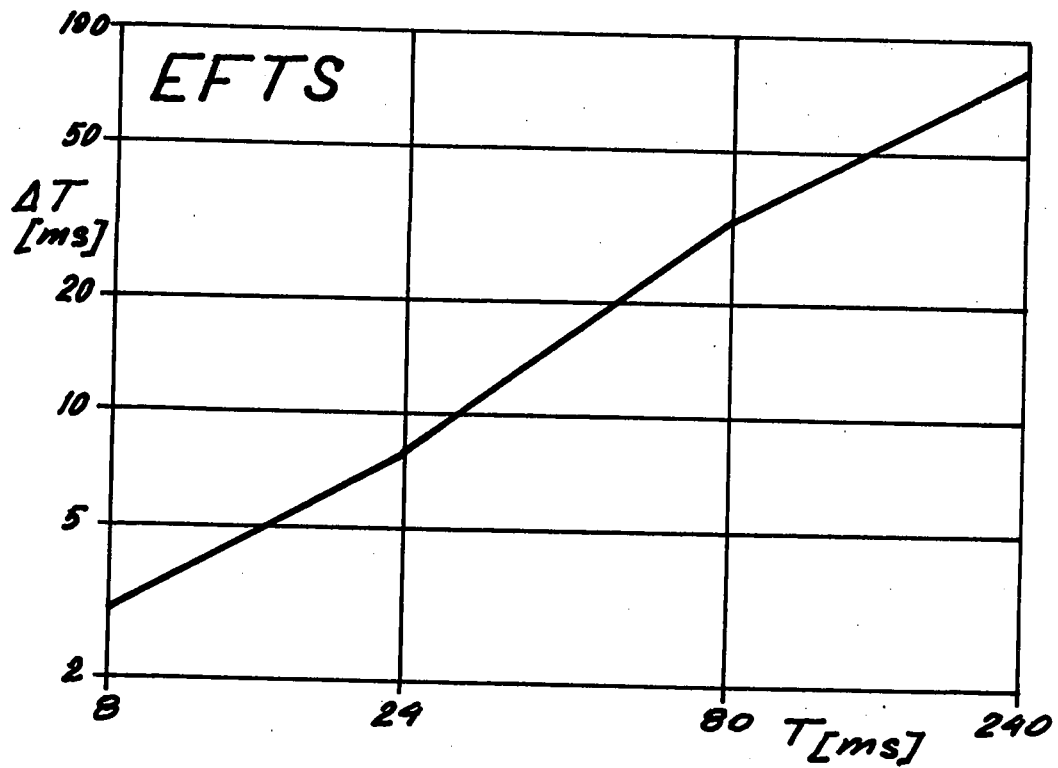
Figure 5.4   Partial experiment EFTS. The just noticeable

increment in duration ΔT as a function of the

duration T of tone stimuli.

Figure 5.5 Partial experiment EFTS. The relative just
noticeable increment in duration ΔT/T and
the relative standard deviation σ as a
function of the duration T of tone stimuli.

E B T S

Source	d.f.	Absolute Data		Relative Data	
		MS	F-Ratio	MS	F-Ratio
S	5	1.712	28.438**	1.712	36.949**
E	2	0.016	0.322	0.016	0.320
B	4	0.118	1.791	0.118	1.793
T	3	193.759	2787.498**	1.063	15.332**
EB	8	0.072	0.845	0.072	0.846
ET	6	0.109	2.609*	0.109	2.606*
BT	12	0.066	1.873	0.066	1.863
EBT	24	0.060	1.804*	0.060	1.792*
SE	10	0.050		0.050	
SB	20	0.066		0.066	
ST	15	0.070		0.069	
SEB	40	0.085		0.085	
SET	30	0.042		0.042	
SBT	60	0.035		0.035	
SEBT	120	0.033		0.034	
S pooled	295	0.046		0.046	

**p<0.01

*p<0.05

Table 5.2   Analysis of variance for the partial
experiment EBTS.

Figure 5.6 Partial experiment EBTS. The just noticeable
increment in duration ΔT as a function of the
duration T of tone and noise stimuli with
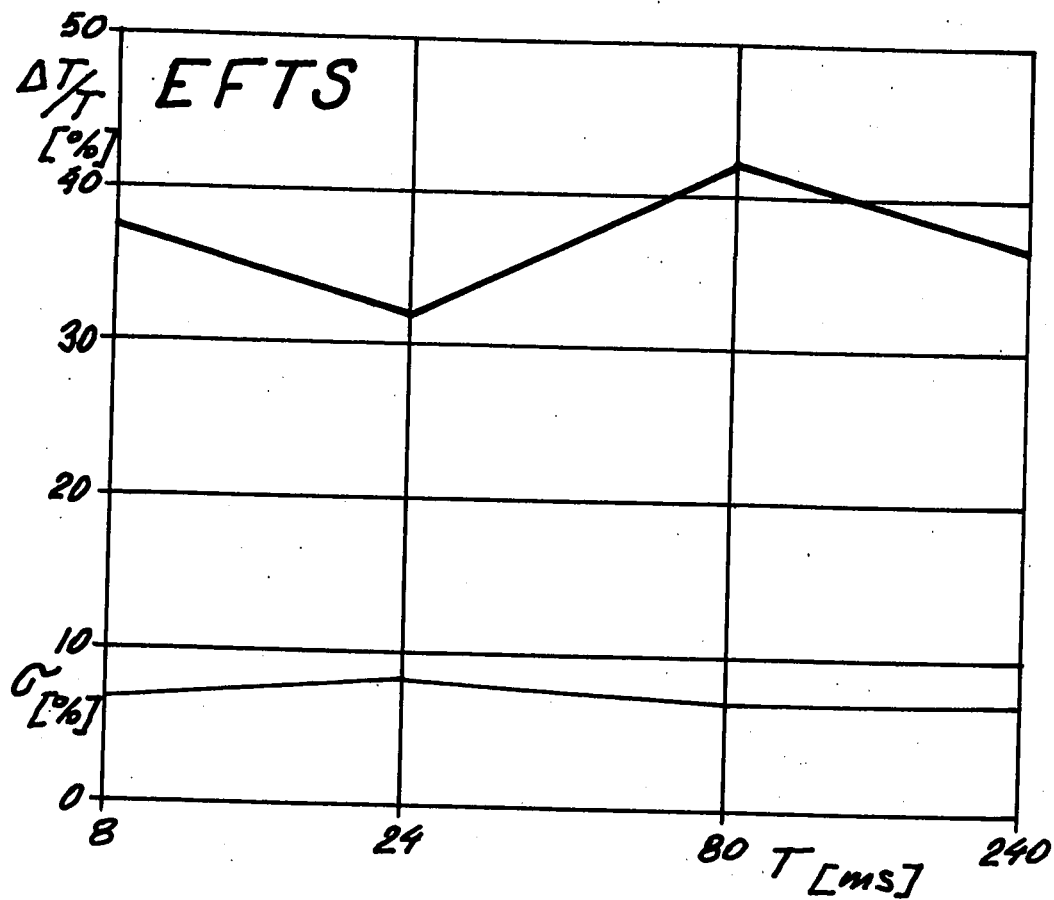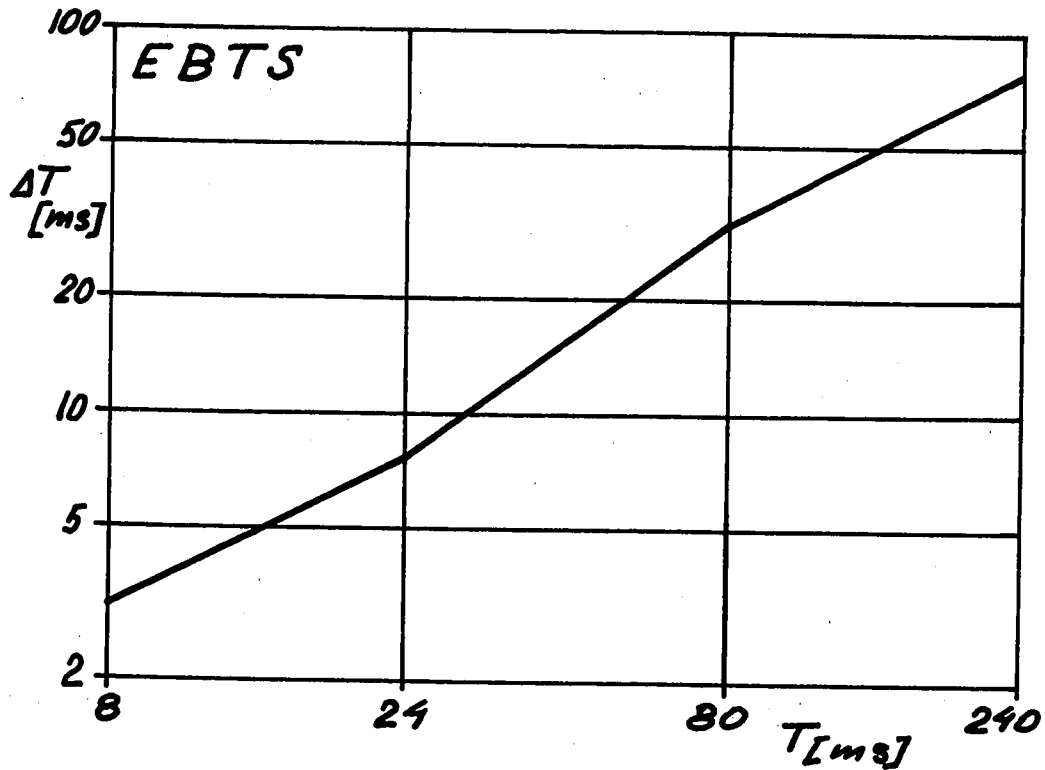carrier central frequency 1 kHz.

Figure 5.7 Partial experiment EBTS. The relative just
noticeable increment in duration $\Delta T/T$ and the
relative standard deviation σ as a function
of the duration T of tone and noise stimuli
with carrier central frequency 1 kHz.

using stimuli of rectangular envelope only. According to Table 5.3 both the carrier bandwidth factor B and the carrier frequency factor F proved to be nonsignificant. Significant are only the factors T, S and the second order interaction BFT. The interaction BT is significant at the 5 per cent level. As usual, the data have been averaged and arranged in Figures 5.8a and 5.9. In order to interpret the interaction BFT, the data displayed in Figure 5.8b are averaged only across the subject factor S. Thus the partial experiment BFTS can be modeled as

$$x_{BFTS} = m_{BFTS} \; t_m \; s_n \; bft_{jkm} \; e_{BFTS}.$$

5.3.d Resulting Duration Discrimination Model

To obtain the resulting model, which takes into consideration all five factors, the factors and factor interactions significant at the 1 per cent level from Tables 5.1, 5.2, and 5.3, were compiled into Table 5.4. In this table letter "s" indicates significance and letter "n" non-significance of the corresponding factor or factor combination. From its last column we can determine the significant factors and factor interactions of the resulting five factor model. The good correlation between the three partial models is apparent.

The resulting five factor model of duration discrimination can be expressed as

$$x_{ijkmn} = m_{EBFTS} \; t_m \; s_n \; bft_{jkm} \; e_{EBFTS}.$$

The same factor analysis as just described was also carried out for the relative jnd in duration and yielded

B F T S

		Absolute Data		Relative Data	
Source	d.f.	MS	F-Ratio	MS	F-Ratio
S	5	1.409	36.069**	1.409	30.936**
B	4	0.114	1.312	0.114	1.312
F	2	0.186	1.812	0.186	1.812
T	3	184.011	3081.750**	1.374	23.108**
BF	8	0.084	1.623	0.084	1.623
BT	12	0.076	2.068*	0.076	2.058*
FT	6	0.030	1.014	0.030	1.003
BFT	24	0.046	2.027**	0.046	2.007**
SB	20	0.087		0.087	
SF	10	0.102		0.102	
ST	15	0.060		0.059	
SBF	40	0.052		0.052	
SBT	60	0.037		0.037	
SFT	30	0.030		0.030	
SBFT	120	0.022		0.023	
S pooled	295	0.039		0.039	

**p<0.01

*p<0.05

Table 5.3   Analysis of variance for the partial
experiment BFTS.
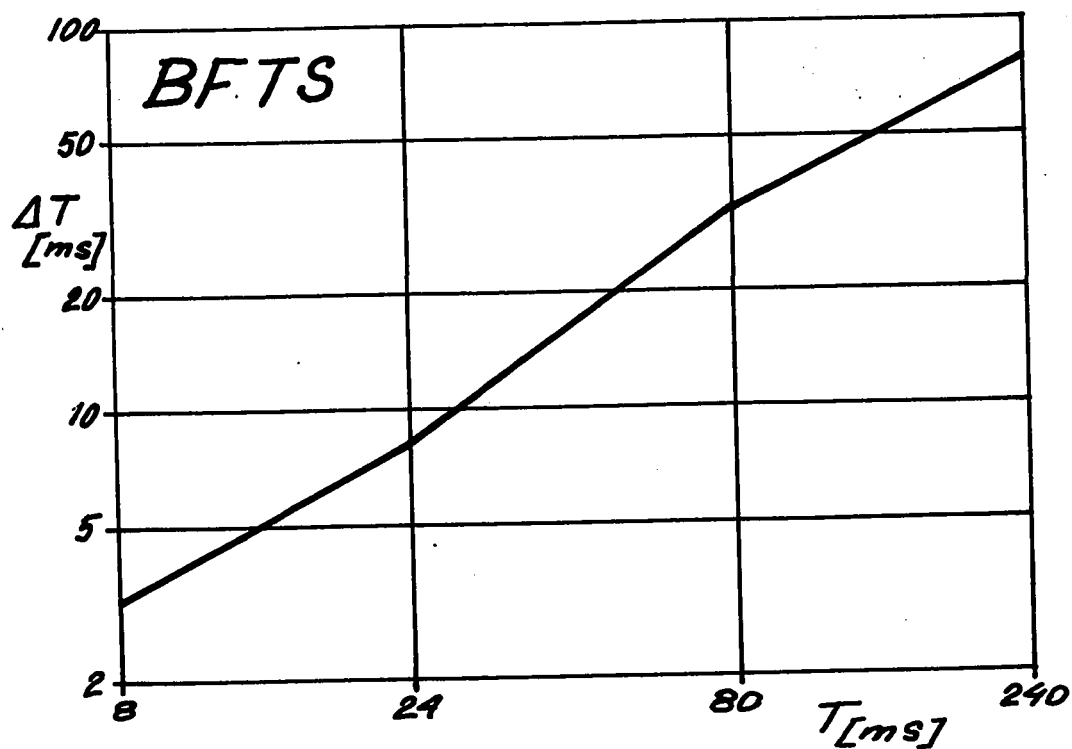
Figure 5.8a Partial experiment BFTS. The just noticeable
increment in duration ΔT as a function of
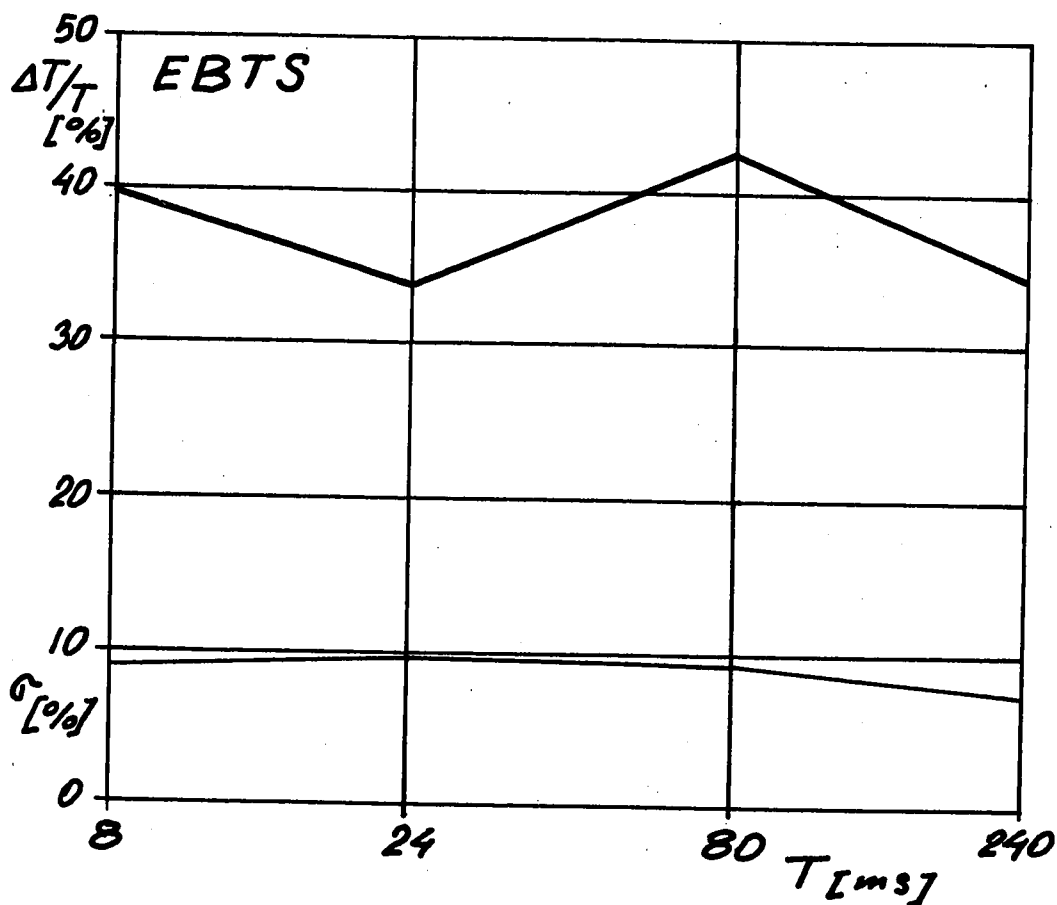the duration T of noise stimuli with
rectangular envelope.
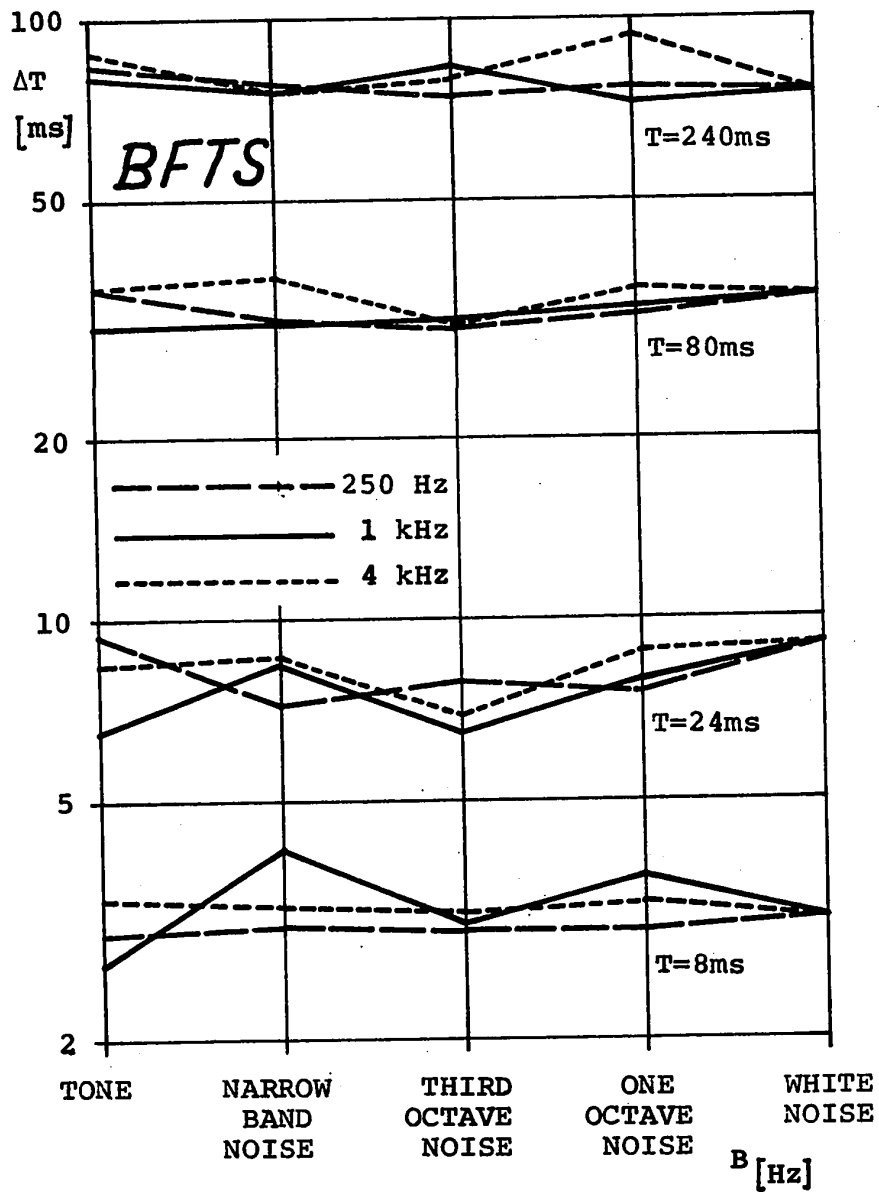
Figure 5.8b   Partial experiment BFTS. The just noticeable

increment in duration ΔT of tone and noise

stimuli as a function of all three signal

factors: carrier bandwidth B, carrier

frequency F, and stimulus duration T.
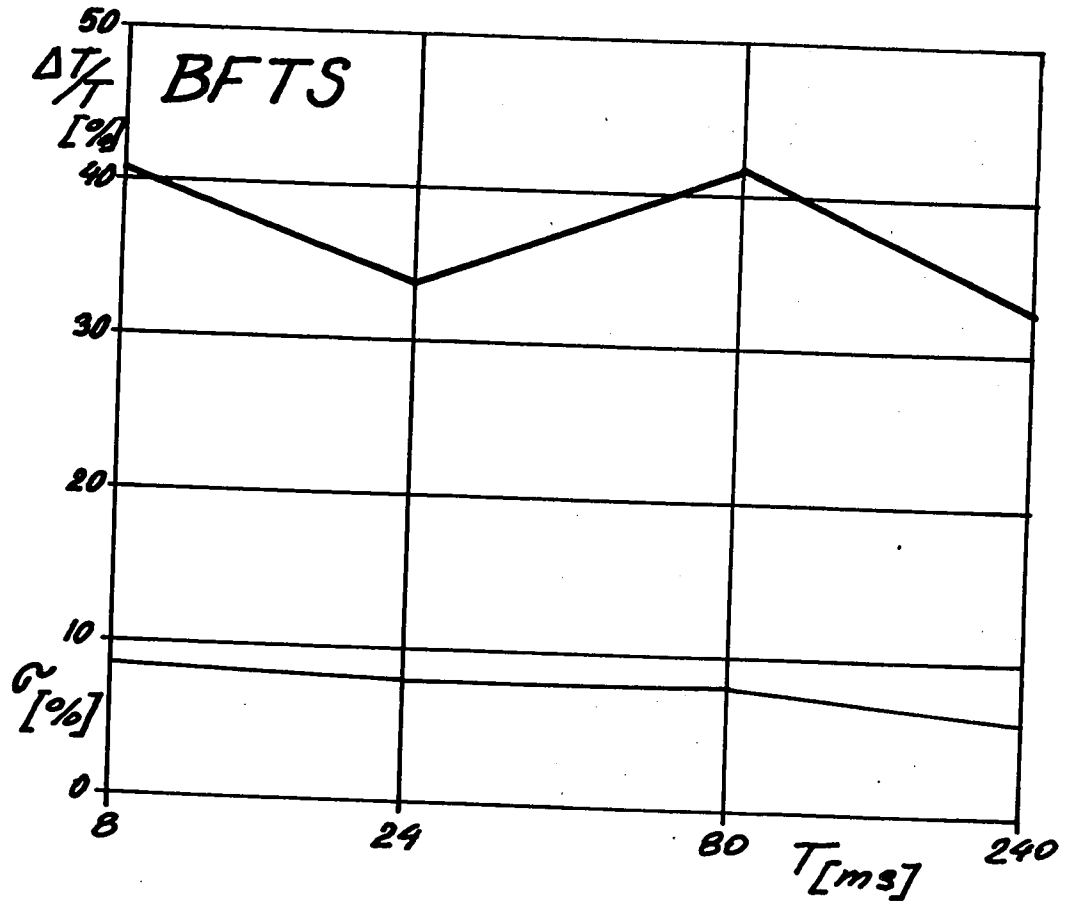
Figure 5.9   Partial experiment BFTS. The relative just
              noticeable increment in duration ΔT/T and
              the relative standard deviation σ as
              a function of the duration T of noise stimuli
              with rectangular envelope.

Source	Partial experiment			Resulting model EBFTS
	EFTS	EBTS	BFTS	
S	s	s	s	s
E	n	n	-	n
B	-	n	n	n
F	n	-	n	n
T	s	s	s	s
EB	-	n	-	n
EF	n	-	-	n
ET	n	n	-	n
BF	-	-	n	n
BT	-	n	n	n
FT	n	-	n	n
EBF	n	-	-	n
EBT	-	n	-	n
BFT	-	-	s	s

Table 5.4   Results of the analysis of variance from
the three partial experiments compiled into
the resulting five factor duration
discrimination model EBFTS.

exactly the same results, included in Tables 5.1, 5.2, and 5.3. The fact that in this case the stimulus duration factor T is also significant indicates that the difference limen for stimulus duration is not a constant fraction of the stimulus duration in the range of stimulus parameters examined in our experiment. In other words, a significant discrepancy with the Weber-Fechner law was observed.

5.4 Experiments of Other Investigators Related

to Duration Discrimination

Our discussion and interpretation of the duration discrimination of human observers will be based, besides on our data, also on the results of the three following experimenters.

L.A. Chistovich (1959) used as stimuli empty time intervals bounded by sharp sound clicks. Subject's task was to determine, in a two-alternative forced-choice (2AFC) procedure, whether the second time interval was of the same duration or longer than the first one. The interstimulus interval of 10 seconds was unusually long for this kind of experiment. The just noticeable increment of the empty time interval duration is plotted in Figure 5.10.

S. Rochester (1970) in her duration discrimination experiments used as one type of stimulus rectangular bursts of noise submerged in noise background. Stimulus and background noise were generated by two independent white noise sources of the same 64 dB sound pressure level. In such an arrangement the stimulus was represented by a 3 dB increment in the continuous background noise level. Both the constant stimulus variant and the adaptive DELTA variant of the 2AFC psychophysical method was used to find the stimulus duration increment required for 75 per cent and 90 per cent correct recognition. These data are also presented in Figure 5.10. In order to compare the results of duration discrimination
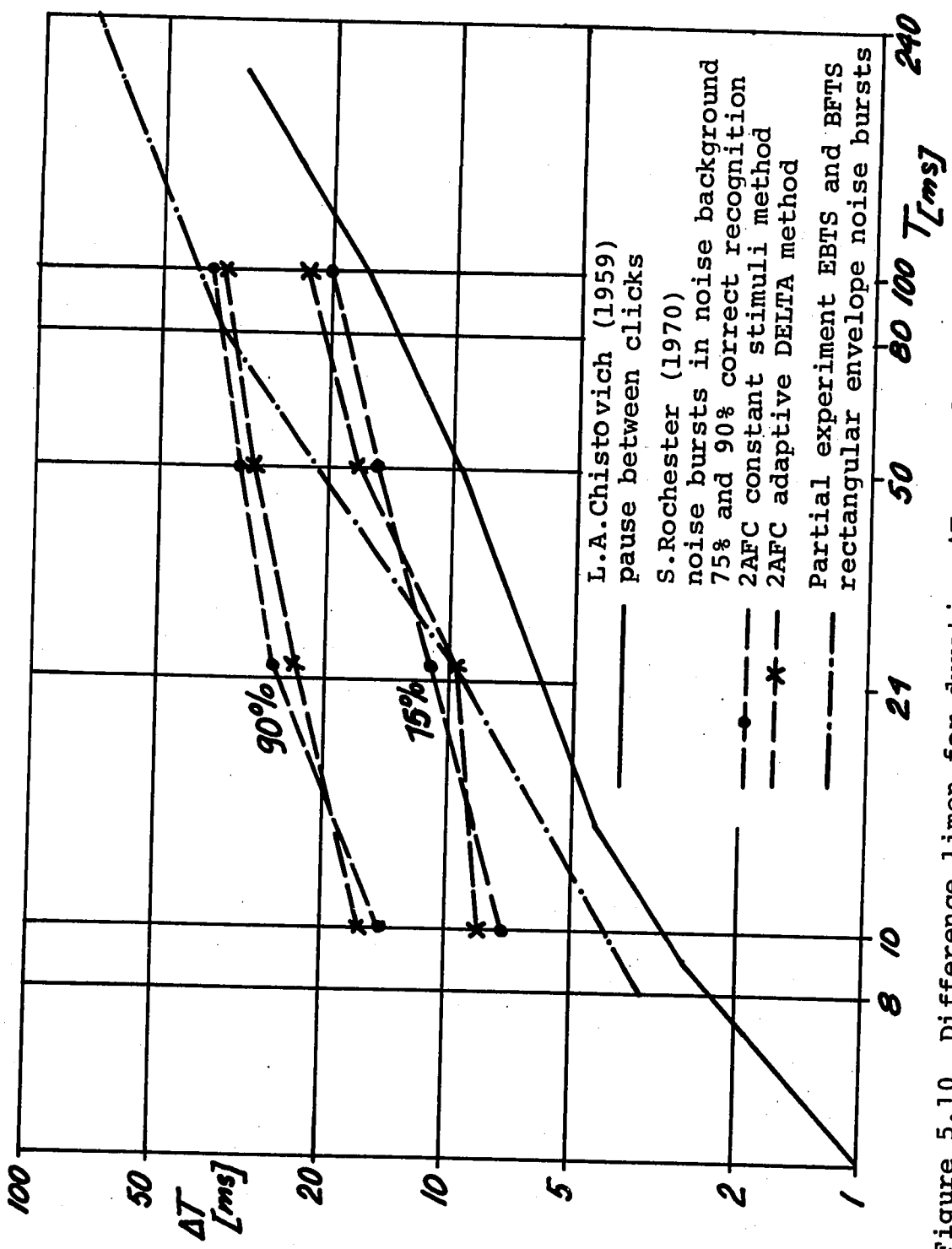
Figure 5.10 Difference limen for duration ΔT as a function of stimulus

duration T. Comparison with previous results.

of white noise bursts with and without background noise, our data for white noise stimuli with rectangular envelope are plotted in the same diagram as well.

