Open Research Online

The Open University's repository of research publications
and other research outputs

Optophone design: optical-to-auditory vision substitution for the blind

Thesis

How to cite:

O'Hea, Adrian Ralph (1994). Optophone design: optical-to-auditory vision substitution for the blind. PhD thesis The Open University.

For guidance on citations see FAQs.

© [not recorded]

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data policy on reuse of materials please consult the policies page.

oro.open.ac.uk

DX179937
UNRESTRICTED

OPTOPHONE DESIGN:

OPTICAL-TO-AUDITORY VISION SUBSTITUTION FOR THE BLIND

by

Adrian Ralph O'Hea

BSc (hon) civil engineering

MSc engineering hydrology

MSc computing systems

submitted on 11 April 1994 for the award of the degree of

Doctor of Philosophy

in

Electronics

Author number: U1702523
Date of submission: 18 April 1994
Date of award: 17 May 1994

OPTOPHONE DESIGN:

OPTICAL-TO-AUDITORY VISION SUBSTITUTION FOR THE BLIND

EXTRACT

An optophone is a device that turns light into sound for the benefit of blind people. The present project is intended to produce a general-purpose optophone to be worn on the head about the house and in the street, to give the wearer a detailed description in sound of the scene he is facing. The device will therefore consist of an electronic camera, some signal-processing electronics, earphones, and a battery. The two major problems are the derivation of (a) the most suitable mapping from images to sounds, and (b) an algorithm to perform the mapping in real time on existing electronic components. This thesis concerns problem (a). Chapter 2 goes into the general scene-to-sound mapping problem in some detail and presents the work of earlier investigators. Chapter 3 discusses the design of tests to evaluate the performance of candidate mappings. A theoretical performance test (TPT) is derived. Chapter 4 applies the TPT to the most obvious mapping, the cartesian piano transform. Chapter 5 applies the TPT to a mapping based on the cosine transform. Chapter 6 attempts to derive a mapping by principal component analysis, using the inaccuracies of human sight and hearing and the statistical properties of real scenes and sounds. Chapter 7 presents a complete scheme, implemented in software, for

representing digitised colour scenes by audible digitised stereo sound. Chapter 8 tries to decide how many numbers are required to specify a steady spectrum with no noticeable degradation. Chapter 9 looks at a scheme designed to produce more natural-sounding sounds related to more meaningful portions of the scene. This scheme maps windows in the scene to steady spectral patterns of short duration, the location of the window being conveyed by simulated free-field listening. Chapter 10 gives detailed recommendations as to further work.

OPEN
UNIVERSITY
28 JUN 1994
LIBRARY
DONATION

ABBREVIATIONS

CIE	Commission internationale de l'Éclairage
CPR	candidate psychophysical representation
dBDL	decibel difference limen
DC	direct current
DL	difference limen
ERB	equivalent rectangular bandwidth
erb	not an abbreviation but a word (Figure 3.2)
ERD	equivalent rectangular duration
erb	not an abbreviation but a word (Section 7.2.1)
FDL	frequency difference limen
FFDL	formant frequency difference limen
FFT	fast Fourier transform
FT	Fourier transform
GDL	frequency difference limen in erbs
GPO	general problem of optophonics
JND	just noticeable difference
KL	Karhunen-Loève
PA	power attenuation
PIA	property of inconsequential ambiguity
PR	psychophysical representation
PRL	power ratio limen
TPT	theoretical performance test

ACKNOWLEDGEMENTS

I most gratefully thank my wife Miriam for her patience and support during the seven years that this work has taken up my spare time, and in particular for her encouragement since my illness began. Without her belief in the optophone I would not have been able to complete this thesis.

My daughter Shanti and supervisor Phil Picton also rallied round selflessly.

My heartfelt thanks go to my mother Rowena who bought me the computer on which the work was done.

TABLE OF CONTENTS

FIGURES

CHAPTER 3 TESTING OF MAPPINGS

FIGURES

FIGURES

CHAPTER 6 PRINCIPAL COMPONENT ANALYSIS

FIGURES

CHAPTER 8 NUMBER OF NUMBERS REQUIRED TO SPECIFY A SPECTRUM

FIGURES

CHAPTER 9 SCHEME 4 - FREE-FIELD PATCH TRANSFORM

FIGURES

CHAPTER 10 CONCLUSIONS AND RECOMMENDATIONS

REFERENCES

CHAPTER 1 INTRODUCTION

1.1 General

This thesis reports on research done in the wider context of a project. In order to set the scene, this introduction first describes the project as a whole, then specifies which part of the project is the subject of the present research and thus of the thesis, and ends by introducing the research.

1.2 The project

An optophone is a device that turns light into sound for the benefit of blind people. The present project is intended to produce a general-purpose optophone to be worn on the head about the house and in the street, to give the wearer a detailed description in sound of the scene he is facing. The device will therefore consist of some kind of electronic camera, some signal-processing electronics, earphones, and a