C.D. Creelman (1962), in an extensive series of experiments, investigated the effects of the stimulus duration and intensity on duration discrimination. The 1 kHz pure tone stimuli were presented in a continuous white noise background. Three stimulus parameters were variable: stimulus base duration, duration increment of the stimulus, and stimulus level. In separate experiments the value of one of these parameters was varied, while the other two were kept constant at arbitrary values. The 2AFC psychophysical method was used to collect the data. The inter-stimulus interval was 0.8 seconds. Unfortunately, the stimulus and background noise levels in this paper are specified only as voltages at the input to the observer's earphone. So we do not know how far the stimuli were situated above the hearing threshold, i.e. above the level of the internal neural noise. All the results were expressed in terms of the sensitivity index d', as the duration discrimination was interpreted here as a signal detection task.

In the variable stimulus level experiment, the detection of the duration increment improved rapidly with increasing stimulus level as long as it was comparable to the level of the background noise. When the stimulus level was high enough to make the stimulus sound loud and clear above the noise background, this dependence disappeared. In the

constant stimulus base experiment, no interaction between
increment duration and stimulus level was observed, although
the lower stimulus level was reflected in poorer performance.
In the constant stimulus increment experiment an apparent
interaction between the stimulus level and the base duration
was found: in general, higher stimulus levels result in
a better duration discrimination and the stimulus base
duration increase brings with it deterioration in detect-
ability. But, the lower the signal level, the less pro-
nounced this dependence is, as the data curve becomes less
steep. By extrapolating the data above the range of the
base durations tested, we can assume that the influence of
the stimulus level would disappear for base durations some-
where around 2 to 3 seconds.

To simulate the observer's performance, C.D. Creelman
modeled the stimulus duration measuring process as a
separate and independent mechanism consisting of a pulse
source and a pulse counter. The pulse source was assumed to
consist of a large number of independent pulse generators
with firing rates governed by the Poisson probability dis-
tribution. By comparison with the experimental results the
mean firing rate of the pulse source was estimated to be
between 2,700 and 10,000 pulses per second. The counter counts
the number of pulses generated during the stimulus duration.
The subject's decision is based on likelihood ratio between
probability distributions for the stimulus of a base dura-
tion and for the stimulus of a base plus increment duration.

The effects of imperfect memory and uncertainty in the
instant of stimulus onset and termination were also incor-
porated in this mathematical model.

The validity of this model was verified in an addi-
tional duration discrimination experiment in which the base
to increment ratio was fixed at 8:1 and the stimulus level
was varied to keep the energy in the stimulus increment
constant. Subjects' performance was very close to the data
predicted by the model. Significant discrepancy appeared
only for duration increments as short as 5 msec. Here the
observers showed much poorer performance than expected
according to the model.

5.5 Discussion of the Duration Discrimination Results

As our measurements indicate, the following factors
are relevant for the size of the just noticeable difference
in duration of acoustic stimuli. As had been expected, the
most pronounced is the influence of the stimulus duration
factor T. Second in order is the factor of individual
differences between observers S. Third is the second order
interaction BFT. All these factor contributions are sig-
nificant at the 1 per cent level. Neither the envelope
curve shape E, nor the spectral bandwidth of the carrier
signal B, nor its central frequency F, exhibit a significant
influence on the duration discrimination.

At the 5 per cent level of significance appear also
the first order interactions ET and BT from the partial
experiments EBTS and BFTS, respectively, as well as the

second order interaction EBT.

The results of the significance analysis for absolute and relative data are identical.

In discussing the performance of the human ear in the duration discrimination experiments, we shall first mention the clues available for this discrimination. Second, we shall delimit ranges of action in which these clues can actually support the discrimination process.

First to be considered is the duration of the stimulus, the elapsed time during which the signal is audible. Usually it is the time between the stimulus onset and offset transients.

Perception of time by human observers has not yet been studied systematically. Up to now, no subjective time scales for the "time sense" have been established that are analogous to psychophysical scales for pitch and loudness, reflecting perception of frequency and intensity, respectively. A subjective time unit has not yet been widely accepted, either. It is known that temporal judgment is subject to large individual differences and that it depends to some degree on organismic variables, on stimulus parameters, and on simultaneous gross stimulation of several sensory channels. The history of research in the duration discrimination field is reviewed in the paper of C.D. Creelman (1962). E. Zwicker (1969/70) published a pilot study recently on the relationship between objective and subjective duration of 4kHz tone bursts, white-noise bursts, pauses delimited by such bursts,

and time intervals delimited by two sharp clicks. He found, that at the same objective duration, the sensation of duration evoked by the pause or by the interclick interval was only about one half of the subjective duration of the tone or noise burst. For short 4kHz tones the subjective duration increased with intensity.

The next clue might be called the detected quantitative change in the temporal structure of the stimulus. Let us suppose the stimulus contains two transients, such as, for instance, the onset and offset step of the rectangular tone pulse. Up to a certain duration, such a stimulus is perceived as a single time event, with no subjective duration at all. Prolonging the stimulus at a particular duration, the observers begin to resolve the two transients as two separate events in time. At this point the quantitative change in the sensation takes place.

Another possible clue is stimulus loudness. To estimate stimulus intensity, hearing integrates neural activity evoked by the stimulus. According to J.J. Zwislocki (1960)(1969) the time constant of this process is between about 200 msec and 100 msec, for near threshold and sufficiently loud stimuli, respectively. Consequently, stimulus loudness can serve as a clue only for short stimuli with upper duration limit of about twice the value of the aforementioned time constant.

The detectability of amplitude modulation, frequency modulation, and also simultaneous amplitude and frequency

modulation was investigated by E. Zwicker (1952), and by D. Maiwald (1966), who also suggested a functional model for the detection of amplitude and frequency modulation (1967a), (1967b). The results of these experiments indicate that independently of the modulation type, the modulation is detected as soon as the variations of the excitation pattern of the basilar membrane at any point along its length exceeds 1 dB.

B. Leshowitz (1971) studied the ability of human observers to discriminate between a single 20 μsec rectangular pressure pulse stimulus and a pair of 10 μsec rectangular pulses. The variable parameter was the time delay between the end of the first pulse in the pair and the start of the second pulse. A typical value of this time separation was 10 μsec, of the same order as the stimulus pulses. His results indicate that the ear is sensitive to differences in spectral contours of the stimuli. For discrimination a 1 to 2 dB difference in the upper frequency region of the spectra was sufficient. This sensitivity remained unchanged even when the stimulus amplitude was randomly varied over a range of 6 dB.

The frequency selectivity mechanism can aid in duration discrimination only in the cases when either the stimulus itself is short, or it contains marked short transient components. Under "short" it is understood in this case either equal to or shorter than the time window involved in the short-time Fourier frequency analysis performed in hearing.

At the level of the mechanical time-frequency analysis, as carried out by the basilar membrane, the duration of the time window is of the order of several milliseconds. As this analysis is constant Q in character, the duration of the time window is inversely proportional to the stimulus frequency (see Equation 2.2). At the level of subsequent neural analysis, which offers further sharpening in frequency selectivity, the time window is longer, about 100 to 200 msec (see Equation 6.1). Thus, at the neural stage the stimulus duration range limitations for the spectral contour clue are about the same as for the loudness clue.

In general, it is very difficult, if not impossible, to separate the duration, loudness, and spectral factors in duration discrimination. For instance, by prolonging the stimulus envelope and simultaneously keeping its amplitude constant, the stimulus energy increases, which leads to increased loudness. At the same time the frequency spectrum of the stimulus becomes narrower.

The general shape of our resulting curves suggests that two different mechanisms contribute to duration discrimination. This is particularly apparent for the data displayed as the relative jnd in duration. One local minimum can be expected around stimulus duration 24 msec. The other one probably lies above 240 msec, which is above the highest stimulus duration used in our experiments. As we have measured the discrimination only in the range from 8 to 240 msec and for four stimulus durations only, we have no knowledge of how the

hearing performs between these points and outside of this range. So the limits of action of these different mechanisms can be specified only roughly and should be regarded with adequate caution.

The region of optimal performance observed in our experiments around base durations 24 msec can be interpreted as the region where the actions of the clock and counter mechanism, on the one hand, and the energy integration mechanism and the discrimination between spectral contours, on the other hand, overlap.

Our 24 msec value of optimum duration discrimination is in very good agreement with experiments on the discrimination of frequency transitions conducted by I. Nábělek and I.J. Hirsh (1969). Their results convincingly indicate that the smallest relative difference limen in duration of linear frequency glides, forming the initial part of a tone burst, lies between 20 and 28 msec. This optimal range of transition durations depends neither on the magnitude of the frequency interval of the glide, nor on the frequency region of the stimulus. This coincidence can be attributed, most probably, to the fact that analyzing properties of hearing were developed in close relationship with the development of speech perception and production. For proper recognition of speech sounds both frequency and amplitude transients of a duration of this order are very important.

5.6 Duration Discrimination of Short Stimuli

In the region of short stimuli, with the optimum duration around 24 msec, the mechanism responsible for the duration discrimination is probably based either on the perceived change in stimulus quality or on the detected change in its energy, or on both.

Obviously, in L.A. Chistovich's experiment (1959), energy integration can not serve as a clue for duration discrimination. Her empty intervals were marked-off by two clicks, and the energy of both clicks remained, for stimuli of different durations, always the same. For short stimuli when the delay between these two clicks is shorter than the time window of the basilar membrane, both clicks are included in the short-term time-frequency analysis and their short-term Fourier spectrum is dependent on their time separation. Here the duration discrimination is based on detected differences between these spectra. Duration of the short-term frequency analysis time window of the mechanical structure of the basilar membrane is inversely proportional to frequency and of the order of milliseconds for high frequencies and of ten milliseconds for low frequencies. For longer separations of clicks the clock and counter mechanism takes over the duration discrimination. In the curve of L.A. Chistovich in Figure 5.10 an apparent change in steepness around click separation 15 msec indicates a transition from the sphere of one mechanism into that of the other.

The fact that we have employed as stimuli pulses having different spectral bandwidths of the carrier signal allows us to isolate the contribution of the spectral clue to duration discrimination. Our stimuli s(t) were generated in the time domain as a product between the carrier signal c(t) and the envelope function e(t)

$$s(t) = e(t) \, c(t).$$

The corresponding Fourier spectra of these signals are

$$E(f) = \mathcal{F}\left\{e(t)\right\},$$

$$C(f) = \mathcal{F}\left\{c(t)\right\}.$$

In order to calculate the frequency spectrum S(f) of the resulting modulated signal s(t), the convolution theorem in the frequency domain can be used. According to this theorem, the frequency spectrum S(f) of a product of two time functions e(t) and c(t),

$$S(f) = \mathcal{F}\left\{e(t) \, c(t)\right\},$$

is the convolution of their Fourier spectra in the frequency domain

$$S(f) = \frac{1}{2\pi} \left\{E(f) * C(f)\right\}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} E(u) \, C(f-u) \, du$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} C(v) \, E(f-v) \, dv \qquad (5.6)$$

In our experiments we have used five different spectral bandwidths of the carrier signal c(t). At one extreme, the carrier was a pure tone of frequency f_c. Thus the resulting spectrum S(f) is equal to the spectrum of the envelope

function E(f), shifted by the carrier frequency f_c

$$S(f) = E(f_c \pm f) .$$

Due to the existence of the upper and lower sidebands around the carrier frequency f_c, the absolute spectral bandwidth of S(f) is double the spectral bandwidth of the envelope E(f). As long as the envelope duration is not extremely short, the spectral bandwidth of E(f), and consequently also of S(f), is narrower than the audio frequency range and thus offers to the receiver a possible spectral, or frequency contour, clue.

At the other extreme, white noise was used as carrier signal. In this case, both C(f) and S(f) cover, uniformly, the whole audio range. Neither S(f) nor any short-term Fourier spectrum of the stimulus offers any spectral clue to the observer. The analyzer has to rely solely on noise level variations in time.

To evaluate the weight of the spectral clue available from stimuli of different carrier bandwidths, let us have a closer look at the significance of the carrier spectral bandwidth factor B. We find it to be significant in the partial experiment BFTS only in the second order interaction EBT, see Figure 5.8b. The influence of this interaction of factors is difficult to interpret, as no systematic trend in the curves can be recognized. But in the same partial experiment BFTS, see Table 5.3, the first order factor interaction BT is significant, but only at the 5 per cent
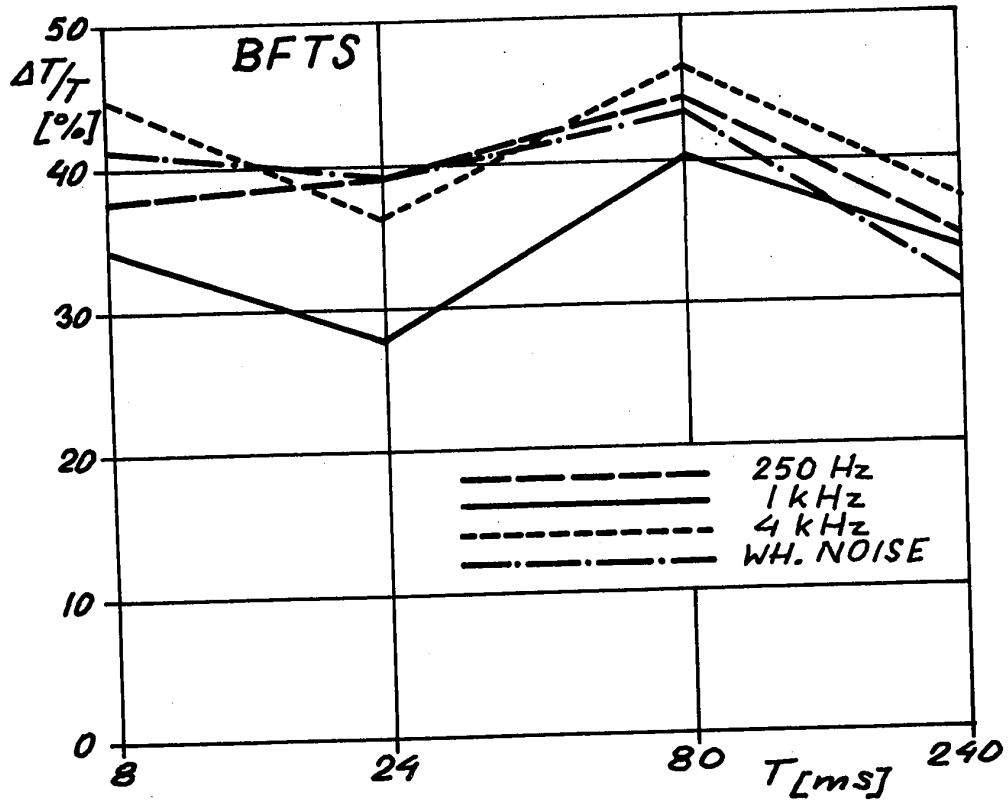
Figure 5.11 Data selected from partial experiment BFTS.
The relative just noticeable increment in
duration $\Delta T/T$ as a function of the duration T
of tone and white-noise stimuli with
rectangular envelope. Carrier signal is the
parameter.

level. Results for tone and white noise stimuli from the BFTS partial experiment are plotted in Figure 5.11. With the exception of one point, the duration discrimination for tone stimuli is better than for white noise at stimulus durations 8 and 24 msec. At 80 msec the average relative ΔT for tone stimuli and for white noise stimuli are equal. For the 240 msec stimuli the duration of white noise stimuli is easier to discriminate than that of tone stimuli of all three frequencies tested. Surprisingly enough, among the tone stimuli the 4 kHz carrier systematically exhibits the poorest duration discriminability, in spite of the fact that the basilar membrane response time decreases with frequency. The best performance is for carrier signal 1 kHz.

Completing this additional data analysis, we can make these inferences regarding the spectral clue: first, it is instrumental in duration discrimination of stimuli shorter than 80 msec and, second, it is most pronounced for carrier frequency 1 kHz.

Experiments, carried out by the Stuttgart group and described by E. Zwicker and R. Feldtkeller (1967), investigated the threshold of detectability of sinusoidal amplitude modulation of pure tones as a function of signal level. For the optimal modulation frequency 4 Hz, a region of maximal sensitivity was found around the same carrier frequency, namely 1 kHz, see Figure 6.6. The frequency location of this sensitivity maximum was stable for all signal levels tested.

As mentioned earlier, the white-noise stimuli are
unique in one respect, that they offer no spectral clue about
the stimulus duration. Their short-term spectrum covers
uniformly all the audio frequency range and is independent
of the stimulus duration. In the range of the shortest
durations white-noise stimuli can be discriminated only on
the basis of the energy integrating mechanism. In all de-
tection tasks, of which detection of the energy increment
resulting from lengthening the stimulus duration is one
case, the index of sensitivity d' of the detector is defined
as the distance Δm between the mean values of the two dis-
tributions, in the observation space, on which the decision
is based, divided by the standard deviation of the detector
noise. In the case of negligible external noise only the
internal neural noise with standard deviation σ_n is in-
cluded in the detector noise and the index of detectability
can be expressed as

$$d' = \frac{\Delta m}{\sigma_n}$$

In the case where the signal itself is submerged in the noise
background, this external noise with standard deviation σ_e
is added to the internal noise of the detector. This results
in decline of the detector sensitivity, reflected in lower
value of detectability

$$d' = \frac{\Delta m}{(\sigma_e^2 + \sigma_n^2)^{1/2}}$$

Two sets of data are available for comparison: results

of the experiment of S.R. Rochester (1970), described earlier, and our data for white noise carrier stimuli with rectangular envelope curve. Both data can be compared in Figure 5.10. As different psychophysical methods were used to obtain these results, we can compare only the trends of the curves. But we can assume that to each of these curves a certain constant value of d' can be ascribed. At the stimulus duration 100 msec the discriminability is probably based to a high degree on the clock and counter mechanism, as the energy detection can play only a secondary role at this signal duration. Let us take the discriminability at this signal duration as a reference point. From this point, in the direction of shorter stimulus durations, the duration increment ΔT decreases much more steeply for our stimuli, with approximately 70 dB signal to noise ratio, than does the increment for stimuli with only 3 dB signal to noise ratio, as used in experiments by S.R. Rochester. This indicates that the shorter the stimulus, the more important is the role played by the detection of the energy increment in duration discrimination.

If the duration discrimination of short stimuli were based on energy detection, essentially the same relative jnd whould be obtained for duration discrimination as was measured for intensity discrimination. The following data reported on intensity discrimination are relevant.

E. Zwicker (1965c) assumed that the simultaneously masked signal is detected when the increment in the integral

of the intensity due to the maskee exceeds 25 per cent of the intensity integral within the same critical band produced by the masker signal alone. The integration interval is in this case 200 msec. But experiments carried out by R.A. Campbell and E.Z. Lasky (1967), D.M. Green and S.T. Sewell (1962) indicate that intensity discrimination is markedly poorer with brief clicks or tone or noise bursts than with longer stimuli of the same type. Several investigators measured the just detectable stimulus intensity increment ΔI over the reference stimulus intensity I for short stimuli and expressed the differential sensitivity as a Weber ratio $\Delta I/I$.

Simultaneous masking between tone pulses of identical frequency and duration was studied by R.A. Campbell and E.Z. Lasky (1967). From the data reported we can make inferences about the intensity discrimination of tone stimuli of 20 msec duration and frequency 1kHz in the sound pressure level range 10 to 90 dB. The two-alternative forced-choice (2AFC) psychophysical method was used with either 1330 or 580 msec interstimuli intervals. In the 30 to 45 dB stimulus intensity range the $\Delta I/I$ ratio was about 1.5. For higher intensity levels the value of $\Delta I/I$ gradually decreased down to 0.55 at the level of 90 dB. In the range of stimulus intensities used in our experiments, which was 70 dB sensation level, or about 80 dB sound pressure level, the relative jnd in intensity was about 0.60.

D.H. Raab and H.B. Taub (1969) studied intensity

discrimination of clicks over a range of sensation levels
from 10 to 90 dB. The differential thresholds were determined using a 2AFC psychophysical method. The interstimulus
interval was 800 msec. In an experiment in which a situation
without any background noise was investigated, the lowest
value of 0.25 for the relative jnd in intensity $\Delta I/I$ was
obtained for click intensity 80 dB. In the range of sensation levels 70 dB used in our experiments, the authors
arrived at the value of $\Delta I/I$ about 0.70. These brief clicks
were most poorly differentiated at click intensity 30 to
40 dB where the required intensity increment exceeded the
intensity of the reference click, up to the value $\Delta I/I = 1.20$.

In a power discrimination experiment of his series of
power and phase discrimination experiments D.A. Ronken (1970)
used as a reference stimulus a transient signal consisting
of two 250 μsec rectangular pulses of the same amplitudes.
Spacing between these two pulses was variable between 1 and
10 msec. The amplitude of the similar test pulse was variable. The 2AFC method was used to find the necessary
amplitude difference between the standard and test pulses
for which the observers correctly identified 75 per cent of
stimulus pairs as either "identical" or "different". The
interstimulus interval was 500 msec. The stimuli were
presented at sensation level about 60 dB. The obtained
mean jnd in amplitude of about 2 to 3 dB required for 75 per
cent correct discrimination corresponds to a value of the
relative increment in intensity $\Delta I/I$ of 0.60 to 1.00.

If we interpret the duration discrimination at short stimulus durations as an energy-detection process, we should expect to obtain a relative differential threshold in duration of the same order as the relative differential threshold in intensity reported in the just mentioned three papers. In our experiments we have measured the mean relative threshold at durations 8 msec to be about 40 per cent. The lowest value, 31 per cent, was obtained for isosceles triangle stimuli with 1 kHz tone carrier. At stimulus durations 24 msec our relative jnd was about 33 per cent. These values are markedly lower than the corresponding values from 0.60 to 1.00 for $\Delta I/I$ obtained in the three previously mentioned papers on intensity discrimination of brief stimuli. It is obvious that energy increment detection mechanism alone can not account for our results in the short stimulus duration range. It is plausible, as the following discussion of the duration discrimination in the range of the long stimuli indicates, that at these base durations the clock and counter mechanism still contributes to the discrimination. Shorter stimuli than 8 msec have not been used. However, the trend of our results between stimulus durations 24 and 8 msec suggests further increase in the relative jnd in duration for stimuli shorter than 8 msec.

Detection of energy differences and spectral contour differences, relevant for duration discrimination of short stimuli, can be modelled by a model presented in Chapter VII.

5.7. Duration Discrimination of Long Stimuli

In the region where the stimulus duration exceeds the auditory integration constant, the duration discrimination can be explained by a pulse counting mechanism as described by C.D. Creelman (1962). If this assumption were correct, we would expect the best performance in this stimulus duration region to be for stimuli with their onset and offset transients exactly defined events in time. From the three forms of stimuli tested in our experiment, the rectangular envelope signal is characterized by the sharpest beginning and termination transients, with no other transients present between these two events. Thus we would expect the rectangular stimuli to have the shortest difference limen in duration in the stimulus duration range where duration discrimination is based on the clock pulse counting mechanism.

Looking back through our data for the significance of the interaction between the signal envelope factor E and the stimulus duration factor T, we find this interaction to be nonsignificant in the partial experiment EFTS, as Table 5.1 shows. But in the partial experiment EBTS the ET interaction is significant, although only at the 5 per cent level, as can be verified in Table 5.2. Figure 5.7 displays the results of the partial experiment EBTS averaged over the subjects and over all the nonsignificant single factors. In contradistinction to this diagram, in Figure 5.12 are plotted the same data, but averaged across the subjects and across the spectral bandwidth only. In this diagram the envelope
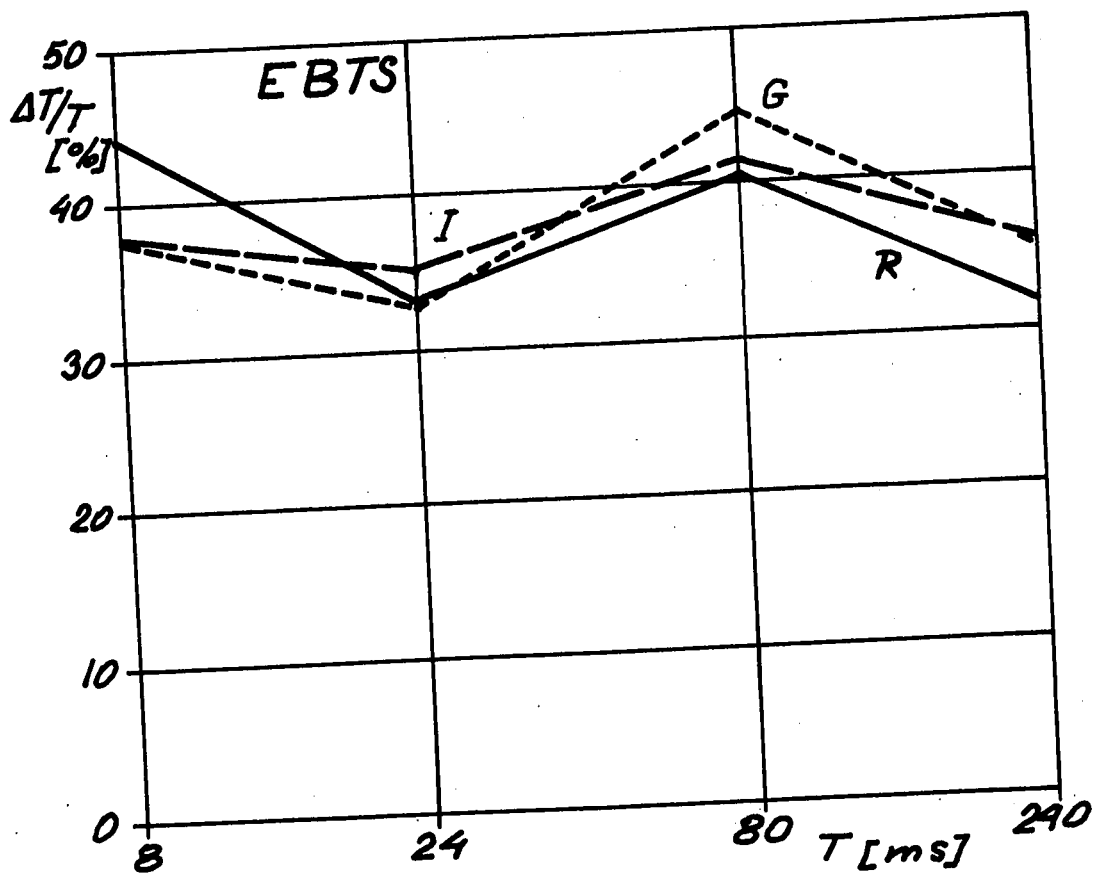
Figure 5.12   Partial experiment EBTS. The same data as in
              Figure 5.7 but not averaged across the envelope
              factor E. The relative just noticeable incre-
              ment in duration ΔT/T as a function of the
              duration T of tone and noise stimuli with
              carrier central frequency 1 kHz. Stimulus
              envelope is the parameter: R - rectangular,
              I - isosceles triangle, G - Gaussian.

curve of the stimuli is the parameter. The data arranged
in this way support our presumption: for the shortest stimuli
tested, the discrimination of the rectangular stimuli is the
worst from all the three envelopes tested. But, beginning
with the stimulus duration of 24 msec, the longer the stimulus
duration the better the discrimination of rectangular pulses,
curve R, as compared with the discrimination of triangular
and Gaussian stimuli, represented by curve I and G, respect-
ively.

C.D. Creelman, discussing his model, stated that the
constantly running "internal clock" will not account for his
data. No reasons were offered. Next we present a modifica-
tion of his model, which will take into account also the
decreased detectability of duration increments at short base
durations. The block diagram of our model is presented in
Figure 5.13. Let us suppose that the mechanism for duration
discrimination depends on a clock pulse source CPS. This
pulse source is formed by a single clock pulse generator.
The clock pulses are issued continuously at time intervals
t_c. This clock interval is, however, random and nonstation-
ary. Its mean value $\bar{t}_c$ over a time period of an order of
a testing session may drift due to such factors as organismic
variables, stimulus parameters, and general level of stimula-
tion. But we can assume that these external and internal
factors remain unchanged within a testing session. In that
case the mean value of the clock interval can be regarded as
time invariant, having a Gaussian distribution with mean
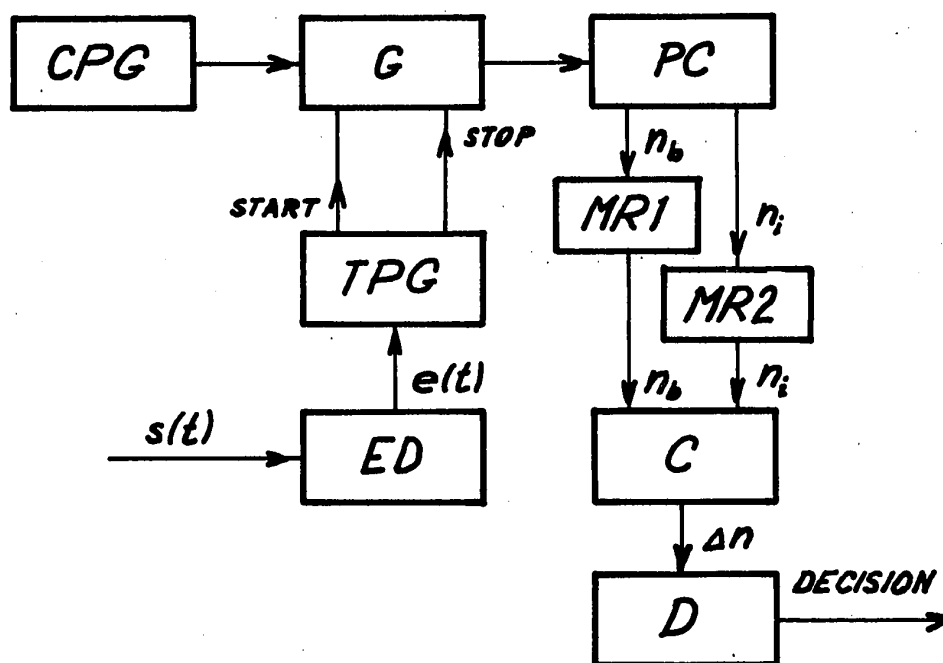
Figure 5.13   Block diagram of the clock and counter model

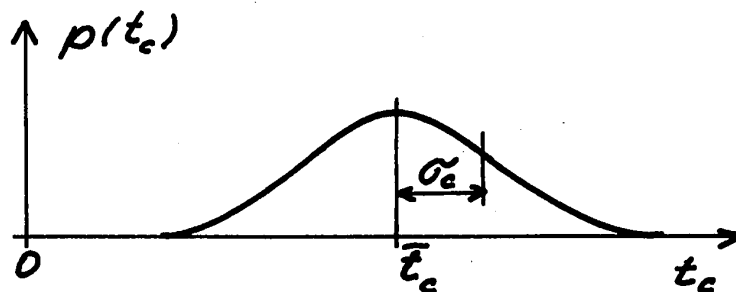for duration discrimination of long stimuli.

Figure 5.14   Probability density distribution $p(t_c)$

of the clock interval t_c.

value $\bar{t}_c$ and standard deviation σ_c, as illustrated in Figure 5.14.

This is the basic difference between C.D. Creelman's and our model. In his model the firing rate of the clock pulses was assumed to be governed by the Poisson probability density distribution

$$p_p(x) = \frac{\lambda^x e^{-\lambda}}{x!} .$$

Only one parameter figures in this formula. Both the mean value and the variance in a probability density distribution of this type have the same value λ and consequently can not be selected independently. It is quite possible that in C.D. Creelman's model the short mean interval of the clock pulse source, 0.10 to 0.37 msec, was only a result of matching the variance of the model clock pulse source to the actual variance in the neural clock pulse generator. And, due to the said property of the assumed Poisson probability density distribution, the clock interval worked out to be too short to agree with the measured drop in observers' performance at short stimulus durations. On the other hand, in the Gaussian probability density distribution

$$p_G(x) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2} .$$

there appear two independent parameters: mean value μ and standard deviation σ. Thus, if we regard the clock time interval to have the Gaussian probability distribution, it is possible to adjust the variance of the clock generator according to the observers' performance at sufficiently long

stimulus durations. At the same time the mean value of the clock time interval can be chosen to model the duration discrimination at short stimulus durations.

The pulses generated by the clock generator pass through the gate G, which is controlled by pulses derived from the incoming stimulus s(t). In the envelope detector ED the envelope e(t) is obtained from the stimulus s(t). In the TPG block triggering pulses are generated to control the gate. A start pulse, generated at the onset of signal e(t), opens the gate and a stop pulse, issued at the termination of the envelolpe e(t), closes it. Thus, at the output of the gate G appear only the clock pulses generated by CPG during the duration of the stimulus. These pulses are counted in a pulse counter PC. The final count at the termination of the first stimulus is stored in the memory register MR1. Let us suppose that this count n_b corresponds to the stimulus of a base duration T. Due to the random character of the clock intervals, this pulse count can be described only by a probability density distribution function. If we assume this distribution to be Gaussian, then its mean value will be

$$\bar{n}_b = \frac{T}{\bar{t}_c}$$

and its standard deviation

$$\sigma_b = \bar{n}_b^{1/2} \frac{\sigma_c}{\bar{t}_c} = \frac{T^{1/2}}{\bar{t}_c^{3/2}} \sigma_c \quad .$$

Similarly, the count n_i of the counter at the termination

of the second stimulus of the incremented duration $T + \Delta T$ is stored in another register, MR2. The pulse count n_i will be distributed with mean value

$$\bar{n}_i = \frac{T + \Delta T}{\bar{t}_c}$$

and standard deviation

$$\sigma_i = \frac{(T + \Delta T)^{1/2}}{\bar{t}_c^{3/2}} \sigma_c .$$

Probability distributions of n_b and n_i are illustrated in Figure 5.15.

After completion of both counts, n_b and n_i are compared in the comparator C. In this unit the difference between the counts is calculated,

$$\Delta n = n_i - n_b .$$

The value of Δn is put into the decision block D, to serve as a basis for the decision process. Subject's decision and his response, either "identical", or "different", are based on the probability density distribution of Δn, which is, again, Gaussian. Assuming that the values of n_i and n_b are uncorrelated, this distribution has a mean value

$$\overline{\Delta n} = \bar{n}_i - \bar{n}_b = \frac{\Delta T}{\bar{t}_c}$$

and a standard deviation

$$\sigma_d = (\sigma_i^2 + \sigma_b^2)^{1/2} = \frac{(2T + \Delta T)^{1/2}}{\bar{t}_c^{3/2}} \sigma_c$$

as illustrated in Figure 5.16.
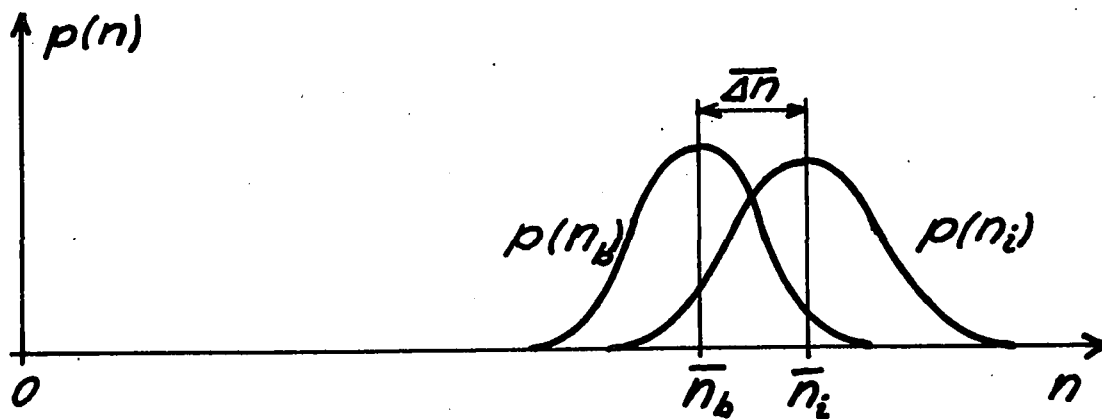
Each observer establishes his own criterion N. For

Figure 5.15   Probability distributions $p(n_b)$ and $p(n_i)$ of
the pulse counts n_b and n_i, corresponding to
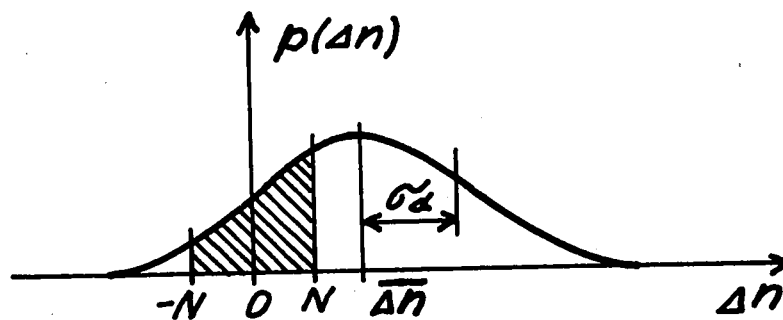the stimulus durations T and ΔT, respectively.

Figure 5.16   Probability distribution $p(\Delta n)$ of the
difference Δn between the pulse counts
n_b and n_i.

$$|\Delta n| < N$$

his response is "stimuli are of equal duration." In the case

$$|\Delta n| \geq N$$

he responds "stimuli differ in duration." The shaded area in Figure 5.16 represents the error probability of responding "equal" to stimuli of different durations.

Finally, the detection index d' takes the form

$$d' = \frac{\overline{\Delta n}}{\sigma_d} = \frac{\bar{t}_c^{1/2}}{\sigma_c} \frac{\Delta T}{(2T + \Delta T)^{1/2}} \quad . \quad (5.7)$$

This detection index depends on the stimulus base T and on the stimulus duration increment ΔT in the same way as in C.D. Creelman's model.

For stimulus durations of the order of the clock interval, this mechanism can not support the duration discrimination as, in this case, the error in duration estimation is of the same order as the duration itself. As the stimulus duration increases, the duration discrimination improves. In C.D. Creelman's model-verifying experiments the significant decrease in subjects' performance in comparison to that predicted by his model appeared for base duration T = 40 msec and increment duration ΔT = 5 msec. This fact determines the duration of the clock interval to be of the order of several milliseconds. Applying our model, described by Equation 5.7, to the data presented in C.D. Creelman's Figure 6, an average value 5.3 msec is obtained for the clock interval $\bar{t}_c$ and 2.6 msec for its standard

deviation σ_C. For stimuli shorter than the mean clock interval the clock and counter mechanism is unable to perform reliably enough. As a consequence, the duration discrimination in that stimulus duration range is based mostly on spectral differences and on differences in stimulus energy.

Chapter VI

ENVELOPE DISCRIMINATION EXPERIMENTS

In this series of experiments, again, the discriminability of two acoustic pulses was investigated. The pulses to be discriminated were of the same effective duration, envelope peak amplitude, and carrier signal, but of different envelope curves. Thus the influence of the differences in duration, loudness, energy, and spectral composition were minimized leaving the different stimulus envelopes as the only clue for discrimination. The effective stimulus duration T was defined by Equation 5.1. The shortest effective duration, sufficient for the pulses to be perceived as different, was sought, as a measure of the envelope discriminability. This limit will be further referred to as the critical duration T_L. The stimuli with different envelopes are indistinguishable if their duration is shorter than this critical duration.

6.1 Variable Factors

Similar stimulus and individual factors as in the duration discrimination experiment, save the stimulus duration factor, were taken into consideration. The following levels have been chosen for the three stimulus factors:

As the envelope factor E selected pairs of five types of envelope curve were used. In addition to the shapes employed in the previous experiment, namely rectangular,

isosceles triangle, and Gaussian (Figures 5.1, 5.2, and 5.3), two new envelopes were used: divergent triangle $e_D(t)$ and convergent triangle $e_C(t)$. These two functions were chosen to allow us to investigate the relevance of the time sequence of such stimulus events as a jump and ramp of the envelope curve. Note that the two last mentioned envelopes are mutually reversed in time and consequently their energy spectral densities $E_D(\omega)$ and $E_C(\omega)$ are identical:

$$e_D(t) = \begin{cases} \dfrac{t}{2T} & \text{for } 0 < t < 2T \\ 0 & \text{elsewhere} \end{cases}$$

$$e_C(t) = \begin{cases} 1 - \dfrac{t}{2T} & \text{for } 0 < t < 2T \\ 0 & \text{elsewhere} \end{cases}$$

$$E_D(\omega) = E_C(\omega) = (T^2 + \frac{1}{\omega^2})\, Sa^2(\omega T) + \frac{1}{\omega^2}[\cos^2 \omega T - 2Sa(\omega T)\cos \omega T] \ .$$

These functions are plotted in Figure 6.1. Also in this case the frequency scale of the spectral function is normalized in terms of the effective stimulus duration T, given by Equation 5.1, and $Sa(x) = (\sin x)/x$. For the actual shape of the envelopes refer to Figure 4.3. Not all possible pair combinations of these five envelopes were tested, as is described later in the text.

The carrier signal bandwidth factor B, as well as the carrier central frequency factor F, were examined at the same levels as in the duration discrimination experiment. The individual factor S was also investigated to the same extent, as six subjects took part in the testing.
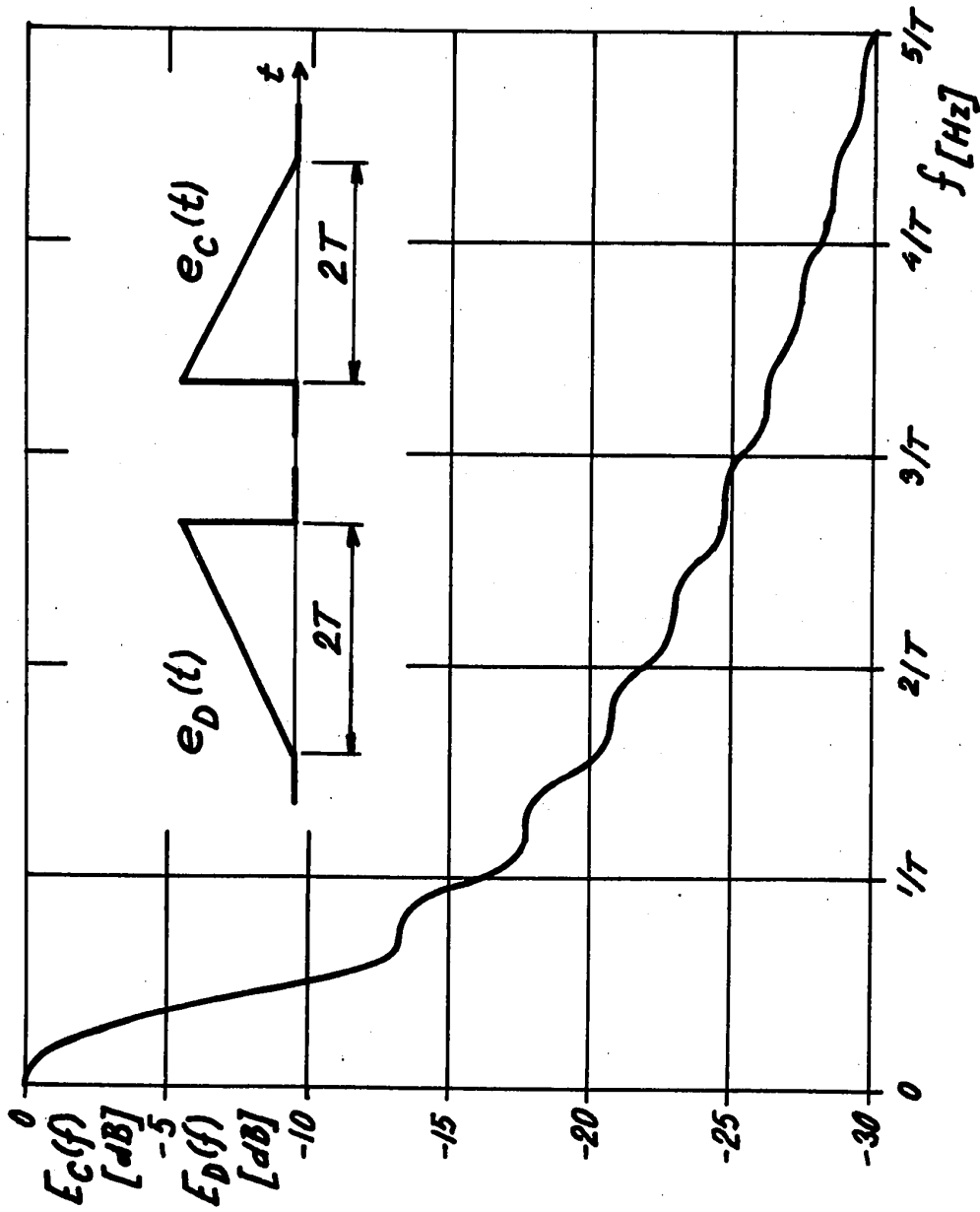
Figure 6.1   The convergent triangle $e_C(t)$ and divergent triangle $e_D(t)$ envelopes and their energy spectral density $E_C(f) = E_D(f)$.

6.2 Results

Three series of measurements were carried out. In the first one the influence of the carrier signal frequency and of the stimulus envelope on the critical duration was determined. The second series was essentially repetition of the first one, but with the number of levels increased for carrier frequency factor F and decreased for envelope factor E and subjective factor S. In the third series, apart from the stimulus envelope, also the influence of the spectral bandwidth of the carrier signal on the envelope discrimination was studied.

The data were processed using the analysis of variance method described in Section 5.2.

6.2.a Envelope Discrimination for Tone Carriers, Experiment A

In the first series, which can be called the EFS series, in accordance with the variable factors, six different pairs of five types of envelopes were originally employed as six levels of the factor E: rectangular with Gaussian (RG), and with isosceles triangle (RI), Gaussian envelope with isosceles triangle (GI), with convergent triangle (GC), and with divergent triangle (GD), and, finally, convergent with divergent triangle (CD). But the initial testing proved subjects' inability to discriminate between pulses of Gaussian and isosceles triangle envelopes even at stimulus effective duration 15 msec. So this envelope combination was abandoned, and consequently the number of levels of factor E

was reduced to five. In this series only pure tones were used as stimulus carrier signals, hence the carrier bandwidth factor B was constant. As an addition, for the sake of comparison, the case of white noise carrier signal was also tested, but these data were not included in the analysis of variance for this series. Six observers were tested. The results of analysis of variance applied to these data are shown in Table 6.1. This analysis revealed as significant at the one per cent level the envelope curves combination factor E and the tone frequency factor F, both as single factors, as well as their joint influence. Thus our model for this series of the envelope discrimination experiment has the form of

$$x_{ikn} = m_{EFS} \, f_k \, v_i \, vf_{ik} \, e_{EFS} \quad .$$

The order of the factor contribution coefficients from left to right reflects the degree of significance of corresponding factors. It is interesting to note that no significant differences between subjects were observed.

The data collected in this series, averaged over the group of observers as the only single nonsignificant factor, are plotted in Figure 6.2.

6.2.b Envelope Discrimination for Tone Carriers, Experiment B

In the second series the envelope discrimination was investigated for half octave steps of the carrier frequency, in the range from 250 Hz to 4 kHz, in order to detect any

E F S	Source	d.f.	MS	F-Ratio
	S	5	0.012	3.176*
	E	4	0.203	54.334**
	F	2	0.434	71.026**
	EF	8	0.164	53.516**
	SE	20	0.00374	
	SF	10	0.00611	
	SEF	40	0.00306	**p<0.01
	S pooled	70	0.00369	*p<0.05

Table 6.1 Analysis of variance for the experiment EFS-A.

E B S	Source	d.f.	MS	F-Ratio
	S	5	0.00264	1.760
	E	1	0.346	112.133**
	B	4	0.106	63.000**
	EB	4	0.0596	63.351**
	SE	5	0.00309	
	SB	20	0.00168	
	SEB	20	0.00094	**p<0.01
	S pooled	45	0.00150	*p<0.05

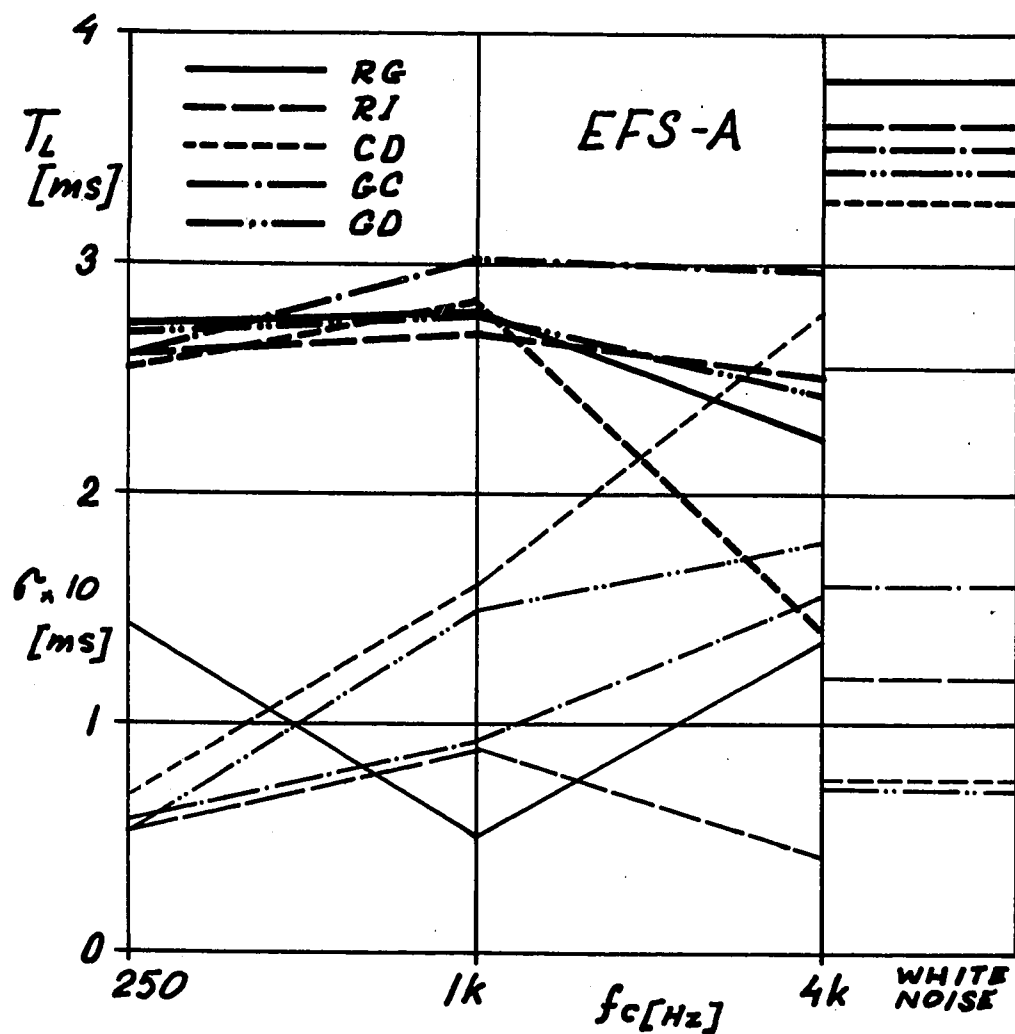Table 6.2 Analysis of variance for the experiment EBS.

Figure 6.2 Experiment EFS-A. The critical duration T_L and the standard deviation σ for tone and white-noise stimuli as a function of carrier frequency f_c. Envelope combination is the parameter.

possible irregularities of the discriminability versus
carrier frequency relationship. We were mainly interested
in confirming of the poorer ability of hearing to detect
envelope variations at the carrier frequencies around 2.8 kHz,
as observed by E.Hojan and A.Rozsypal (1967). Their results,
discussed in paragraph 6.3 below, are given in Figure 6.13.
For this series of measurements two envelope pairs were
selected, namely rectangular with Gaussian envelope (RG),
and convergent with divergent triangle (CD). Two subjects,
Ko and Ch, took part in this series.

As can be seen from the data plotted in Figures 6.3 and
6.4, no irregularity in envelope discrimination around
carrier frequency 2.8 kHz was revealed by these measurements.
In general, increasing the frequency of the carrier tone
above 1 kHz, the critical duration gradually decreases. Both
subjects perform better in the case of the envelope pair CD
rather than pair RG.

6.2.c Envelope Discrimination for Noise Carriers

In the third or EBS series, which investigated the
influence of the carrier signal bandwidth factor B on the
critical duration T_L, the central frequency of the carrier
signal, i.e., factor F, was kept constant at 1 kHz. The
envelope curve factor E was studied at two levels: rectan-
gular with Gaussian envelope (RG), and convergent with
divergent triangle (CD). The following five levels of the
carrier bandwidth factor B were investigated: pure tone,
narrow-band noise, third-octave noise, one-octave noise,

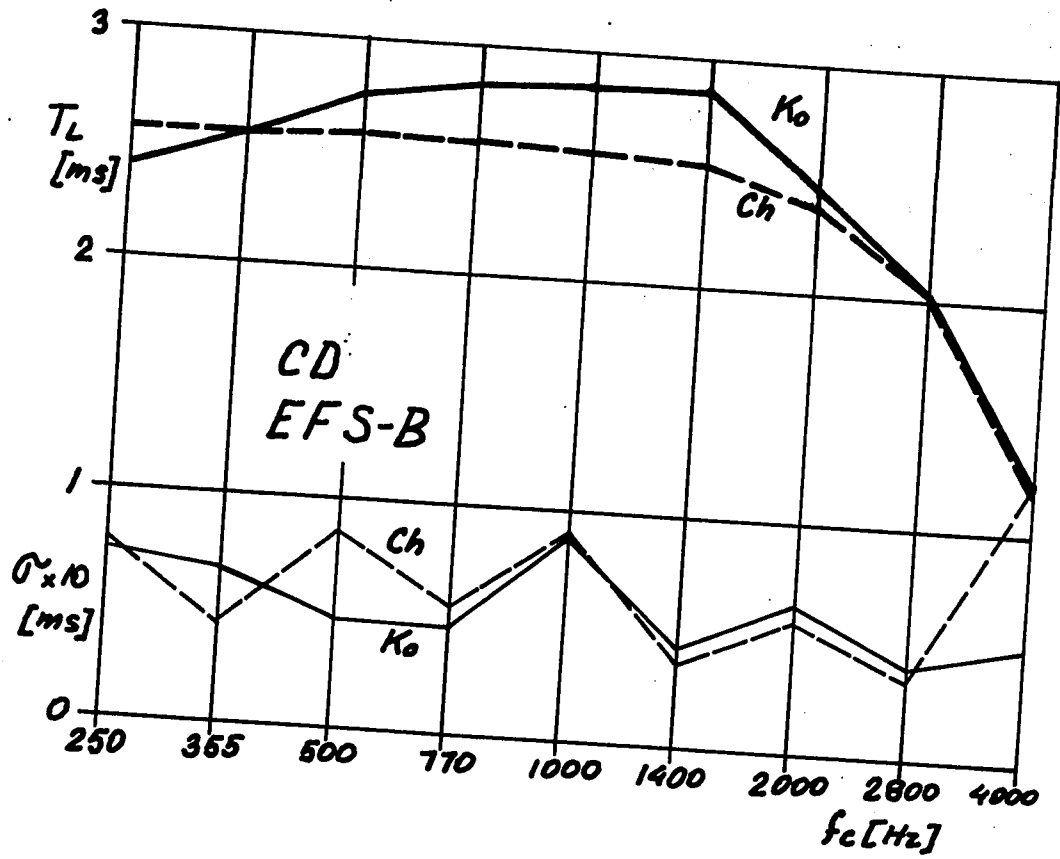Figure 6.3 Experiment EFS-B, convergent and divergent triangle envelopes. The critical duration T_L and the standard deviation σ for tone stimuli as a function of carrier frequency f_c. Subjects Ko and Ch.
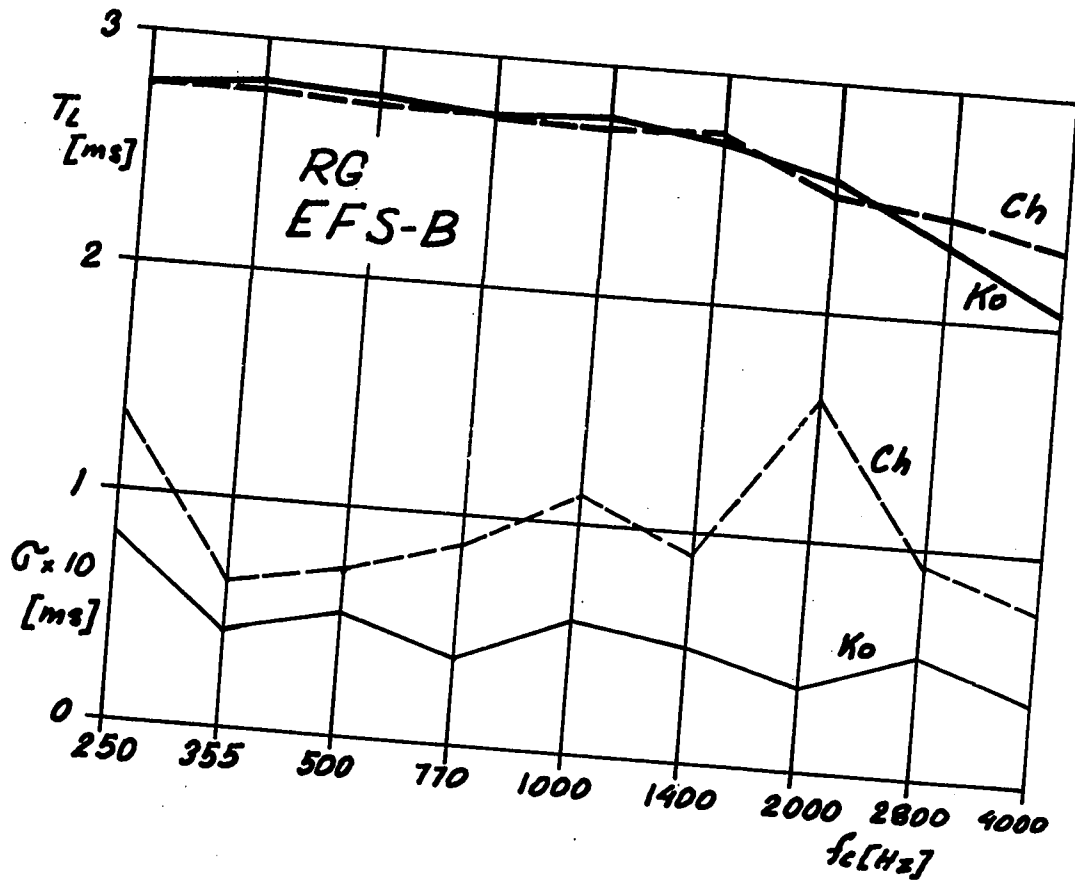
Figure 6.4   Experiment EFS-B, rectangular and Gaussian
envelopes. The critical duration T_L and the
standard deviation σ for tone stimuli as a
function of carrier frequency f_c.
Subjects Ko and Ch.

and white noise. Six observers were tested in this series. The analysis of data revealed, according to Table 6.2, that each trial in this series can be modeled by an equation

$$x_{ijn} = m_{EBS} \; v_i \; b_j \; vb_{ij} \; e_{EBS} \; .$$

Here again, as in the first series of the envelope discrimination experiment, the subjective factor S remained insignificant. Both signal factors, namely the carrier spectral bandwidth B and the combination of envelopes E, as well as their interaction, turned out to be significant.

The results, averaged over the group of observers, are shown in Figure 6.5.

6.3 Experiments of Other Investigators Related to Envelope Discrimination

Much is revealed about the response of hearing to transient stimuli by the experiments carried out by the group of E.Zwicker and R.Feldtkeller (1967). Specifically related to our experiments are their experiments investigating how the detectability of amplitude modulated tone and noise carriers depends on the frequency of the carrier signal and on the frequency and shape of the modulation function.

The dependence of the just noticeable degree of sinusoidal amplitude modulation m on the modulation frequency f_m is indicated by heavy lines in Figure 6.6. In this diagram different carriers are compared: pure tones of the frequency 250, 1000, and 4000 Hz and white noise. The sound pressure level is in all cases 40 dB. The course of all three curves for the tone carriers is similar, especially in the region
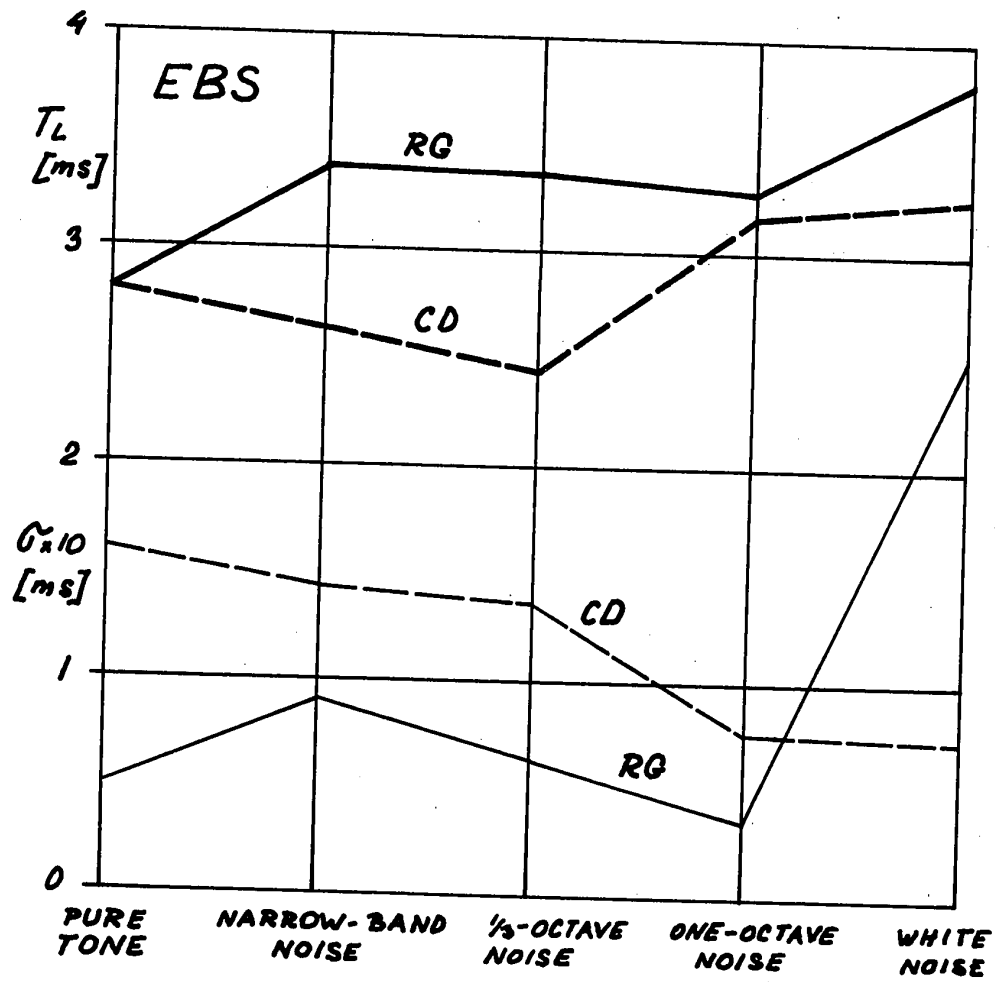
Figure 6.5   Experiment EBS. The critical duration T_L and the standard deviation σ as a function of the spectral banwidth of noise bursts with central frequency 1 kHz.   Combination of envelopes is the parameter: RG - rectangular and Gaussian, CD - convergent and divergent triangle.
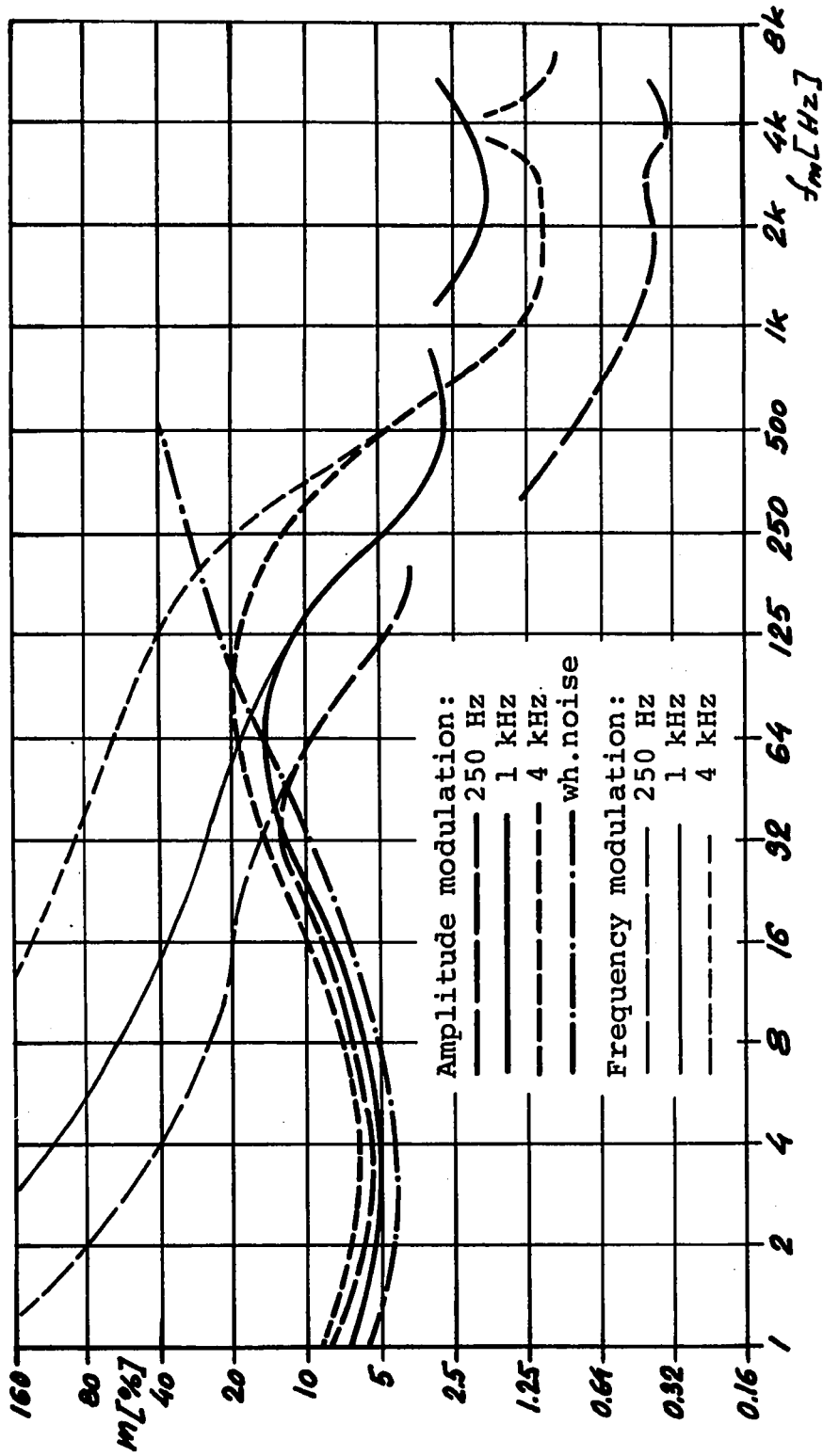
Figure 6.6 The just detectable degree m of amplitude and frequency modulation

of tone and white-noise carriers as a function of the modulating

frequency f_m. The sound pressure level is 40 dB. Carrier signal is

the parameter.   E.Zwicker and R.Feldtkeller (1967,p.172-173,98).

of the low modulation frequencies. The region of optimal sensitivity of hearing to sinusoidal amplitude modulation can be found around the modulation frequency 4 Hz, at which the hearing is able to follow fully the envelope curve variations. For the lower modulation frequencies the hearing lacks the ability of direct comparison of the two extreme signal levels due to imperfect short-term memory. So the degree of amplitude modulation has to be increased with decreasing modulation frequency in order for the modulation to be detected. Tracing the course of these curves in the direction of increasing modulation frequency above its optimal value, the amplitude modulation detectability declines, but, for tone carriers, only up to a certain point. The modulation frequency at which the improvement trend sets in depends on the carrier frequency. The higher the carrier frequency, the later the turnover point takes place: at the modulation frequencies of about 40, 64, and 100 Hz for the carrier frequencies 250, 1000, and 4000 Hz, respectively. In these peak points the highest depth of modulation is required for its detection. In this modulation frequency region the amplitude variations are no more discerned and the amplitude modulation is detected only as a roughness of the tone. By further increasing the modulation frequency, sensitivity to amplitude modulation improves again. Namely, the carrier frequency f_c and the two modulation products, side frequencies $f_c - f_m$ and $f_c + f_m$, become gradually so far apart, that they are perceived as three separate components of a complex

sound. The amplitude variations are not perceived at all, especially in the modulation frequency region in which the ear is unable to distinguish between amplitude modulation and frequency modulation. This region spreads where the threshold curves for amplitude modulation detection (heavy lines) and for frequency modulation detection (thin lines) merge. These merging frequencies are closely related to the widths of the critical bands of hearing for the particular carrier frequency. It appears that the width of the merging frequency. So for a carrier modulated by its merging frequency, no matter whether it is amplitude or frequency modulation, the frequency difference between the upper and lower modulation side spectral line always equals the value of the critical band. The qualitative change of perception takes place when the modulation products fall either in the same or in a different critical band as the carrier signal. As long as the three spectral components representing the modulated signal fall into one critical band, the phase relationships between these components are preserved in the process of perception. The ear is able to discriminate between the amplitude and frequency modulation as the modulated signal is perceived as one either amplitude-varying or frequency-varying tone. In the opposite case, when the three components fall into different critical bands, the phase information is lost and the components are perceived as three isolated tones.

In the same figure is also plotted the just notice-
able degree of amplitude modulation of white noise. The re-
gion of optimal sensitivity is identical as for the case of
pure tone carriers. Near the modulation frequency of about
10 Hz the amplitude modulated noise appears as having a
rough and seething characteristic. The marked distinction
between the modulation threshold curves of noise carriers
against tone carriers consists in the absence of the rever-
sal of the curve for the high modulation frequencies. This
can be explained by the fact that in the case of amplitude
modulated tone the ear is able to resolve the carrier tone
and the two side frequencies. On the other hand, the
spectra of the carrier white noise and of the modulation
sidebands do overlap, as they cover the whole audio fre-
quency range. So the modulation sidebands cannot serve as
a clue for detection of the amplitude modulation by chang-
ing the quality of the perception as is the case of tone
carriers and high modulation frequencies.

The rate of amplitude variations is related to the
absolute spectral bandwidth of the noise signal. The wider
this bandwidth, the more rapid the envelope variations are.
In the case of white noise the envelope variations due to
the spectral bandwidth of the signal are so fast that the
ear perceives it as a homogeneous signal of a constant in-
tensity. When such a signal is amplitude modulated the
ear ascribes all the detected amplitude changes to the am-
plitude modulation. In the case of a narrow band carrier

noise the hearing is unable to discriminate the amplitude changes caused solely by the limited bandwidth of the carrier and the amplitude variations resulting from the process of amplitude modulation. Thus a deeper modulation is required for its detection. Figure 6.7 represents the relationship between the spectral bandwidth of the carrier noise B and the just detectable amplitude modulation depth m for sinusoidal (heavy line) and for rectangular (thin line) amplitude modulation of the frequency 4 Hz.

Another remarkable difference between tone and noise carriers is in the dependence of the just noticeable degree of amplitude modulation on the intensity of the carrier signal. In Figure 6.8 the just detectable degree of sinusoidal amplitude modulation m for different carriers is plotted as a function of the sound pressure level of the carrier signal L. The just noticeable depth of amplitude modulation of white noise carrier signal was measured also using the square wave as a modulation signal. The modulation frequency again is 4 Hz, at which the hearing is most sensitive to amplitude changes. It can be seen that the sensitivity of hearing to amplitude variations consistently increases in the whole intensity range for all sinusoidal carriers. On the other hand, for the white noise, the sensitivity improves only up to the sound pressure level of about 30 dB. Above this signal level the ear maintains constant sensitivity to amplitude modulation of wideband noise carriers.
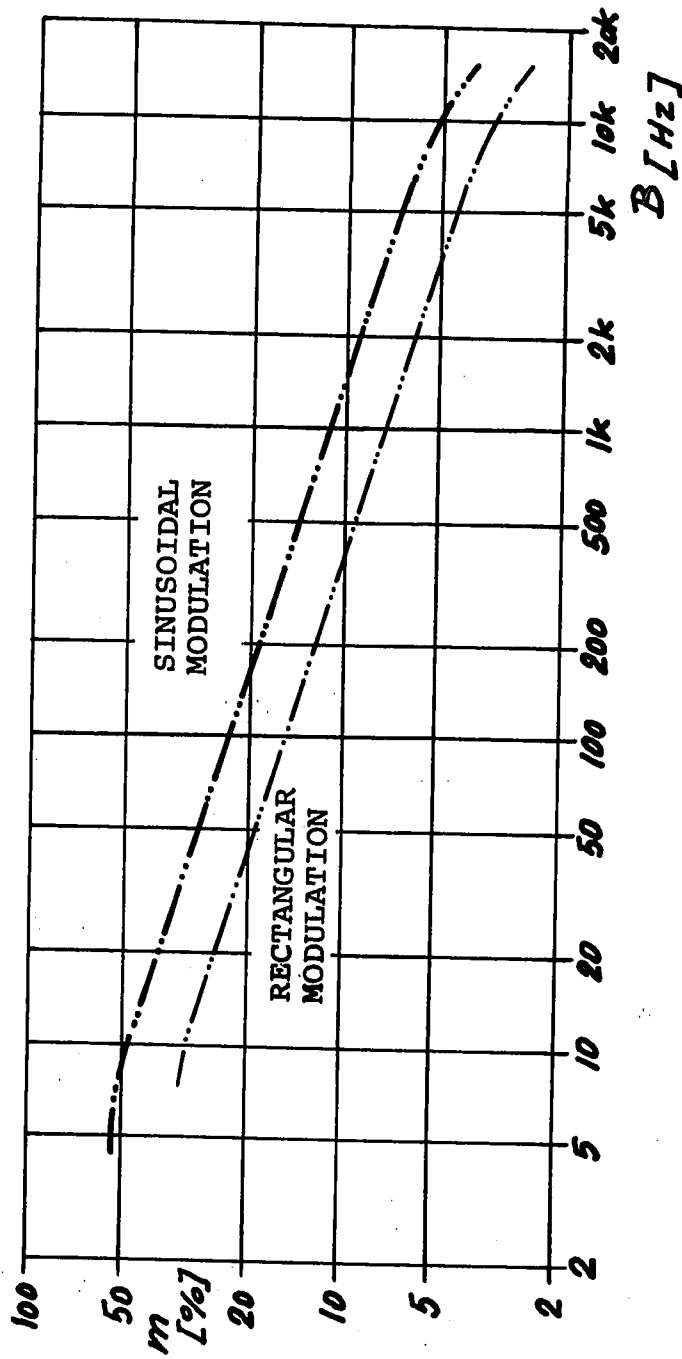
Figure 6.7 The just detectable degree m of sinusoidal and rectangular amplitude modulation as a function of the spectral bandwidth B of the band-limited noise carrier. The modulation frequency is 4 Hz.

E.Zwicker and R.Feldtkeller (1967, p.100).
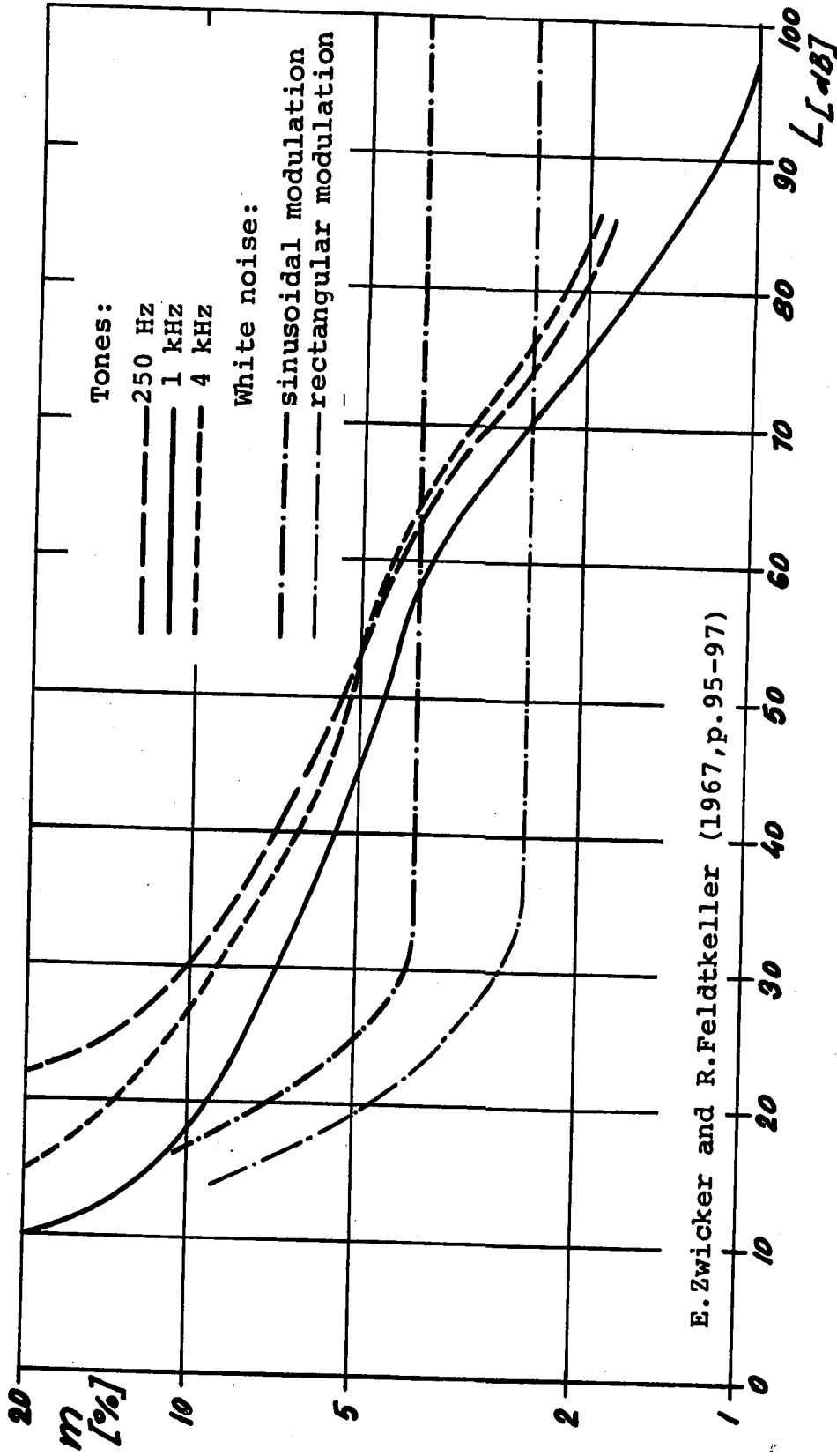
Figure 6.8 The just detectable degree m of amplitude modulation of tone and
white-noise carriers as a function of the sound pressure level
of the carrier signal L. Sinusoidal modulation, for the white-noise
carrier also rectangular modulation. The modulation frequency is 4 Hz.

Also N. A. Dubrovskij and L. N. Tumarkina (1965)
measured the threshold of perception of amplitude modulated
wideband noise and its dependence on the frequency of the
sinusoidal modulation signal over the range 0.5 through
100 Hz. They also considered the influence of subjects'
practice in psychoacoustical experiments on their differen-
tial sensitivity. The first experiment was started as soon
as the subjects gave sufficiently consistent responses.
These results are plotted in Figure 6.9, where curve I re-
presents the averaged just noticeable degree of amplitude
modulation m as a function of the modulation frequency f_m.
Curve II represents the results of the same experiment repeat-
ed with the same group of three observers after a six months'
break. During this period the subjects in question parti-
cipated in experiments testing the detection of amplitude
modulation of pure tones and wideband noise at various loud-
ness levels. The sound pressure level of the noise was 76 dB
in both experiments. According to curve I the maximal sensi-
tivity to sinusoidal amplitude changes lies in the range
between 1 and 2 Hz. The six months practice approximately
doubled the subjects' sensitivity in the optimal sensitivity
range, widened and shifted this optimal region towards high-
er modulation frequencies, from 2 to 5 Hz. The authors
assume that the detection of amplitude modulation is based
primarily on the perception of the periodical variations
in the level of the wideband noise. In the region of optimal
sensitivity on curve II they have obtained nearly the same
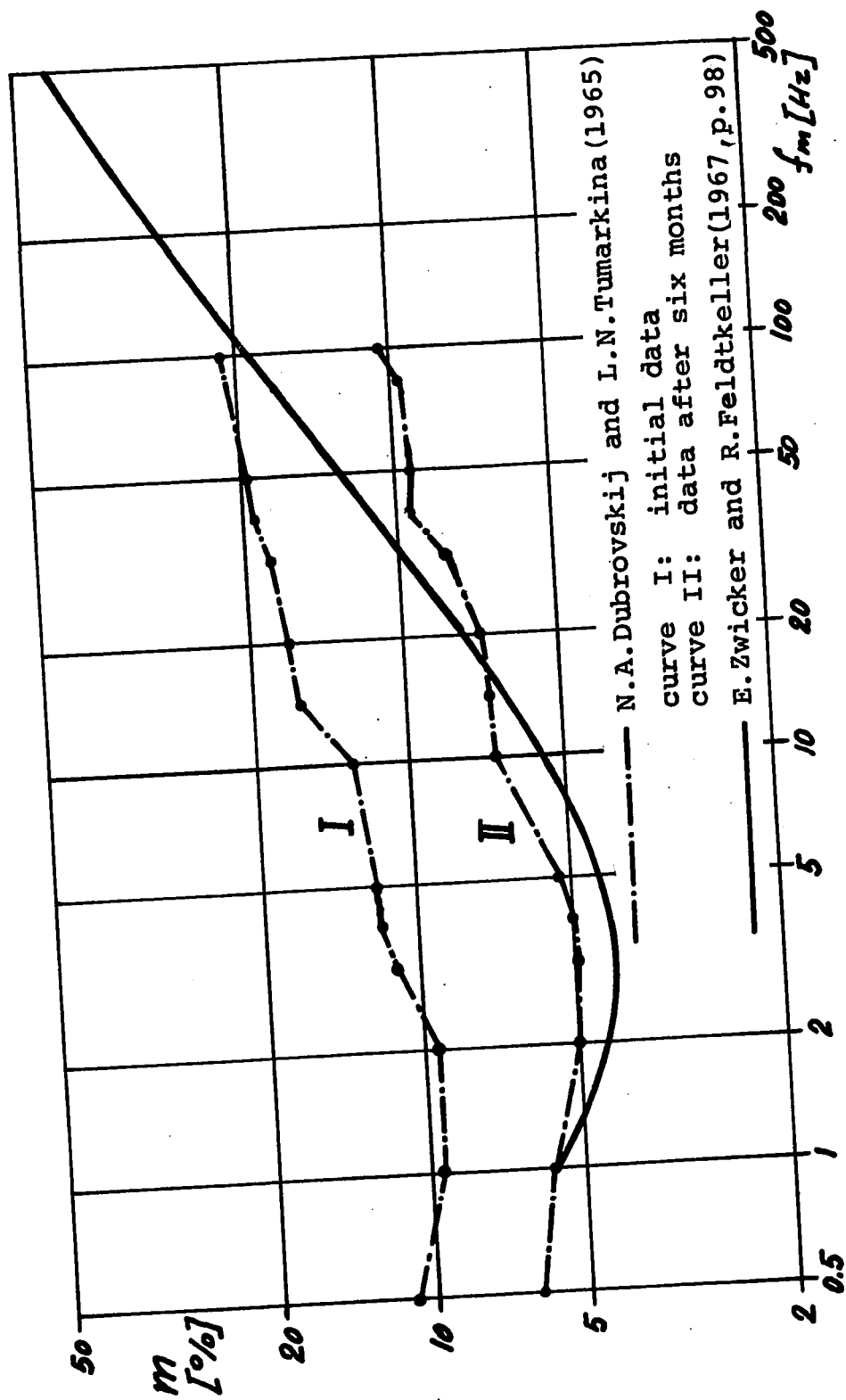
Figure 6.9 The just detectable degree m of sinusoidal amplitude modulation of a white-noise carrier as a function of the modulation frequency f_m.

results as the data for white noise from Figure 6.6, which

are plotted for comparison in the same diagram. On curves

I and II the authors recognize four sections with different

slopes. In the region of the lowest measured modulation

frequencies the rates of sound pressure level changes are

too small for hearing to detect them easily. Due to inade-

quate memory, hearing is obviously unable to retain the ex-

treme absolute levels of pressure for accurate comparison of

events so far apart in time. So the degree of modulation

in order to be detected must be increased with modulation

frequency decreasing below the optimal modulation frequency.

A differentiation property can be thus ascribed to hearing

in the lowest modulation frequency region. The second re-

gion is the flat bottom of both curves, where the differen-

tial threshold is constant. The hearing is able to follow

exactly the changes of the envelope of the stimulus here.

In the third region the hearing still detects the amplitude

modulation on the basis of amplitude changes of the stimulus,

but the limited response-time of the system prevents the

hearing from accurate tracing of the signal envelope. Here

the hearing behaves as an integrating system, as the thresh-

old is proportional to the modulation frequency. In the

fourth region, which begins above 15 Hz on curve I and about

above 40 Hz on curve II the detection criterion is changed,

although the threshold curves have the same positive slope

as in the third region. Here the amplitude modulation is

detected as roughness of perception. On both curves the

transition between the third and fourth section appears as
a short horizontal plateau. The integration time constant
of hearing derived from these experiments by the authors
was initially 10.7 msec and by virtue of the subjects'
practice was reduced to a value of 4.7 msec. The differen-
tiation time constant was decreased only slightly, from the
initial value of 494 msec to a value of 478 msec.

In the last mentioned work the existence of two time
constants, differentiation and integration, has been shown.
Both constants affect envelope curve perception. As white
noise was used as a carrier signal, these experiments did
not reveal whether or not these time constants depend on
the carrier signal frequency.

The dependence of one integration time constant of
hearing on frequency was investigated by H. Scholl (1962c).
In his experiments the masking threshold of a continuous
tone was measured. Periodic rectangular bursts of wideband
noise with a duty cycle equal to one half of the switching
period served as the masker. The masker switching frequen-
cy was varied within the range from 1 to 2000 Hz. The fre-
quency of the maskee tone was 300, 1000, 3000, and 9000 Hz.
The data obtained are plotted in Figure 6.10. All four
curves, with the frequency of the masked tone f_t as a para-
meter, exhibit the same course. Below the critical value
f_{mc} of the masker noise switching frequency f_m the masking
threshold L_m increases about 6 dB per octave of f_m. Above
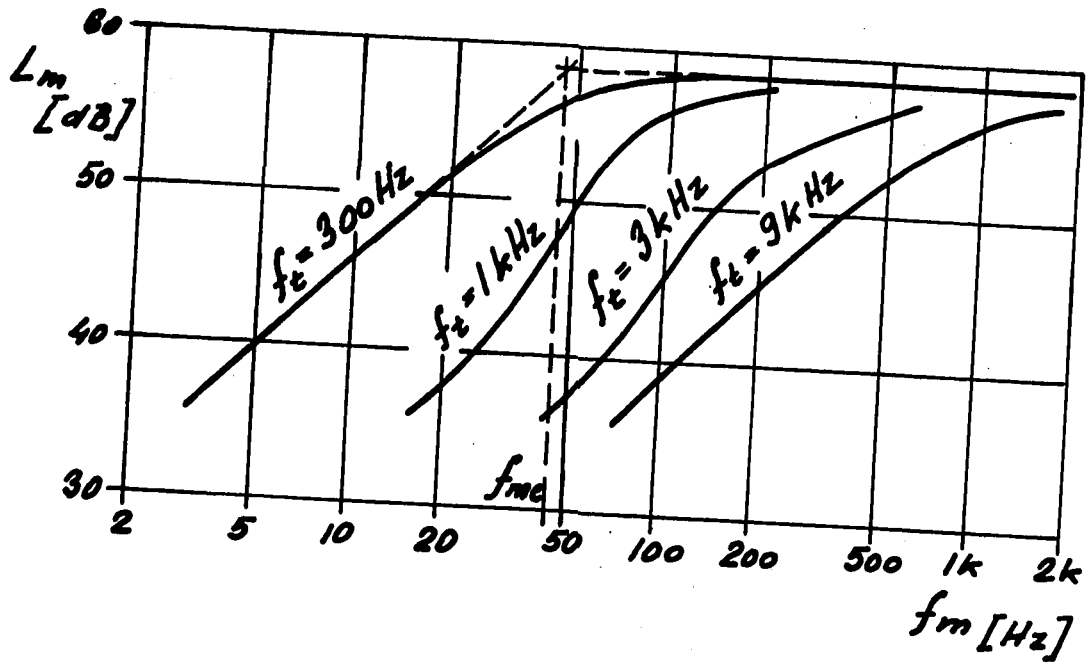f_{mc} the threshold is constant as the masker in this region

Figure 6.10 The masking threshold L_m of continuous tones
of frequency f_t as a function of the pulse
frequency f_m of the rectangular bursts of
white-noise masker. The integration constant
of hearing τ_i is derived from the critical
pulse frequency f_{mc} of the masker.
H.Scholl (1962c).

is fully integrated so that the duration of the pause between the masking noise bursts does not affect the masking threshold. An integration constant τ_i of hearing is derived from these data as the duration of the pause between the masking noise bursts, $\tau_i = 1/(2f_{mc})$. The reciprocal value of the constant τ_i is indicated by crosses in Figure 6.11 as a function of the tone frequency f_t. In the same figure also the width of the critical bands of hearing is plotted by a solid line. The similar dependence on frequency f_t of these two parameters of hearing analyzer, of which the integration time constant can be said to specify its power to resolve signal events in time and the width of the critical band to govern its frequency selectivity, is apparent.

In one part of his extensive publication W. Türk (1940) tried to determine what he calls the physiological build-up time of the hearing analyzer by investigating to what degree hearing discriminates onset rates of tonal pulses. Two tone pulses of the same carrier frequency with stationary portion of 35 msec duration were compared. The amplitude of the stationary portion of both pulses was identical. Both pulses were terminated by a linear tone decay of 20 msec duration, which is slow enough not to evoke a perceptible click. The steepness of the pulse onset was the only clue for discrimination. The first pulse was switched on with the shortest rise time allowed by the apparatus, which was about 0.05 msec. The second pulse was of the exponential onset. In the first test this onset was
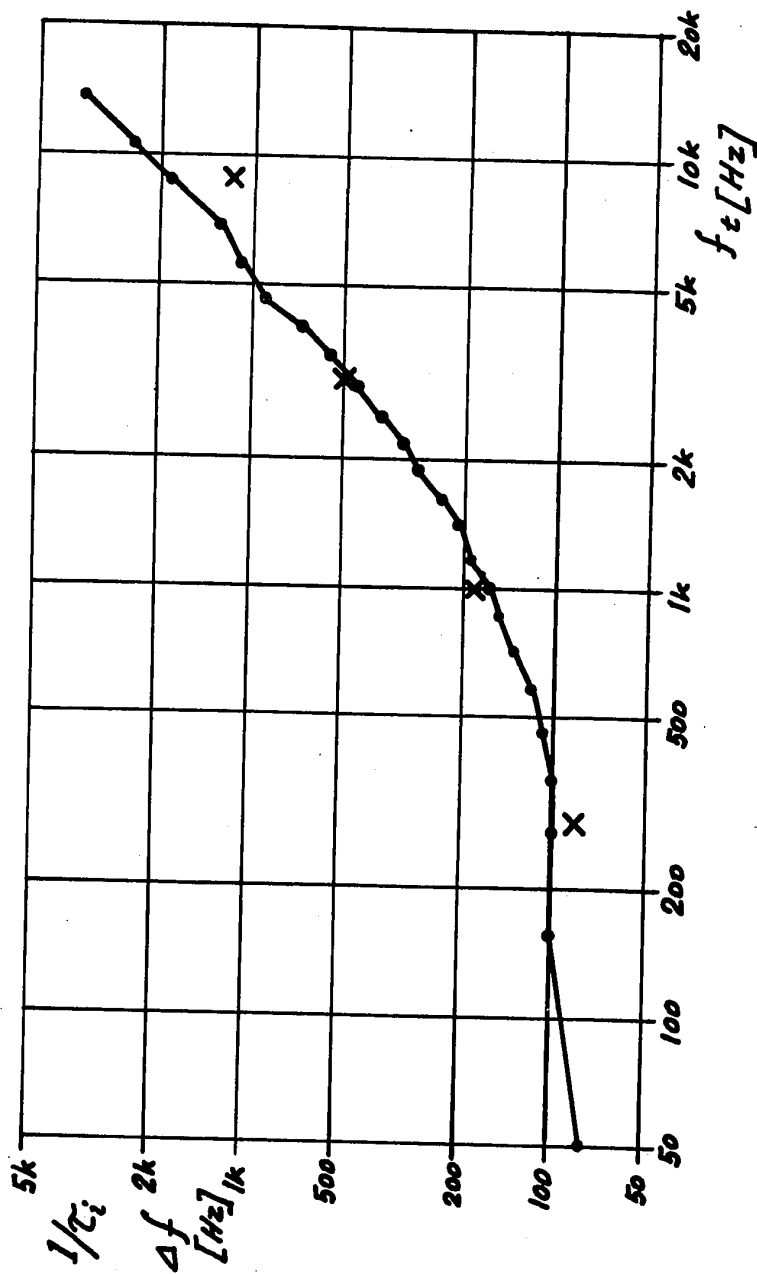
Figure 6.11 The reciprocal value of the integration time constant τ_i (crosses) from Figure 6.10 as a function of the tone frequency f_t, compared with the critical band-widths Δf of hearing (solid line). H.Scholl (1962c).

of the duration of 0.5 msec. All subjects were able to discriminate both pulses mainly by the quality of the onset transient. The most distinctive differences were observed in the range of carrier tone frequencies 50 to 800 Hz and above 6 kHz. In the frequency range 1200 to 5000 Hz the differences were less perceptible. In the second test the rise time of the second pulse was shortened to 0.25 msec. As a result the subjects ceased to discriminate pulses of the carrier frequency around 3 kHz, while the discriminability for the rest of the carrier frequencies remained preserved. To further increase the discriminability of the onset transients, the stationary part of the pulses, which obviously backward masked the onset transient, was left out, so that the decay immediately followed the pulse onset. This step made the discrimination possible in the whole audio range of carrier frequencies, including 3 kHz. As the apparatus used at that time did not allow the exponential rise time to be decreased reliably below 0.25 msec, the author was compelled to conclude that the physiological build-up time of hearing for all the frequencies in the region 50 Hz to 10 kHz is certain to be below 0.25 msec.

In the same paper the experiments investigating the discriminability between different envelope curves of short tone pulses are also described. The subjects were alternatively exposed to two pulses, which featured either different envelope curves, or the same envelopes, but reversed in time. Both pulses of the pair were of the same carrier

frequency, peak amplitude, and duration. Their duration was gradually diminished till the observers were unable to detect any difference between them. In the carrier frequency region of 100 Hz to 10 kHz, the observers were able to discriminate some pulse shapes even down to a duration of 2 msec. On the basis of these experiments, W. Türk deduced that: two mechanisms take part in envelope curve discrimination, namely sound quality perception, and detection of loudness; for the pulses longer than about 10 msec discrimination relies mainly on the perception of the time order of the events of different loudness in time; the stimuli which are shorter than about 10 msec are perceived as one unit; the envelope discrimination of such short pulses depends on their spectrum, as the discrimination in this case is based mostly on the quality of the perception.

How the detection of the envelope curve variations depends on the carrier frequency was investigated by E. Hojan and A. Rozsypal (1967). They used as stimuli tone pulses with a smooth onset and a stationary portion of the duration 4 sec. These pulses were terminated with one of two types of offset transient. The undulated transient (Figure 6.12 a) was of the envelope type

$$\int \frac{\sin x}{x} \, dx \quad ,$$

and the smooth transient (Figure 6.12 b) of the envelope type
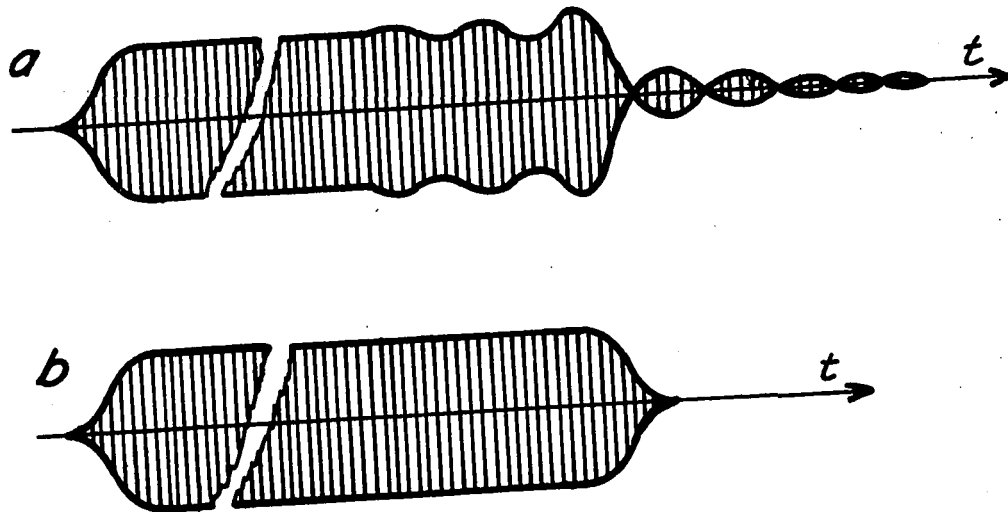
$$\int e^{-x^2} \, dx \quad .$$

Figure 6.12   The undulated transient stimulus (a) and the smooth transient stimulus (b) used in envelope recognition experiments.

E.Hojan and A.Rozsypal (1967).
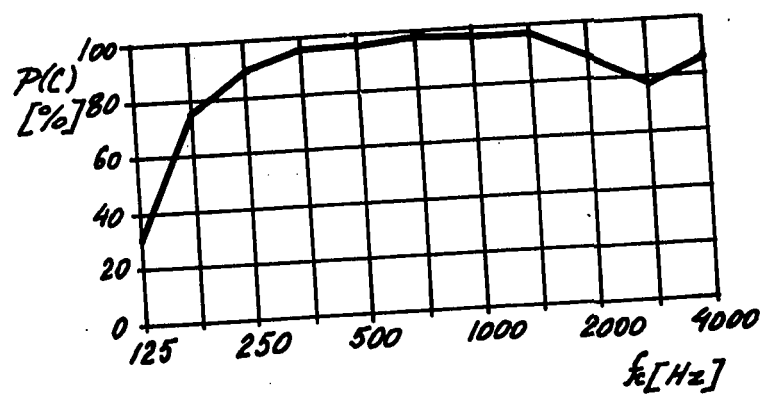
Figure 6.13   The average rate of correct identification P(C) of the undulated transient from Figure 6.12 as a function of the carrier frequency f_c.

E.Hojan and A.Rozsypal (1967).

The decay time of both transient types was the same, namely 20 msec. Thus in the case of the undulated decay, the undulations corresponded to modulation frequency approximately 25 Hz. The stimuli were presented in a random order regarding both pulse decay type and carrier frequency, which was in the region from 125 to 4000 Hz in half-octave steps. The pulses were reproduced by a loudspeaker system in an anechoic chamber at the loudness level of 60±5 Ph. The interval between stimulus presentations was 6 sec. The observer's task was to determine, after listening to each pulse, whether it was terminated by a smooth or undulated transient. The results of this experiment revealed the dependence of the undulated transient correct recognition on the carrier frequency of the stimulus. This recognition is at its best in the central region of the audio frequency range, as can be seen in Figure 6.13. At 2800 Hz a local minimum was observed for six subjects out of ten taking part in the testing. These results suggest that two separate mechanisms might take part in envelope detection, with overlapping around 2800 Hz. The perception of the undulated decay varied for different carrier tone frequencies. For the lowest frequencies measured, namely 125 and 180 Hz, the subjects perceived the transient as a roughening of the tone instead of detecting of the separate peaks of the envelope curve. By increasing the carrier frequency the tone in the decay portion of the pulse became purer and the transient was perceived as clear variations of the tone intensity,

especially in the range from 710 to 2000 Hz. The duration of the transient for frequencies 2800 and 4000 Hz seemed to be shorter than for lower carrier frequencies, which might indicate that not only the envelope perception, but also the subjective time scale depends on the frequency spectrum of the stimulus.

6.4. Discussion of the Envelope Discrimination Results

In principle, the same clues are available for envelope discrimination as were mentioned in the discussion on duration discrimination, namely: duration of the stimulus, temporal structure of the stimulus envelope, loudness of the stimulus, and spectral composition of the stimulus. The duration of both stimuli to be discriminated was varied simultaneously to make their effective duration equal. This fact eliminated the stimulus duration clue and reduced the weight of the stimulus loudness clue.

Our results indicate that the time-frequency structure of the stimulus contributes to envelope discrimination in the range of durations obtained in our envelope discrimination experiment. For most of our tone stimuli the critical duration falls into the interval from 2.5 to 3.0 msec. For noise carriers this value is higher, between 3.3 and 3.8 msec. Our actual and effective stimulus durations were identical only for the rectangular pulse. For all types of triangular pulses the actual duration and consequently also the separation of the onset and offset transients was twice the effective duration of the stimulus. The longest value

of this time interval was 7.2 msec for the combination of
rectangular and isosceles triangle envelopes modulating the
white-noise carrier.

The results of W.Türk (1940), described earlier in Sec-
tion 6.3, suggest that the stimuli shorter than 10 msec are
perceived as one time event. But his subjects were able to
discriminate time-reversed signals even of duration 2 msec.

The same values were reported also by D.A.Ronken (1970).
In his series of signal power and signal phase discrimination
experiments a transient signal consisting of two 250 μsec
rectangular pressure pulses was used as the basic stimulus.
The time delay between the two pulses was one of the stimulus
parameters. In the case that these two pulse components were
separated by more than about 10 msec such a stimulus lost its
unitary character and was perceived as two successive clicks.
In a phase discrimination experiment the two pulse components
were of different amplitudes. For 2 msec delay between
pulses the observers distinguished the temporal order of such
stimuli when the amplitude difference between the pulses was
about 6 dB. For 1 msec delay between the pulses, some of the
trials, because of the adaptive psychophysical procedure
used in the testing, were unsuccessful. Successful trials
resulted in a required amplitude difference between pulse
components of about 8 dB.

The experiments of G.A.Gescheider (1966)(1967) on
auditory and cutaneous temporal resolution of clicks have
shown that two successive clicks of equal amplitude

presented monaurally were resolved as temporally discrete
when they were separated by 1.6 msec at 60 dB sensation level
and by 3.8 msec at 10 dB sensation level. The optimal
temporal resolution was obtained for the amplitude of the
first click 5 to 10 dB lower with respect to the amplitude
of the delayed click, although the curve showing the depend-
ence on the relative amplitudes of the clicks was rather flat.

Also J.H.Patterson and D.M.Green (1970) arrived at the
just discriminable stimulus durations similar to our critical
durations. They have recently presented data on discrimi-
nability of brief transient stimuli with identical energy
spectra but different phase spectra. These stimuli were com-
puter generated using the procedure of D.A.Huffman (1962),
which involves passing a pair of digital impulses through an
all-pass digital filter having a flat amplitude characteristic
and a variable phase characteristic. Such a filter delays only
that portion of the signal energy which is contained in the
frequency bands where the rapid transitions of the phase
characteristic occur. Two stimuli with transitions of the
phase characteristic located at different frequencies were
compared. Highly experienced subjects averaged 75 per cent
correct discrimination for stimulus duration of about 3 msec.
For stimulus duration 10 msec the correct discrimination
rate was almost 100 per cent. The authors attribute the
discrimination to the differences in temporal order of
arrival of the energy from different frequency bands, or, in
other words, to the differences in the short-term spectra

of the stimuli.

D.M.Green (1971) recently presented a general review of temporal resolution in the auditory system. He indicated, that the shortest time interval within which the ear can discriminate two stimuli of identical energy spectra but of different phase spectra is about 1 to 2 msec.

We believe that in our envelope discrimination experiments, as well as in the experiments just mentioned, the discrimination of the stimuli in the millisecond duration range was based on detected differences in the short-term spectra. In our case of the discrimination between stimuli with convergent and divergent triangle envelopes, as in the experiments of the other investigators in which stimuli of identical energy spectral densities were discriminated, neither the Fourier amplitude spectrum, nor the Fourier energy spectral density can indicate any difference between the signals at all. The discrimination can be based only either on the phase spectra, or on the short-term spectra with the memory function (Equation 2.1) of an effective duration in the range of several milliseconds.

The results of our first experimental series, where tones were used as carriers, lead us to several interesting inferences. For the lower frequency carriers, the critical effective duration is nearly the same for all measured envelope combinations, namely 2.7 msec for carrier frequency 250 Hz, and about 2.8 msec for 1 kHz. That means that, except for rectangular envelope stimuli, the actual stimulus lasted

about twice that long. The period of the 250 Hz carrier signal is 4 msec. Evidence based on neurophysiological data indicate that neural firings occur only on one polarity of basilar membrane motion. Bearing this in mind we realize that the envelope was sampled at most twice per actual stimulus. Even if these two samples are regarded as independent, they can carry only a small amount of information about the original envelope function.

As Figure 6.14 illustrates, at these short durations the stimuli were in fact reduced to what can be more properly called pressure pulses, either monopolar or bipolar, depending on the relative phase between carrier and envelope. Recalling the fact that the apparatus used for stimulus generation did not provide for phase lock between these two signals, we realize that the discrimination in this case was based on the actual signal waveform, which strongly depended on the random phase between the carrier and modulation signals. The subjects probably signalled "no difference" when the signals to be discriminated were short enough to contain one peak only. On the other hand, during the ascending test run, the observers detected the difference between two signals as soon as they contained two pressure peaks, since there was high probability that the time order of these peaks of different amplitudes varied between two successive stimuli. The fact that the critical duration at carrier frequency 250 Hz is practically independent on the envelopes combination speaks in favor of this explanation.
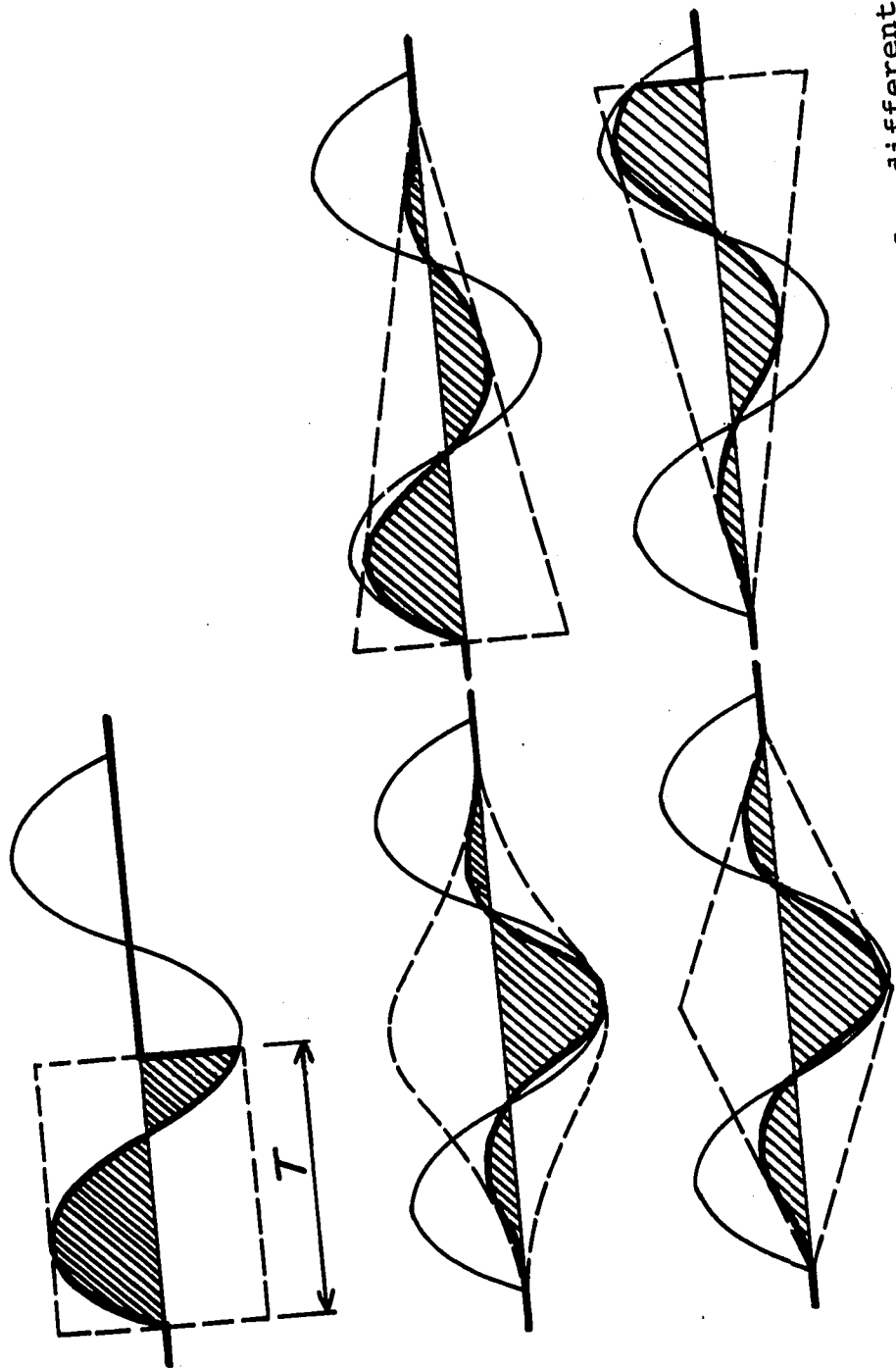
Figure 6.14 Examples of the stimulus waveforms (heavy lines) for different envelopes (intermittent lines) and carrier frequency 250 Hz. The effective stimulus duration T is 2.7 msec.

For the carrier frequency 1 kHz the envelope sampling is four times denser than at 250 Hz. At the critical duration the envelope is sampled about five times per stimulus. In view of Figure 6.14, this is apparently not dense enough to produce different critical durations for different envelope pairs. The only exception is the envelope pair CG, i.e, convergent triangle and Gaussian envelope. This is the first indication, from our experiments, that the ear is unable to obtain as much information from the stimulus onset transient as from the offset one. For this envelope combination the value of critical duration is 3.0 msec, as compared with 2.8 msec for the combination of divergent triangle and Gaussian envelopes. Obviously, a transient at the beginning of the convergent triangle stimulus is less detectable than the same transient placed at the termination of the divergent triangle stimulus. For the carrier tone 4 kHz, a marked spread of critical durations was observed, from 1.4 to 3.0 msec. In general, envelope discrimination is at its best at this frequency, the average critical duration being 2.3 msec. Here again, as in the case of 1 kHz carrier, we obtained the poorest discrimination for the envelope pair: convergent triangle, Gaussian envelope. The 3.0 msec value of critical duration for this combination of envelopes shows no improvement from the result at 1 kHz. Next in order of improving discriminability is 2.5 msec critical duration for the envelope pair: rectangular, isosceles triangle; 2.4 msec for the pair: divergent triangle,

Gaussian envelope; and 2.2 msec for the pair: rectangular, Gaussian envelope. Finally, the shortest critical duration 1.4 msec occurred with the combination: convergent, divergent triangle. Also for white noise carrier does this envelope pair show the shortest critical duration, 3.3 msec. In general, for white noise as carrier signal, the critical durations are longer: from 3.3 to 3.9 msec.

In the tone carrier series of our envelope discrimination experiment all three possible pair combinations of the envelope triplet: Gaussian pulse, convergent triangle, divergent triangle, were tested. From these results we can conclude that, in relation to the smooth Gaussian envelope, the abrupt envelope jump at the stimulus leading edge is less detectable than a jump of the same size located at the stimulus trailing edge.

Our envelope discrimination experiment is not the only experiment indicating that, in hearing, the initial portion of the transient stimulus conveys less information than does the subsequent portion of the stimulus. I.V. Nábĕlek, A.K. Nábĕlek, and I.J. Hirsh (1970), interpreting the results of their experiments in which the frequency of a steady-frequency comparison tone burst was matched to a test tone burst containing a linear frequency glide between initial and final frequency, arrived at similar conclusions. As long as the frequency difference and the duration of the glide were not too large, a single pitch was assigned to such stimuli. In the cases where the frequency glide spread either over the

whole stimulus duration or over the central portion of it, the pitch judgments were consistently shifted toward the final frequency of the gliding frequency stimulus. This shift was more pronounced for rising than for falling frequency glides. On the other hand, the subjects judged the central frequency of the glide only in the cases where the frequency glide lasting one quarter of the stimulus duration took place near the end of the stimulus. The authors concluded that, for the pitch judgement, the final portion of the stimulus has a larger weight and that the excitation pattern of the receptor at the end of the stimulus was decisive for pitch perception.

The data obtained by P.T. Brady, A.S. House and K.N. Stevens (1961) also indicate that the initial part of the transient stimulus does not contribute to the perception as much as its final one. Using the method of matching they compared the perceptions of two single formant speech-like sounds. The stimuli were generated by a resonant circuit excited by five pulses simulating the glottal wave of a 100 Hz pitch. During the generation of the test stimulus the resonant frequency of the resonant circuit was shifted either upward or downward, between frequencies 100 and 1500 Hz. This frequency glide was of 20 msec duration. Six different positions of the glide in relation to the glottal pulses were tested. Observers controlled the tuning of another resonant circuit generating the comparison stimulus. Their task was to match the perception evoked by the time-varying

test stimulus by appropriate adjustment of the time invariant resonant frequency of the comparison stimulus. The subjects exhibited a strong tendency to tune the resonant frequency of the comparison pulse to a value corresponding to the final frequency of the gliding frequency stimulus, particularly in the cases when the frequency glide took place in the initial phase of the test stimulus.

A very interesting paper, relevant to the interpretation of our envelope discrimination experiment, was published recently by T.S. Korn (1969/70), in which he investigates the resolving power of hearing in time and frequency domains. He introduces the concept of the "Elementary Messages of Discrete Frequency", EMDIF in short. These are derived as the inverse Fourier transforms of the extrapolated masking curves, as obtained by R.L. Wegel and E.E. Lane (1924) for sustained pure tones. The object of the extrapolation was to remove the irregularities of the masking curves caused by the beats between the maskee and the masking tone and its harmonic frequencies. As the masking curves supply no information about the phase relationships of the spectral components constituting the EMDIF signal, this phase information was arbitrarily supplied. From the several possible alternatives of the EMDIF signals found in this way, the best results in listening tests were obtained for the time function

$$S_0(t) = t^2 e^{-\alpha t} \cos 2\pi f_c t, \qquad (6.1)$$

where f_c is the carrier frequency of the stimulus and α is the parameter to be chosen in order to obtain the optimal

auditive effect for different carrier frequencies f_c. We may note that the part of this function determining the envelope of the EMDIF signal has the same form as the basilar membrane weighting function $g(t)$, given by Equation 2.2. The only difference between these two functions consists in the about 30 times smaller value of the parameter α in the EMDIF envelope, which means that the EMDIF signal is about 30 times longer.

The amount of pitch information and the click content of signals of this type depend to a large degree on the value of α. When the value of the parameter α during listening tests was gradually decreased, i.e. the duration of the stimulus was increased, the signal was first perceived as a "white click" with no pitch information. For signals of longer durations the audible effect was described as a "percussion on a wooden board", later as a "xylophone sound" with more and more pronounced pitch. Eventually, for certain α, depending on the carrier frequency f_c, the stimulus was perceived as an instantaneous appearing of the pure sine tone and its subsequent instantaneous disappearing, with no click component and no duration information about the stimulus. The signal of this duration was defined as an EMDIF signal. For even longer durations the stimuli were perceived as gradually appearing and disappearing pure tones with definite durations. Signals longer than EMDIF signals obviously do not bring to the listener more frequency information than EMDIF signals, in spite of the fact that their spectra are

more selective than the spectra of EMDIF signals. So, once
the full frequency slectivity of hearing is reached, as
given by the steady state masking curves, the further pro-
longation of the stimulus does not provide the listener any
more spectral information about the stimulus, and the audi-
tory analyzer obviously switches its attention to register-
ing the duration of the stimulus and to detection of its
envelope changes. T.S. Korn's experiments at carrier fre-
quencies 400, 1200, and 4000 Hz indicated that the duration
of the EMDIF signals is roughly inversely proportional to
the frequency of the carrier tone. In other words, the
EMDIF signals contain equal number of carrier signal periods,
independent of the carrier frequency.

Precisely speaking, the EMDIF signal is of infinite
duration, but its effective duration is limited and depends
on the level of background noise. About 52 carrier signal
periods are spaced between the envelope onset and offset
points 20 dB below the envelope peak level. The EMDIF
envelope function $e_E(t)$ and its energy spectral density
$E_E(\omega)$, expressed in terms of the effective stimulus duration
T, are given by

$$e_E(t) = \begin{cases} \dfrac{e^6}{16} \left(\dfrac{t}{T}\right)^2 \; e^{-\frac{e^2}{2}\frac{t}{T}} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

$$E_E(\omega) = T^2 \; \frac{e^{12}}{(e^4 + 4\omega^2 T^2)^3}$$

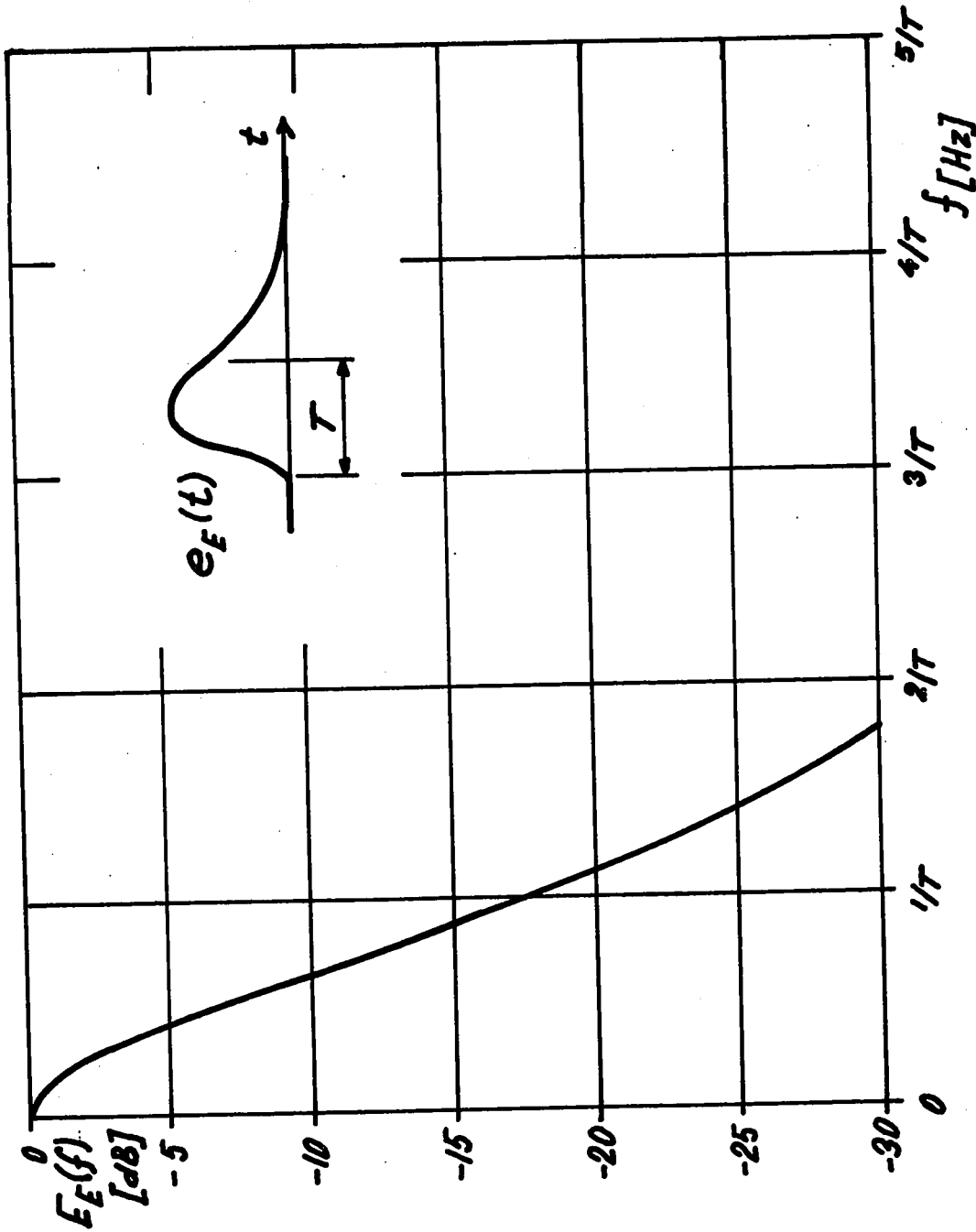and illustrated in Figure 6.15.

Figure 6.15 The EMDIF (elementary message of discrete frequency) envelope

$e_E(t)$ and its energy spectral density $E_E(f)$.

Thus the EMDIF signal is the shortest signal supplying discrete frequency information to the listener and represents the useful length of observation for pure tone signals. The EMDIF signals seem more natural elementary signals to the auditory organ than the "audio information quanta" with Gaussian envelopes as suggested by D. Gabor (1946) (1947). Namely the time symmetrical Gaussian envelope signals evoke a perception which appears much longer in its onset part than in its decay one. T.S. Korn ascribes to hearing some phase sensitivity, as the time unsymmetrical EMDIF signals, if presented to the observers reversed in time, evoke the perception of a tone of certain duration gradually increasing in intensity and terminated by a click.

At this point the suggestion may be put forward that the time and frequency resolution of transient signals, as performed by hearing, should not be derived from the masking curves measured for sustained stimuli. To the best of the author's knowledge the masking curves when both masker and maskee signals are of a transient character and over-lapping in time have not yet been measured, so the exact evolution of the masking curves in time is not known.

Evolution of masking curves in time appears to be a crucial factor in auditory time-frequency analysis of transient stimuli. In Chapter VII we will present further data on the dynamical properties of hearing and eventually give a model of evolution of critical bands following the stimulus onset.

The data observed in noise carrier envelope discrimination experiment suggest that the critical durations tend to increase with increased carrier spectral bandwidth, see Figure 6.5. Apparently, the wider the spectral bandwidth of the carrier signal, the less reliable is the cue offered by the differences in perceived stimulus quality, and the more the hearing has to rely on the time waveform of the stimulus. Corresponding to this uncertainty in the frequency domain, the random carrier signal introduces uncertainty into the time domain of the stimulus representation, as the signal envelope becomes smeared. The rate of random changes superimposed on the original envelope function is proportional to the bandwidth of the noise carrier. In the extreme case, the stimuli formed by modulating a continuous white noise offer no spectral clue to the receiver at all. Spectra of amplitude modulated stimuli are given by a convolution between the spectra of the carrier and the envelope, see Equation 5.6. When the spectrum of one of these signals is uniform over the whole audio range, so is the spectrum of the resulting signal. This explains why, for noise carriers, especially for the one-octave and white noise, the critical duration is longer than for pure tone carriers. The duration of wide-band noise stimuli had to be extended in order to bring them into the range where the detection of temporal changes of the stimulus envelope alone, without the help of spectral clues, can offer enough information to make the observers' decision switch from "identical" to "different."

Chapter VII

FUNCTIONAL MODEL OF TIME ORGANIZATION

OF THE TIME-FREQUENCY ANALYZER

As already mentioned several times in the previous chapters, the hearing analyzer seems to gradually adjust its frequency selectivity to the spectrum of the incoming signal. In this chapter a functional model of this adaptable filtering process will be presented. This model is intended to simulate the results of several psychoacoustical experiments on the dynamical properties of hearing, to be described in the following section.

7.1 Frequency Selectivity of Hearing as a Function of Stimulus Duration

Dependence of the frequency selectivity of hearing on the frequency and duration of sinusoidal stimuli was investigated by B.L. Cardozo (1962), Liang-Chian and L.A. Chistovich (1961), R. Oetinger (1959), and D.A. Ronken (1971). All these investigators measured the frequency discrimination of the human ear as a function of the stimulus duration. The results of these three experiments are plotted together in Figure 7.1, where Δf represents the measured differential threshold in frequency and Δt the stimulus duration. The parameter of these curves is the carrier frequency and the envelope type of the stimulus. Even though the curves from
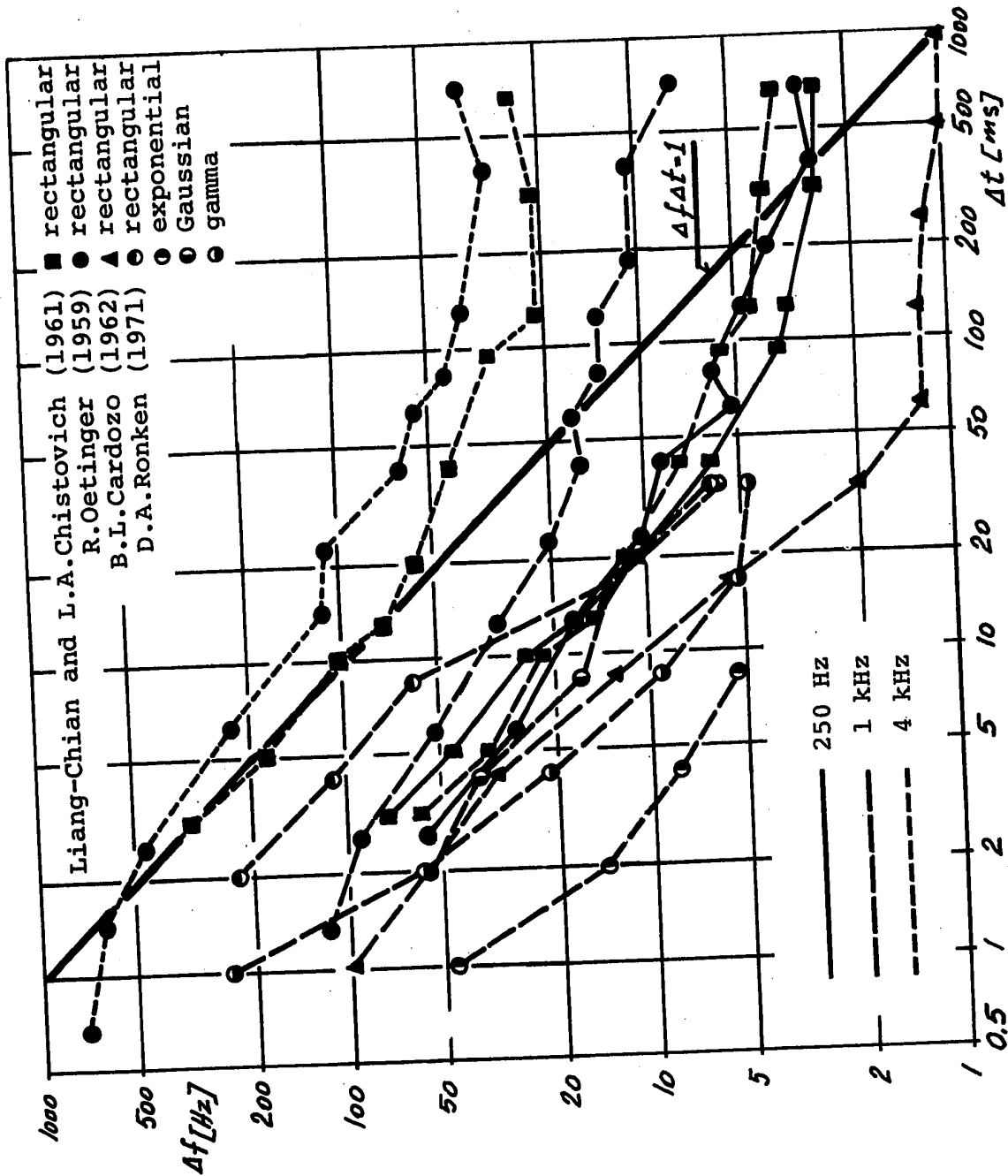
Figure 7.1 Just noticeable difference in frequency Δf as
a function of the tone stimulus duration Δt.
Tone frequency and envelope are the parameters.

different experiments corresponding to a given carrier frequency are displaced relative to one another, presumably due to different definitions of the effective stimulus duration and other experimental differences, the course of all curves is similar. Up to a certain stimulus duration all curves follow approximately the uncertainty principle of observation, Equation 2.4, as the data curves in this region are nearly parallel to the straight line $\Delta f . \Delta t = 1$ plotted in the same diagram. It is evident that, for short signals and lower carrier frequencies, frequency discrimination by the human ear is more than one order better than by linear systems, as some of these data lie below the uncertainty principle limit.

Prolonging the stimulus duration above some critical value, which in this case lies within the interval between 50 to 300 msec, brings no substantial improvement in frequency discrimination. The critical duration of the stimulus represents the duration of the time window of the structure carrying out the frequency analysis. Liang-Chian and L.A. Chistovich recognized on their curves even three linear sections distinguished by different steepnesses, delimited by two critical durations T_1 and T_2. While the critical duration T_1, with value between 7 and 41 msec, decreased with increasing carrier frequency, the value of T_2 depended on frequency to a lesser degree as its value lay between 123 and 188 msec, decreasing with frequency. In the region of the shortest stimulus durations, up to duration T_1, the spectral bandwidth of the

stimulus is wider than the bandwidth of the frequency analyzing mechanism of hearing. Consequently, the authors explained, the frequency difference limen is inversely proportional to the stimulus duration here, as the frequency difference limen is expected to be proportional to the stimulus spectral bandwidth. The steepness of the curves in this region, on a double logarithmic plot, is minus one. The critical stimulus duration T_1 is a measure of the spectral selectivity of the mechanical frequency analyzing elements of hearing and can be regarded as a time needed to reach the steady state of the mechanical structures of the basilar membrane. On the other hand, the authors ascribed the critical duration T_2 to the processes in the auditory neural network. They were of the opinion that any extension of the stimulus duration above the critical value of T_2 does not contribute to a more accurate determination of the stimulus frequency since during the time interval T_2 the frequency analysis, both at its mechanical and neural stages, is fully completed. Consequently, the jnd in frequency is constant for stimulus durations greater than the critical value T_2.

In the intermediate region of stimulus durations between T_1 and T_2 the frequency difference limen is inversely proportional to the square root of the stimulus duration. The curves in this range can be approximated with a straight line of steepness minus one half. The authors explain this relationship in the following way: the central auditory system averages the continuously arriving fluctuating data about the stimulus

frequency over a time interval T_2, so that the variance of the final frequency estimation is inversely proportional to the number of frequency readings taken during the stimulus duration.

As observed by G. von Békésy (1960, p.454), the mechanical structure of the basilar membrane acts as an approximately constant Q tuned system with distributed parameters. Its absolute tuning bandwidths are narrow at low frequencies, and wide at high frequencies. As the temporal resolution is inversely proportional to the frequency resolution, the temporal resolution improves with increasing frequency. This fact accounts for the wide interval and decreasing trend with frequency of the time constant T_1 observed by Liang-Chian and L.A. Chistovich (1961).

7.2 Temporal Summation of Loudness

Integration times comparable with T_2 were measured also by W.R. Garner and G.A. Miller (1947) and by D.M. Green, W.P. Tanner Jr. and T.G. Birdsall (1957) in experiments investigating the masked threshold of pure tones as it depends on stimulus intensity and stimulus duration. Their data have shown that hearing integrates the intensity of the received signal within the time interval of the duration from about 100 to about 200 msec. The same integration times were found by I. Pollack (1958) while measuring the loudness of bursts of white noise as it depends on their duration and also by L.A. Chistovich and V.A. Ivanova (1960) in experiments dealing with jnd in intensity as a function of signal duration.

Some researchers report the time constant of temporal summation of loudness to be dependent on stimulus intensity and frequency. G.A. Miller (1943) noticed that for short bursts of noise its value is about 200 msec for stimuli near the threshold of audibility. Increasing the stimulus intensity into the range of moderate and high suprathreshold loudness levels the time constant decreases to less than 100 msec.

C.S. Watson and R.W. Gengel (1969) studied the dependence of the loudness summation time constant on stimulus frequency. They measured the detection threshold of tones of durations 16 to 1024 msec in the presence of a masking noise of sound pressure level 30 dB in the contralateral ear. The integration time constant was found to range from 125 to 175 msec at low stimulus frequencies from 125 to 250 Hz. But it systematically decreased with increased frequency, down to values from 40 to 60 msec at stimulus frequencies 3 to 9 kHz.

7.3 Lateral Inhibition as the Contrast Enhancing Mechanism in the Frequency Domain

The frequency selectivity of the auditory analyzer and the masking curves are obviously closely related. So are also the gradual increase in frequency selectivity of both in time.

The process of time-frequency analysis is carried out in two stages. The first stage is performed in the inner ear by the mechanical system of the basilar membrane. Here the selectivity in both domains is determined by the previously-mentioned basilar membrane weighting function calculated by J.L. Flanagan and given by Equation 2.2. So in the initial

phase of the stimulus onset, until the steady state of the
basilar membrane is reached, the "frequency window" and so
also the corresponding masking curve for transient stimuli
of comparative duration is obviously rather broad, broader
than the frequency selectivity obtainable by the basilar mem-
brane acting as a linear mechanical system in response to
sustained stimuli. Only when the preanalysis performed by
the basilar membrane reaches its frequency resolution limits,
after a duration of stimulus equalling the effective duration
of the basilar membrane time window has elapsed, does some
other mechanism set in. This mechanism, located most probably
in the auditory neural network, could be based, for instance,
on lateral inhibition described by G. von Békésy (1967),
W. Reichardt and G. MacGinitie (1962), A. Rozsypal, V.
Majerník and V. Balko (1969). A less likely mechanism is
coincidence filtering suggested by R. Schief (1963). The
latter mechanism can hardly explain edge effects observed by
A. Rakowski and A. Rozsypal (1968) in experiment in which the
pitch of low-pass noise bands were matched to that of a pure
tone. In this study the observers most frequently adjusted
the frequency of the comparison tone to the cut-off frequency
of the noise filter, which had roll-off steepness 1 dB/Hz.
In any case, both such mechanisms can be easily visualized
as performed by properly interconnected neural elements, as
is known from neurophysiological investigations to be the case
for other sensory systems. Due to delays, introduced mainly
by the integration properties of synaptic junctions, the

response of the neural network is slower than the response of the basilar membrane. The masking curve is gradually built up and its bandwidth is progressively narrowed, eventually reaching the frequency selectivity as given by the masking curves for steady tones. This state is reached in a time interval corresponding to the duration of the EMDIF signals. The required skewness of the envelope curve of the EMDIF signals, and hence the sensitivity of hearing to the phase of the EMDIF signals, can be attributed to the fact that in the initial phase of the stimulus the evolving masking curve has not yet reached its steady state frequency selectivity. Because at this stage the masking curve is broader than in its steady state, the auditory analyzer tolerates the onset steepness of the stimulus envelope to be higher, i.e. to have a wider short-term spectrum, than the steepness of the decay part of the stimulus, without producing a perceptible click. Hence the phase sensitivity of hearing to transient stimuli can be explained as a time-varying process during which the receiver system is being adapted to the spectral composition of the stimulus.

Lateral inhibition is a common feature of sensory organs, G. von Békésy (1967). It is a consequence of the lateral inhibitory interconnections of sensory nerve fibers. Lateral inhibition may occur at several neural levels. Its mechanism is not known in complete detail. Probably, it is a combination of peripheral and central processes. The resulting effect of lateral inhibition is that the neural excitation as a function

of stimulus frequency has a much sharper maximum than the vibration pattern of the basilar membrane in the cochlea, so that the frequency resolution of hearing at its neural levels is much higher than the resolution calculated from the mechanical properties of the basilar membrane.

The mechanism of lateral inhibition will be explained in terms of a mathematical model suggested by A. Rozsypal, V. Majerník, and V. Balko (1969).

Let us assume a row of m neurons N_i, see Figure 7.2. Excitation y_i of these neurons is proportional to the deflection of the basilar membrane and comes from the outputs of the sensory cells S_i arranged on the basilar membrane. Let us assume a stationary stimulus causing the basilar membrane to deflect in a pattern $s(\ell)$, which is a function of the basilar membrane spatial coordinate ℓ. Spacing of the sensory cells on the basilar membrane is $\Delta\ell$. Let us further assume that the firing rate s_i of the sensory cells S_i is proportional to the deflection of the basilar membrane at the place $i(\Delta\ell)$

$$s_i = s \ (i(\Delta\ell))$$

This neural activity is transmitted by a excitatory synaptic junction to a corresponding neuron N_i in a direct neural path. But this neuron is also inhibited by the spatial neighbors of S_{i-1} and S_{i+1}. Signals from these sensory cells are fed into inhibitory synapses of neuron N_i. Assuming linearity between the input and output of the neurons, the output z_i of the neuron N_i can be expressed as

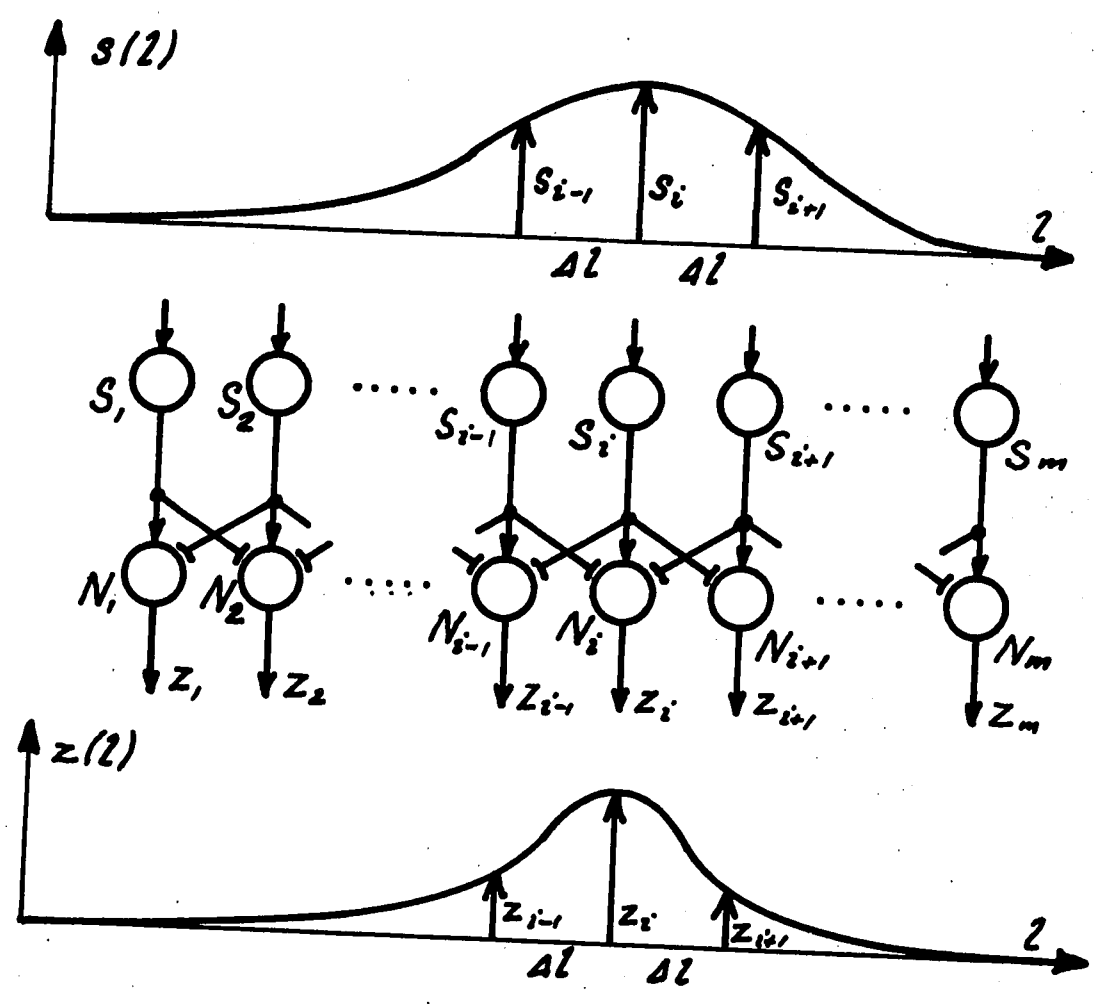$$z_i = c \ [s_i - k(s_{i-1} + s_{i+1})], \tag{7.1}$$

Figure 7.2 Model of the lateral inhibition neuron network.

where c is a coefficient depending on the sensitivity of the sensory cells and on the ratio between the input and output rates of the neurons. Coefficient k characterizes the weight of the inhibitory inputs with respect to the excitatory input in the neurons.

Using the first four terms of the Taylor series to express the excitation function $s(\ell)$ in points (i-1) and (i+1) and substituting these expansions into Equation 7.1, we obtain

$$z_i = c[s_i(1-2k) - k(\Delta\ell)^2 \frac{d^2s(i(\Delta\ell))}{d\ell^2}] \qquad (7.2)$$

This result shows that the output z_i of the row of neurons N_i is a scaled superimposition of the original deflection function $s(\ell)$ of the basilar membrane and its negative second derivative with respect to the length of the basilar membrane. The contrast enhancing effect of such superimposition is illustrated in Figure 7.3. The value of the coefficient k determines the ratio between these two components. For k = 1/2 the output consists only of the second derivative component.

7.4 Organization Time of the Auditory Analyzer

The following experiments offer valuable data about the time course of the organization of the auditory analyzer after the stimulus onset and about the time required for the decomposition of this organization following the stimulus termination.

In a study on a masked thresholds of tone, noise, and pressure pulses H. Scholl (1962a) concludes that the masked threshold of pulses shorter than 5 msec is determined by the
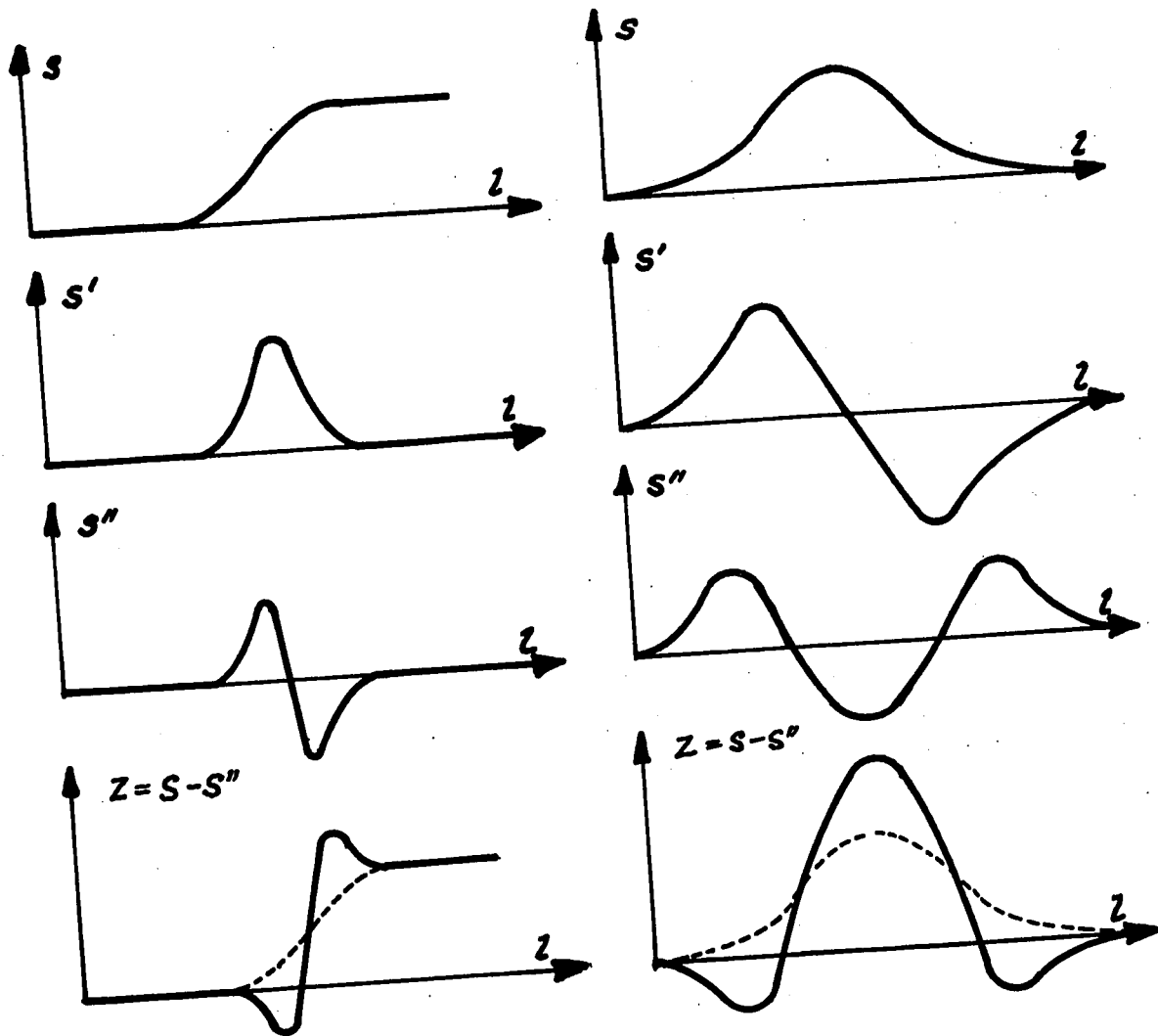
Figure 7.3 Improvement in frequency selectivity between the deflection pattern of the basilar membrane s and and the neural firing rate z, in a lateral inhibition neuron network.

energy of the pulse only, and is independent of its frequency spectrum. The effect of the formation of critical bands begins to appear only for stimuli longer than 5 msec. In another paper H. Scholl (1962b) investigated in more detail the time course of the composition of critical bands following the stimulus onset. In masking experiments he measured the build up of frequency selectivity as a function of stimulus duration. He found the time constant of the initial period of the critical band formation process to be about 10 msec. The full organization of critical bands was reached in 300 msec.

E. Zwicker (1965a)(1965b) in his experiments on temporal effects in simultaneous masking investigated to what degree a short signal pulse is masked by a longer masking burst. Relevant to this discussion are those of his experiments in which the masked threshold was measured as a function of the time delay between the onsets of the masker and the maskee bursts. The most pronounced dependence of the masked threshold on this delay was measured for short maskee signals of the duration 2 or 3 msec, masked by 600 msec masking bursts. In the cases when the spectral bandwidth of both maskee and masker covered the same number of critical bands, the masked threshold was independent of the delay. This was shown in experiments in which both maskee and masker were white-noise pulses, or, in which the masker was a narrow-band burst and the maskee signal a short tone pulse of the masker central frequency sent through the same filter as the masker noise.

Lack of dependence on the delay time was also observed when the narrow-band noise masker and the tone pulse maskee were located at different frequencies. Quite another situation appeared in the cases when the spectral bandwidths of the maskee and masker differed. The masked threshold of a 2 msec maskee tone pulse of frequency 5 kHz located at the onset of the white-noise masker pulse was 12-13 dB higher than for onset delay 300 msec, see Figure 7.4. The wider the spectral bandwidth of the maskee pulse, the smaller the difference observed between the masked thresholds for the maskee in the masker onset position and for the maskee delayed by 300 msec. In all cases the masked threshold for 300 msec delay was equal to the threshold for continuous masking noise. So the 300 msec delay can be regarded as the time interval required by the analyzer system to reach the steady-state conditions.

The data from one of E. Zwickers experiment are displayed in Figure 7.4, which represents the masked threshold shift ΔL_m of a 5 kHz tone pulse of 2 msec duration masked by a white-noise masker pulse of 600 msec duration. The independent variable Δt is the delay between the onset of the masker and the onset of the maskee. The data (1965a - circles, 1965b - crosses) were replotted in a log-lin graph in order to faciliate the estimation of the time constants of the threshold shift. The time constant of the initial 6 dB exponential decay is about 24 msec, the following 6 dB the threshold decays with time constant about 188 msec. The steady state -12 dB is reached in 300 msec. Design of our model of time organization of

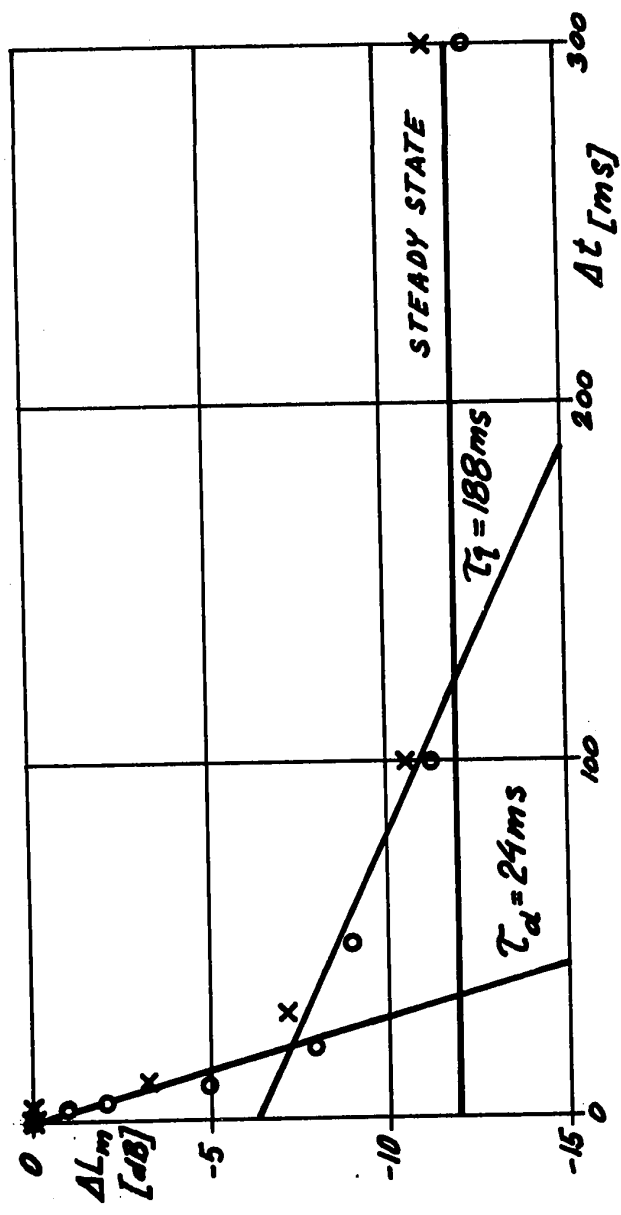Figure 7.4 Shift of the masked threshold ΔL_m of a 2 msec tone of 5 kHz frequency as a function of the delay Δt between onsets of the 600 msec white-noise masking burst and the maskee. Data taken from E.Zwicker (O-1965a, ✗-1965b).

the auditory analyzer is based on these data.

L.L. Elliott (1967) investigated the time required by the auditory analyzer to develop steady-state frequency contours as a response to a narrow-band stimulus. In four separate experiments forward, backward, and simultaneous masking situation was studied. The masked threshold for tones of different frequencies and durations was measured using the method of adjustments. Bursts of noise were used as a masker, either wide-band noise, or narrow-band noise of bandwidth approximately one critical bandwidth.

J.J. Zwislocki (1960) has hypothesized that a time interval on the order of 200 msec is necessary for the complete cessation of residual neural activity after stimulus termination.

The results of these experiments indicate that an organization time on the order of 250 to 300 msec is required by the auditory system to get tuned-in to the narrow-band stimulus. This tuning time is shorter in the cases where the frequency contours established by the preceding stimulus of the same or similar frequency composition are not completely decomposed at the time of arrival of the new stimulus. The organization of the auditory system due to the narrow-band stimulus persists for about 300 msec after stimulus termination.

Although the loudness summation time constant and the time required by hearing to organize its frequency contours seem to be of the same duration, it is difficult to assert, at the present state of knowledge, whether these two time

characteristics specify one process or two different and independent processes.

7.5 Short-Term Adaptation as the Contrast Enhancing Mechanism in the Time Domain

It is quite probable that the ear's sensitivity to amplitude changes, as long as the observer is not primarily interested in the frequency content of the signal, is derived directly from the output of the sensory cells located on the basilar membrane. The basilar membrane responds about 30 times faster than the subsequent neural stages, as reflected in the different values of the parameter α in the formulae for the basilar membrane time window and the envelope of the EMDIF signals, as well as from the differences between the two time constants T_1 and T_2 measured by Liang-Chian and L.A. Chistovich (1961). Only a system with response time of the duration comparable to the basilar membrane time window is able to discriminate between stimulus envelopes as short as the critical duration obtained in our experiments. This process may be further aided by the short-time adaptation, actively participating in the perception of transient signals. In order to achieve the optimal intensity resolution, the sensory transducers are biased to match the range of the maximal differential sensitivity of the receiver to the stimulus level. This adaptation process represents amplitude normalization of the incoming signal in the mathematical sense, or automatic gain control, familiar in electronic engineering.

In our experiments, both in envelope discrimination and

in duration discrimination, we have found an indication
supporting this idea. Our narrow-band noise carrier signal
resembled a sinusoidal wave with slow random variations
of its envelope curve. Within several quasi-periods of such
a signal the envelope amplitude could be regarded as constant.
For instance, under the condition of stimulus duration of
only a few periods of a carrier signal central frequency and
of a sufficiently long interstimulus interval, the rectangular
pulses generated by time filtering from continuous narrow-band
noise appeared on the oscilloscope screen as a series of pulses
of sinusoidal carrier with constant amplitude within each
single pulse. However, this amplitude varied at random from
pulse to pulse. Very often the successive stimuli in a series
differed in amplitude as much as by 10 dB.

In spite of the fact that in both experiments the sub-
jects were instructed to signal any perceived difference between
stimuli, their behavior indicated that they were neglecting
the stimulus level variations while concentrating on the qual-
ity or, as they usually expressed it, on the "sharpness" of the
stimulus. The experimenters, while monitoring the stimuli via
the loudspeaker system, detected the loudness variations
between successive stimuli quite easily.

Another hint about such signal level normalization can
be mentioned. Modeling the observer's performance in signal
detection tasks, D.M. Green and J.A. Swets (1966, p. 226)
had to modify their basic energy-detection model to make it
account for Weber's law. One of their plausible modifications

was the assumption that the magnitude of internal noise is proportional to the mean value of the stimulus. However, this is equivalent to the situation where the stimulus level is normalized and the level of internal noise is constant.

E. Eijkman, J.M. Thijssen, and A.J.H. Vendrik (1966) also conclude that their data, obtained in experiments on category judgement of the loudness of a 500 msec tone of frequency 1 kHz, can be interpreted under the assumption that the observers adjust the sensitivity of their limited-dynamical-range amplitude analyzer to match the range of stimulus amplitudes. As such behavior is analogous to voltage measurement using a multirange voltmeter set to the appropriate range of expected signal amplitudes, the authors call the model simulating such loudness discrimination the multirange meter model. But there is a distinction in application of this model: while in the just mentioned paper the ear's sensitivity is adjusted to match the intensity of the set of stimuli within an experiment run, our single stimuli seem each to be normalized separately.

Short-term adaptation is the most likely mechanism responsible for adjusting the sensitivity of the auditory analyzer to make the stimulus fall into the optimal differential sensitivity range of the receiver. These sensitivity shifts are executed with a delay, which can be characterized by a time constant of short-term adaptation. Its actual value can be estimated from the results of the following studies.

The first study, published in a series of papers by

W.D. Keidel and his colleagues, deals with different aspects of short-term adaptation. According to W.D. Keidel (1963), and M. Spreng and W.D. Keidel (1964), this continuous process of sensitivity adjustment seems to be a result of chemical and metabolic processes within the hair cells and adjoining neural pathways rather than a consequence of the mechanical properties of the tectorial membrane. The transient states of the cochlear microphonic and summated nerve action potentials, recorded in cats as the response to an amplitude jump in an auditory stimulus, was investigated in a series of experiments. W.D. Keidel and M. Spreng determined the time interval, in which the summated nerve action potential as collected from a gross electrode placed at the round window reaches 90 per cent of the new steady-state value, to be 3.24 ± 0.25 msec. This readaptation time was found to be strongly dependent upon the blood supply, concentration of oxygen in the inhaled air, and local body temperature. The degree of adaptation and its time course depends also upon prolonged exposure to high-level sound stimuli and upon treatment of the animals by drugs. G. Stange, M. Spreng, and U.O. Keidel (1964) investigated the influence of streptomycinsulfate, which affects the formation and decomposition of the excitatory substance within the sensory cells. The effect of the drug ephedrine, affecting the sympathetic nervous system was studied by G. Stange, and P. Beickert (1965).

J.J. Zwislocki (1969) is the author of another study which indicates the value of the short-term adaptation time

constant. Relevant to our discussion is that part of his study which describes the neural activity evoked by a rectangular stimulus, particularly its transient overshoot corresponding to the stimulus onset. According to the explanation offered by J.J. Zwislocki, the initial overshoot and subsequent decay of neural group activity at the onset part of the response stems from the refractory period of neurons and from the statistical character of neural firings. After a long enough interstimulus interval practically all the neurons are ready to respond. Supposing the majority of these fibers fired in the first volley, due to the neural refractory period only the remaining fibers are available for the second volley. Stimulation continuing, eventually steady-state firing rate is reached as an equilibrium between neural fibers which respond to stimulation and fibers which are in the process of recovery from the refractory period. This state should be reached within a few milliseconds, as in the peripheral auditory neurons the refractory period is short. The ratio between the onset peak of the firing rate and its steady-state level is estimated to be about 50. It depends on the stimulus rise time, intensity, and frequency. For the single unit this overshoot to steady-state ratio is between 2 and 3. The details of the fast decay of the firing rate are not known.

Data on microelectrode recordings from single cochlear nucleus units of Mongolian gerbils published by R.L. Smith and J.J. Zwislocki (1971), also show a decay of the firing rate with time after the stimulus onset. In this study the stimulus was a rectangular tone burst at the unit's

characteristic frequency. The decay of the firing rate was approximately exponential, the steady-state firing rate was reached within about 150 msec. The asymptotic firing rate was lower than the initial peak rate by a factor between 2 and 3.

Upon closer examination of this short-term adaptation with time constant about 3 msec, we can see that this process is in fact differentiation of the signal envelope, resulting in sharper definition of envelope changes. This envelope sharpening in the time domain is a direct anology to the lateral inhibition process, which increases the frequency resolution of hearing in the frequency domain.

7.6 Model Design

The basis for the design of our model for simulating the time organization of the time-frequency analyzing neural network is the structure of the model of A. Rozsypal, V. Majerník and V. Balko (1969), described in Section 7.3 and illustrated by Figures 7.3 and 7.4. Its contrast enhancing properties are given by Equation 7.2. However, this model neglects the time dependence of the lateral inhibition process.

In order to model the time evolution of the frequency selectivity sharpening process, the coefficient k in Equation 7.2 will be replaced by some time-dependent operator. Also, some differentiating units will be added to simulate the short-term adaptation.

7.6.a Block Diagram of the Model

Our model simulating the time organization of the

time-frequency analyzing neural network is given in Figure
7.5. In this block diagram only three parallel paths of this
model are illustrated. To model properly the neural inter-
action within the area of one critical band on the basilar
membrane, at least 49 parallel branches are required,
corresponding to the 24 critical bands of hearing. In that
case only the immediate spatially neighbouring branches
interact. Spacing of the "characteristic frequency" of the
parallel branches, can be equivalently expressed in terms of
the length of the basilar membrane, or in mel scale, Bark
scale, or in the scale of the just noticeable differences in
frequency. In our case of 49 branches this spacing is
0.65 mm, or 50 mel, or 0.5 Bark (0.5 of the critical band),
or 12.5 jnd in frequency.

The frequency-place transformation, carried out by the
distributed mechanical system of the basilar membrane, is
modeled by a simultaneous spectrum analyzer working on the
principle of multiple filtering, carried out by the bank of
bandpass filters M_i, tuned to the characteristic frequency
of each particular branch. These filters simulate the
resonance properties of the basilar membrane, as observed by
G. von Békésy (1960, Figure 11-49). The impulse response
corresponding to these frequency characteristics is given by
Equation 2.2. In our investigations we will neglect the
onset time of these broadly tuned filters. This onset time, of
the order of milliseconds, is negligible in comparison to
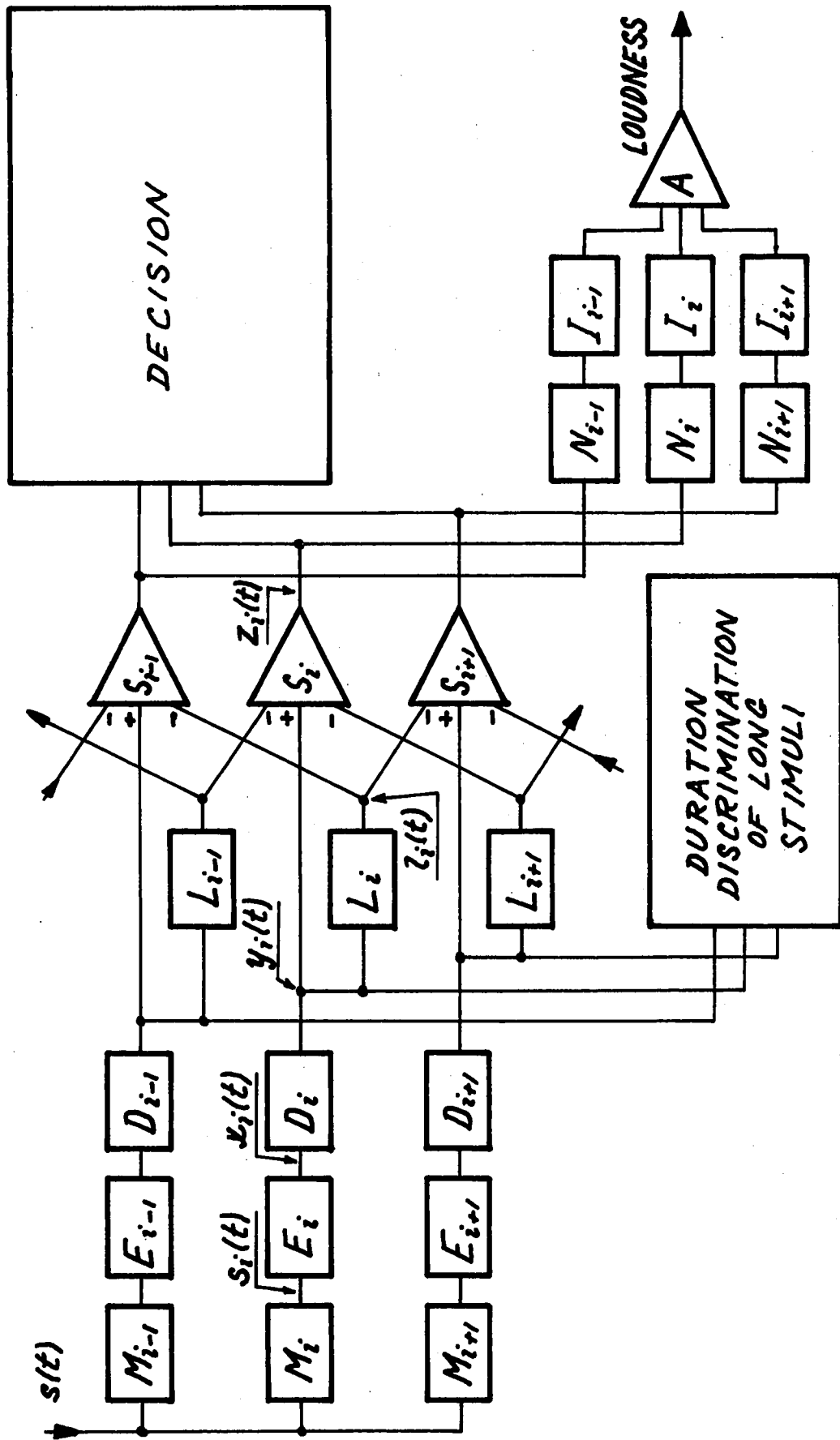the other two time constants which will be encountered in the

Figure 7.5 Model of time organization of the time-frequency auditory analyzer.

design of this model.

The input signal $s(t)$, representing the acoustic pressure waveform is input simultaneously into all filters M_i connected in parallel. The output signal $s_i(t)$ of these filters is demodulated in an envelope detector unit E_i. The signal $s_i(t)$ is a band-pass signal. On the other hand, the output signal $x_i(t)$ of the envelope detector is a low-pass signal. The information about the carrier frequency of the signal $s_i(t)$ lost in the process of envelope detection in the unit E_i is, however, encoded in the "place" subscript i of the parallel branch.

Signal $x_i(t)$ from the envelope detectors is differentiated in unit D_i. Parameters of this differentiator, simulating the short-term adaptation of hearing, will be determined later. Output $y_i(t)$ from the differentiator branches into the positive input of the summation amplifier S_i and into the integrating unit L_i, representing the time dependent element of the lateral inhibition. The same signal is also used for generation of the start and stop pulse for opening and closing of gate G, Figure 5.13, in the clock and counter mechanism for duration discrimination of long stimuli, as described in Section 5.7. Parameters of unit L_i will be determined later. Output from unit L_i is carried by the "lateral connections" into the negative inputs of the summation amplifiers S_{i-1} and S_{i+1} of the neighbouring branches.

Output $z_i(t)$ of the summation amplifier is brought into the auditory decision unit. The same signal is also input into

the cascade connection of the squarer unit N_i and the integrator I_i of the loudness evaluation section of the model.

Noise inherent in the neuron network is not investigated in this model. Its effect may be simulated by a threshold criterion which must be exceeded for detection to occur.

No attempt was made to model the processes taking place in the decision block.

7.6.b Integration Unit

It is intuitively expected that the L_i unit should have some integration properties in order to slow down the process of organization of critical bands.

Let us assume that the shift of the masked threshold ΔL_m, displayed in Figure 7.4, represents decay of excitation at some unspecified neural level corresponding to characteristic frequency 5 kHz. E. Zwicker (1965a)(1965b) obtained different masked threshold shifts for masked signals of different frequencies. In this first state of modeling, however, let us assume the same decay of excitation for all parallel branches of our model, namely, at the output of the summation amplifier S_i.

The decay in Figure 7.4 is a response to a white-noise unit step stimulus. The decay is a superposition of two exponentials with time constants $\tau_d = 24$ msec and $\tau_\ell = 188$ msec and of a steady state response. The decay component with the shorter time constant is in our model caused by the differentiation unit D_i. We will temporarily neglect it in this investigation of L_i. As the difference between the peak and steady state levels of the decay with the slower time constant

is about 6 dB, the decay at the output of the summation amplifier S_i as a response to the unit step of white-noise input signal is, assuming normalized signal,

$$z_i(t) = \frac{u(t)}{2}\left(1 + e^{-\frac{t}{\tau_\ell}}\right) \qquad (7.2)$$

where $u(t)$ is the unit step function. This signal is composed of the component from the direct path, from which is subtracted two components from the spatial neighbor paths

$$z_i(t) = x_i(t) - [\ell_{i-1}(t) + \ell_{i+1}(t)]$$

Because for white-noise stimulus the corresponding signals in all branches are identical, Equation 7.3 becomes

$$z_i(t) = x_i(t) - 2\ell_i(t) \qquad (7.3)$$

and we can drop the superscript i in the following calculations.

In the absence of the differentiating element D

$$x(t) = u(t)$$

and according to Equations 7.2 and 7.3

$$\ell(t) = \frac{u(t)}{4}\left(1 - e^{-\frac{t}{\tau_\ell}}\right) ,$$

which is the required step response of the integration unit L . The corresponding impulse response is

$$h_\ell(t) = \frac{d[\ell(t)]}{dt} = \frac{u(t)}{4\tau_\ell} e^{-\frac{t}{\tau_\ell}}$$

The system function $L(\omega)$ of the unit L is the Fourier transform of the impulse response $h_\ell(t)$

$$L(\omega) = \int_{-\infty}^{\infty} h_\ell(t) e^{-j\omega t} dt = \frac{1}{4}\frac{1}{1 + j\omega\tau_\ell} \qquad (7.4)$$

Integrator L can be realized by a simple leaky integrator circuit with time constant $\tau_\ell = 188$ msec.

The critical frequency of this circuit is

$$f_\ell = \frac{1}{2\pi\tau_\ell} = 0.85 \text{ Hz}.$$

The amplitude characteristic $|L(\omega)|$ is indicated in Figure 7.6.

7.6.c. Differentiator Unit

In order to simulate the sharp onset peak with the time constant $\tau_d = 24$ msec in Figure 7.4, a differentiating unit D_i will be incorporated into the direct path of our model, Figure 7.5.

For white-noise step input signal, again omitting the superscript i

$$x(t) = u(t)$$

the required output from D , in accordance with Figure 7.4, is

$$y(t) = u(t) \left(1 + e^{-\frac{t}{\tau_d}} \right) \tag{7.5}$$

From this step response the impulse response can be calculated as

$$h_d(t) = \frac{d[y(t)]}{dt} = 2\delta(t) - \frac{u(t)}{\tau_d} e^{-\frac{t}{\tau_d}},$$

where $\delta(t)$ is the impulse function. From the impulse response we obtain the system function $D(\omega)$ of the differentiator D by using the Fourier transform

$$D(\omega) \int_{-\infty}^{\infty} h_d(t) e^{-j\omega t} dt = \frac{1 + 2 j\omega\tau_d}{1 + j\omega\tau_d}. \tag{7.6}$$

As the time constant of the differentiator D is much shorter, $\tau_d = 24$ msec, than the integration time constant of the integrator L, $\tau_\ell = 188$ msec, the addition of unit D has negligible effect on the lateral inhibition mechanism.
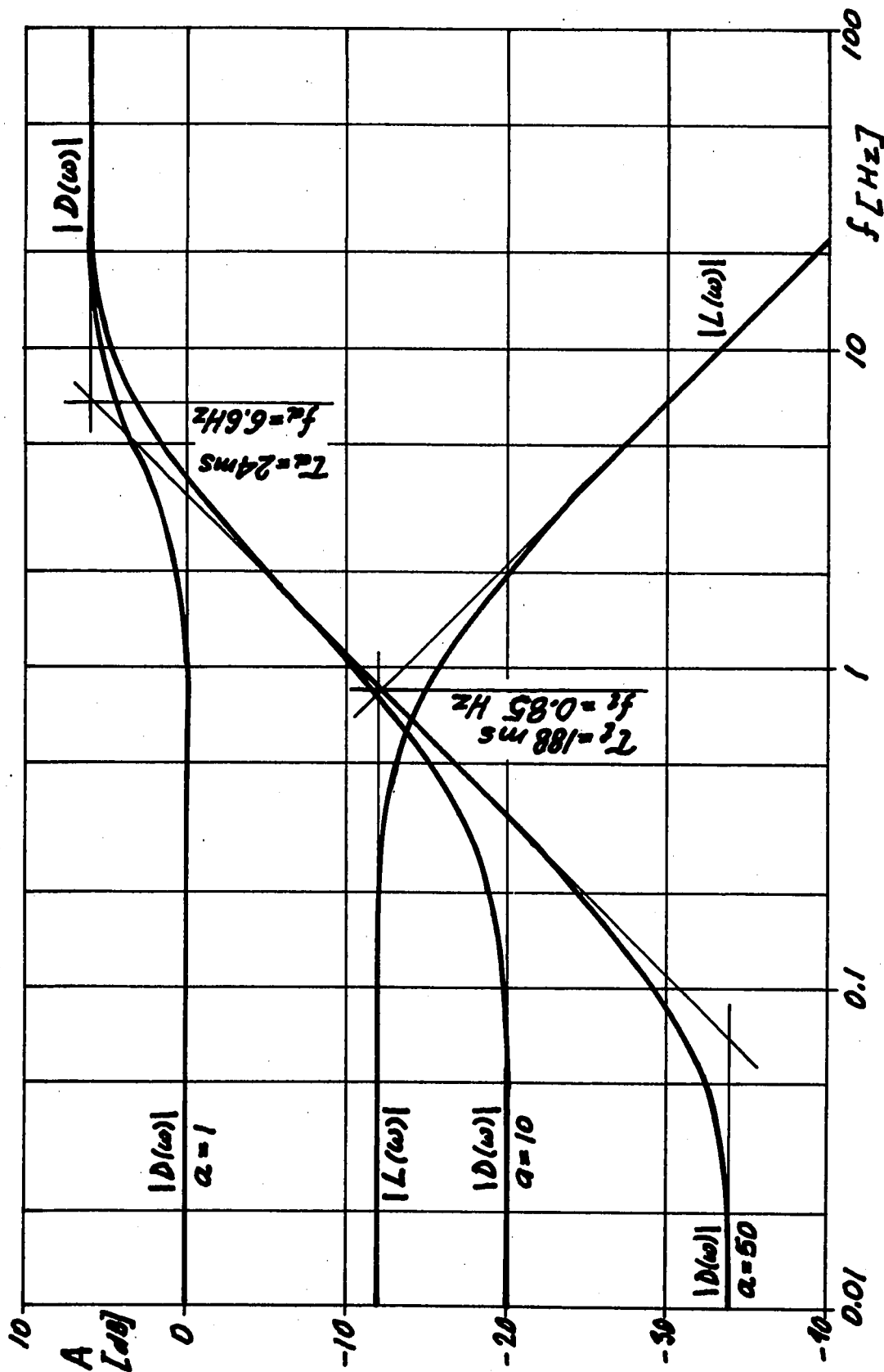
Figure 7.6 Amplitude characteristic |D(ω)| of the differentiator D for different a

from Equation 7.7 and amplitude characteristic |L(ω)| of the integrator L.

As already mentioned in Section 7.5 in connection with investigations of J.J. Zwislocki (1969), the details of the fast decay of the group firing are now known. Some phenomena observed in dynamical masking experiments may be best simulated, by the presented model, for some optimum proportion between the components of the signal y(t) in Equation 7.5. Introducing ratio coefficient a, then

$$y(t) = u(t) \left(1 + ae^{-\frac{t}{\tau_d}} \right) \qquad (7.7)$$

with the value of coefficient a between one and fifty.

The critical frequency of differentiator D, with time constant $\tau_d = 24$ msec, is

$$f_d = \frac{1}{2\pi\tau_d} = 6.6 \text{ Hz}$$

and its amplitude characteristic $|D(\omega)|$ is plotted in Figure 7.6. This unit shows differentiating properties only in a narrow frequency range. By choosing a value of the coecoefficient a in Equation 7.7 greater than one, the low frequecny leveling of the amplitude characteristic will be shifted in the direction of lower frequencies and the differentiation range will be wider, as can be seen for the case of a = 1, 10, and 50 in Figure 7.6. The characteristics for all a have the same attenuation at low frequencies as the characteristic for a = 1, and different boost of high frequencies.

7.6.d Loudness Evaluation

J.J. Zwislocki (1969) developed a quantitative theory based on psychophysical and neurophysiological evidence that loudness, as a function of stimulus intensity and duration, is a result of neural summation performed within the central

auditory nervous system. Hearing apparently exhibits temporal summation of acoustic energy in the process of loudness analysis. But this summation is preceded, at higher sound levels, by a nonlinear transformation between stimulus intensity and neural firing rate. The compression of the range of neural response parallels the loudness function, in which the stimulus intensity, according to S.S. Stevens (1955) is raised to the power of about 0.27. By an analysis of single-unit and group neural responses, J.J. Zwislocki demonstrated that the temporal decay of neural firing, mentioned in Section 7.5, if followed by a linear temporal summation, makes the loudness analyzer respond in the same way as if it integrated acoustic energy with some finite time constant.

Assuming that loudness is directly proportional to the integrated neural activity, the loudness estimation system visualized by J.J. Zwislocki consists of three stages:

First, a nonlinear transformation, simulating the loudness function, second, a temporal decay of neural firing as a response to stepwise stimulus, and third, a linear temporal integrator with a time constant 200 msec.

In our model the loudness estimation process is carried out in the same steps, only the order of the nonlinear transformation and of the decay of the neural activity is reversed. The neuronal network of integrators L and differentiators D simulates, at the outputs of the summation amplifiers S_i (signal $z_i(t)$), the decay of the neural activity. This signal

is input into the nonlinear element N_i modeling the loudness function. The nonlinear characteristic of the element N_i is approximated by a power function with an exponent 0.27. The output from the nonlinear element is integrated in a linear temporal integrator I_i with a time constant of 200 msec. Outputs from the bank of integrators are summed in unit A to give a signal corresponding to the loudness of the input signal s(t).

Output from unit A can be used as a relevant signal for detection of energy differences between stimuli, for example in the case of duration discrimination of short stimuli with wide-band carrier signal.

Chapter VIII

CONCLUSIONS

To sum up the duration discrimination experiment we may say that for a given observer and for a given stimulus duration, none of the remaining signal factors taken into consideration in our experiment influenced the results significantly: neither the envelope shape of the stimulus, nor the spectral bandwidth of the carrier signal, nor its central frequency affected, significantly, the duration discrimination.

Summarizing the envelope discrimination experiment, we may conclude that the envelope pairs ceased to be discriminable when the effective duration of the stimuli was shorter than about 2 to 3 msec. Above this critical duration the stimuli of different envelopes were perceived as different, even if their Fourier spectra were identical as in our case of the convergent and divergent triangle envelope pair. This suggests that even in the range of stimulus durations which can be regarded as very short, the time course of the envelope plays an important role in the envelope discrimination. All three signal factors considered in this experiment, namely the central frequency and the spectral bandwidth of the carrier signal, and the combination of envelopes, proved to be significant.

The variance of performance between subjects in envelope discrimination was insignificant, in distinction to the duration discrimination experiment in which the subjective factor S was significant in all three partial experiments. One possible explanation for this finding is that in the duration discrimination experiment two stimuli of different durations were compared. Under these circumstances the duration, loudness, and short-term spectrum clues were available for discrimination. According to our model, three different mechanisms process these clues. On the other hand, in the envelope discrimination experiment the duration of both compared stimuli was identical, as it was varied simultaneously. Consequently, the energy of both pulses was either the same, or at least of a constant ratio. Therefore the short-term spectrum analysis was the only mechanism contributing to discrimination. So, while in the envelope discrimination only one decision was made, in the duration discrimination the final decision was based on partial decisions from three separate mechanisms. In that case, even if the criteria for the partial decisions were the same from subject to subject, these partial decisions could be weighted in the final decision differently by different observers. This can account for the significant variability between observers in the duration discrimination experiment.

The functional model can serve as a starting point for computer modeling of the time-frequency analysis as well as of the loudness evaluation performed by the hearing. The model simulates the gradual evolution of the time-frequency

neural activity patterns. The enhancing of the contrast in the frequency domain manifests itself by gradual superimposition of the negative second spatial derivative of the excitation function on the original excitation function. This process reaches steady state in about 300 msec after onset of a stepwise stimulus. Contrast in the time domain is improved by differentiators in the direct paths of the model.

Besides these main functions the model can support also the mechanism for duration discrimination of long stimuli. Output signals from the differentiators can be formed into pulses which open and close the gate passing the clock pulses into the counter. Duration discrimination of short wide-band carrier signals, which offer no spectral contour clue, can be based on the output signal from the summator A, simulating the loudness of the perceived signal.

On the example of the model of the time evolution of the time-frequency analyser it was shown that one system can exhibit good resolution both in the time and frequency domains. Such a system has initially a flat frequency characteristic, which allows good resolution in the time domain. For long stimuli this system gradually adapts its frequency characteristic to match the spectral composition of the input signal, in order to attain good resolution in the frequency domain. This indicates that envelope detection and carrier signal analysis in hearing are not necessarily two separate processes, as one system can accomplish both functions.

REFERENCES

Bauch,H.(1956)."Die Bedeutung der Frequenzgruppe für die
    Lautheit von Klängen," Acustica 6, 40-45.

von Békésy,G.(1960). Experiments in Hearing (McGraw-Hill,
    New York).

von Békésy,G.(1967). Sensory Inhibition (Princeton Univ.
    Press, Princeton).

Black,H.S.(1953). Modulation Theory (Van Nostrand, New York).

Bos,C.E., and de Boer,E.(1966). "Masking and Discrimination'"
    J.Acoust.Soc.Amer. 39, 708-715.

Brady,P.T., House,A.S., and Stevens,K.N.(1961). "Perception
    of Sounds Characterized by a Rapidly Changing Resonant
    Frequency," J.Acoust.Soc.Amer. 33, 1357-1362.

Campbell,R.A., and Lasky,E.Z.(1967)."Masker Level and
    Sinusoidal-Signal Detection," J.Acoust.Soc.Amer. 42,
    972-976.

Cardozo,B.J.(1962)."Frequency Discrimination of the Human
    Ear," IVth Int.Congr.Acoust., Copenhagen.

Chistovich,L.A.(1957)."Perception of Time Interval between
    Two Short Sound Pulses," Soviet Physics: Acoustics,
    480-484.

Chistovich,L.A., and Ivanova,V.A.(1960)."Critical Duration
    for Loudness Estimation of Sounds (in Russian),"
    Fiziol.Zh.SSSR 46, 20-25.

Creelman,C.D.(1962)."Human Discrimination of Auditory
    Duration," J.Acoust.Soc.Amer. 34, 582-593.

Dubrovskij,N.A., and Tumarkina,L.N.(1965). "Thresholds for
    Perception of a Modulation of White Noise,"
    Vth Int.Congr.Acoust., Liège.

Eijkman,E, Thijssen,J.M., and Vendrik,A.J.H.(1966).
    "Weber's Law, Power Law, and Internal Noise,"
    J.Acoust.Soc.Amer. 40, 1164-1173.

Elliot,L.L.(1967)."Development of Auditory Narrow-Band
    Frequency Contours," J.Acoust.Soc.Amer. 42, 143-153.

Flanagan,J.L.(1965). Speech Analysis, Synthesis, and Perception (Springer Verlag, New York).

Gabor,D.(1946). "Theory of Communication," Jour.IEE, Part III, 93, 429.

Gabor,D.(1947). "Acoustical Quanta and the Theory of Hearing," Nature 59, No. 4044, 591.

Garner,W.R., and Miller,G.A.(1947). "The Masked Threshold of Pure Tones as a Function of Duration," J. Exp. Psychol. 37, 293-303.

Gässler,G.(1954). "Über die Hörschwelle für Schallereignisse mit verschieden breitem Frequenzspektrum," Acustica 4, 408-414.

Gescheider,G.A.(1966). "Resolving of Successive Clicks by the Ears and Skin," Jour. Exp. Psychol. 71, 378-381.

Gescheider,G.A.(1967). "Auditory and Cutaneous Temporal Resolution of Successive Brief Stimuli," Jour. Exp. Pscychol. 75, 570-572.

Goluzina,A.G., Lyublinskaya,V.V., Rohtla,M.K., and Chistovich, L.A.(1966). "Subjective Interstimulus Distance Scaling as a Method of Investigation of the Auditory Perception (in Russian)," Probl. Fiziol. Akust. SSSR 5, 96-110.

Green,D.M., Tanner,W.P.Jr., and Birdsall,T.G.(1957). "Signal Detection as a Function of Signal Intensity and Duration," J. Acoust. Soc. Amer. 29, 523-531.

Green,D.M. and Sewell,S.T.(1962). "Effect of Background Noise on Auditory Detection of Noise Bursts," J. Acoust. Soc. Amer. 34, 1207-1216.

Green,D.M., and Swets,J.A.(1966). Signal Detection Theory and Psychophysics (J. Wiley, New York).

Green,D.M.(1971). "Temporal Auditory Acuity," Psychological Review 78, 540-551.

Greenwood,D.D.(1961a). "Auditory Masking and the Critical Band," J. Acoust. Soc. Amer. 33, 484-502.

Greenwood,D.D.(1961b). "Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane," J. Acoust. Soc. Amer. 33, 1344-1356.

Hamilton,P.M.(1957). "Noise Masked Thresholds as a Function of Tonal Duration and Masking Noise Bandwidth," J. Acoust. Soc. Amer. 29, 506-511.

Heisenberg,W.(1927). "Uber den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," Zeit. Physik 43, 43.

Hojan,E., and Rozsypal,A.(1967). "Recognition of an Undulated Envelope Transient as a Function of Carrier Frequency (in Polish)," Archiwum Akustyki 2, 267-276.

Huffman,D.A.(1962). "The Generation of Impulse-Equivalent Pulse Trains," IRE Trans. IT 8, S10-S16.

Keidel,W.D.(1963). "Tuning between Central Auditory Pathways and the Ear," IEEE Trans. Milit. Electr., 131-144.

Kharkevich,A.A.(1960). Spectra ans Analysis (transl. from Russian) (Consultants Bureau Enterprises, New York).

Korn,T.S.(1969/70). "Theory of Audio Information," Acustica 22, 336-344.

Leshowitz,B.(1971). "Measurement of the Two-Click Threshold," J. Acoust. Soc. Amer. 49, 462-466.

Liang-Chian, and Chistovich,L.A.(1961). "Frequency Difference Limens as a Function of the Tonal Duration," Soviet Physics: Acoustics 6, 75-80.

Lyublinskaya,V.V., and Goluzina,A.G.(1966). "Effect of the Duration of Tone Signals on the Magnitude of Subjective Interstimulus Distance (in Russian)," Probl. Fiziol. Akust. SSSR 5, 111-117.

Maiwald,D.(1966). Zusammenhang zwischen Mithörschwellen und Modulationsschwellen (doctoral dissertation, Technical University Stuttgart, W. Germany).

Maiwald,D.(1967a). "Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz- und Amplituden-änderungen," Acustica 18, 81-92.

Maiwald,D.(1967b). "Die Berechnung der Modulationsschwellen mit Hilfe eines Funktionsschemas," Acustica 18, 193-207.

Miller,G.A.(1947). "Sensitivity to Changes in the Intensity of White Noise and its Relation to Masking and Loudness," J. Acoust. Soc. Amer. 19, 609-619.

Nábělek,I.(1965). "Discriminability of the Rise Time for Noise and Tone Pulses," Vth Int. Congr. Acoust., Liège.

Nåbělek,I., Rozsypal,A., and Balko,V.(1965). <u>The Loudness of Tone Pulses as Influenced by the Rate of their Exponential Onset and Decay</u> (in Slovak) (research report, Inst. of Physics SAS, Bratislava, Czechoslovakia).

Nåbělek,I., and Hirsh,I.J.(1969). "On the Discrimination of Frequency Transitions," J. Acoust. Soc. Amer. <u>45</u>, 1510-1519.

Nåbělek,I.V., Nåbělek,A.K., and Hirsh,I.J.(1970). "Pitch of Tone Bursts of Changing Frequency," J. Acoust. Soc. Amer. <u>48</u>, 536-553.

Niese,H.(1960). "Subjektive Messung der Lautstärke von Bandpassrauschen," Hochfreq. und Elektroakustik <u>68</u>, 202-217.

Niese,H.(1961). "Die Lautstärkebildung bei binauralem Hören komplexer Geräusche," Hochfreq. und Elektroakustik <u>70</u>, 132-141.

Oetinger,R.(1959). "Die Grenzen der Hörbarkeit von Frequenz- und Tonzahländerungen dei Tonimpulsen," Acoustica <u>9</u>, 430-434.

Patterson,J.H., and Green,D.M.(1970). "Discrimination of Transient Signals Having Identical Energy Spectra," J. Acoust. Soc. Amer. <u>48</u>, 894-905.

Pollack,I.(1952). "Comfortable Listening Levels for Pure Tones in Quiet and Noise," J. Acoust. Soc. Amer. <u>24</u>. 158-162.

Pollack,I.(1958). "Loudness of Periodically Interrupted White Noise," J. Acoust. Soc. Amer. <u>30</u>, 181-185.

Raab,D.H., and Taub,H.B.91969). "Click-Intensity Discrimination with and without a Background Masking Noise," J. Acoust. Soc. Amer. <u>46</u>, 965-968.

Rakowski,A., and Rozsypal,A.(1968). "Pitch of Noise Filtered by Low-Pass Filter with Steep Roll-Off (in Polish)," Archiwum Akustyki <u>1</u>, 3-10.

Reichardt,W., and MacGinitie,G.(1962). "Zur Theorie der lateralen Inhibition," Kybernetik <u>1</u>, 155.

Reichardt,W., and Niese,H.(1970). "Choice of Sound Duration and Silent Intervals for Test and Comparison Signals in the Subjective Measurement of Loudness Level," J. Acoust. Soc. Amer. <u>47</u>, 1083-1090.

Rochester,S.R.(1970). "Direct Measure of Duration Discrimination," J. Acoust. Soc. Amer. 47, 118(A).

Rohtla,M.K., and Rozsypal,A.(1966). "Subjective Interstimulus Distance between Narrow-Band Noise Signals in Dependence on Duration and on Central Frequency Increment," Probl. Fiziol. Akust. SSSR 5, 118-127.

Ronken,D.A.(1970). "Monaural Detection of a Phase Difference between Clicks," J. Acoust. Soc. Amer. 47, 1091-1099.

Ronken,D.A.(1971). "Some Effects of Bandwidth-Duration Constraints on Frequency Discrimination," J. Acoust. Soc. Amer. 49, 1232-1242.

Rozsypal,A.(1966). Generator of Functions of Arbitrary Shape (in Slovak) (technical report, Inst. of Physics SAS, Bratislava, Czechoslovakia).

Rozsypal,A.Majerník,V., and Balko,V.(1969). "Some Nonlinear Functional Models of Frequency Discrimination of the Human Ear," Fyzilálny Časopis 19, 85-92.

Schafer,T.H.,Gales,R.S.,Schewmaker,C.A., and Thompson,P.O. (1950). "Frequency Selectivity of the Ear as Determined by Masking Experiments," J. Acoust. Soc. Amer. 22, 490-497.

Scharf,B.(1959a). "Critical Bands and the Loudness of Complex Sounds near Threshold," J. Acoust. Soc. Amer. 31, 365-370.

Scharf,B.(1959b). "Loudness of Complex Sounds as a Function of the Number of Components," J. Acoust. Soc. Amer. 31, 783-785.

Scharf,B.(1961). "Loudness Summation under Masking," J. Acoust. Soc. Amer. 33, 503-511.

Scharf,B.(1962). "Loudness Summation and Spectrum Shape," J. Acoust. Soc. Amer. 34, 228-233.

Schief,R.(1963). "Koinzidenz-Filter als Modell für das Menschliche Tonhöhenunterscheidungsvermögen," Kybernetik 2, 8-15.

Scholl,H,(1962a). "Uber die Bildung der Hörschwellen und Mithörschwellen von Impulsen," Acustica 12, 91-101.

Scholl,H.(1962b). "Das dynamische Verhalten das Gehörs bei der Unterteilung das Schallspectrums in Frequenzgruppen," Acustica 12, 101-107.

Scholl,H.(1962c). "Verdeckung von Tönen durch pulsierende Geräusche," IVth Int. Congr. Acoust., Copenhagen.

Smith,R.L., and Zwislocki,J.J.(1971). "Responses of Some Neurons of the Cochlear Nucleus to Tone-Intensity Increments," J. Acoust. Soc. Amer. 50, 1520-1525.

Spreng,M., and Keidel,W.D.(1964). "Adaptation: Nicht-mechanische Eigenschaft des Innenohres," Pflügers Arch. ges. Physiol. 279, R7.

Stange,G., Spreng,M. and Keidel,U.O.(1964). "Die Wirkung von Streptomycinsulfat auf Erregung und Adaptation der Haarzellen des Cortischen Organs," Pflügers Arch. ges. Physiol. 279, 99-120.

Stange,G., and Beickert,P.(1965). "Adaptationsverhalten des Cortischen Organs nach Ephedrin und Beschallung," Arch. Ohr.-Nas.-Kehlkopfheilk. 184, 483-495.

Stevens,S.S.(1955). "The Measuremenr of Loudness," J. Acoust. Soc. Amer. 27, 815-829.

Stewart,G.W.(1931). "Problems Suggested by an Uncertainty Principle in Acoustics," J. Acoust. Soc. Amer. 2, 325-329.

Sunstein,D.E.(1949). "Photoelectric Waveform Generator 'Photoformer'," Electronics 22, 100-103.

Thomas,I.B., Dinicola,P.D., Ella,M.R., and Pasierbiak,C.S. (1970). "Auditory Memory in a Pitch-Discrimination Task," J. Acoust. Soc. Amer. 48, 1383-1386.

Trees, van H.L. (1968). Detection, Estimation, and Modulation Theory (J. Wiley, New York).

Türk,W.(1940). "Uber die physiologisch-akustischen Kennzeiten von Ausgleichsvorgängen," Akust. Zeitschr. 5, 129-145.

Vaytulevich,S.F., Kozhevnikov,B.A., Lebedev,A.P., and Soroko, V.I.(1962). "Electronic Gate with Low Spurious Signal Level (in Russian)," Pribory i Tekhnika Exp. 5, 185-186.

Watson,C.S., and Gengel,R.W.(1969). "Signal Duration and Signal Frequency in Relation to Auditory Sensitivity," J. Acoust. Soc. Amer. 46, 989-997.

Weber,C.L.(1968). Elements of Detection and Signal Design (McGraw-Hill, New York).

Wegel,R.L., and Lane,E.E.(1924). "The Auditory Masking of One Sound by Another and its Probable Relation to the Dynamics of the Inner Ear," Phys. Rev. 23, 266.

Winer,B.J.(1971). Statistical Principles in Experimental Design (McGraw-Hill, New York).

Zwicker,E.(1952). "Die Grenzen der Hörbarkeit der Amplituden-modulation und der Frequenzmodulation eines Tones," Acustica 2, 125-133.

Zwicker,E.(1954). "Die Verdeckung von Schmalbandgeräuschen durch Sinustöne," Aucustica 4, 415-420.

Zwicker,E., and Feldtkeller,R.(1955). "Über die Lautstärke von Gleichförmigen Geräuschen," Acustica 5, 303-316.

Zwicker,E., Flottorp,G., and Stevens,S.S.(1957). "Critical Bandwidth in Loudness Summation," J. Acoust. Soc. Amer. 29, 548-557.

Zwicker,E., and Maiwald,D.(1963). "Über das Freifeldübertragungsmass des Kopfhörers DT 48," Acustica 13, 181-182.

Zwicker,E.(1965a). "Temporal Effects in Simultaneous Masking by White-Noise Bursts," J. Acoust. Soc. Amer. 37, 635-663.

Zwicker,E.(1965b). "Temporal Effects in Simultaneous Masking and Loudness," J. Acoust. Soc. Amer. 38, 132-141.

Zwicker,E., and Feldtkeller,R.(1967). Das Ohr als Nachrichten-empfänger (S. Hirzel Verlag, Stuttgart, W. Germany).

Zwicker,E.(1969/70). "Subjektive und objektive Dauer von Schallimpulsen und Schallpausen," Acustica 22, 214-218.

Zwislocki,J.J.(1960). "Theory of Temporal Auditory Summation," J. Acoust. Soc. Amer. 32, 1046-1060.

Zwislocki,J.J.(1966). "Analysis of Some Auditory Characteristics," in Handbook of Mathematical Psychology, vol. III, R.D. Luce, R. R. Bush, and E. Galanter, Ed. (J. Wiley, New York).

Zwislocki,J.J.(1969). "Temporal Summation of Loudness: An Analysis," J. Acoust. Soc. Amer. 46, 431-441.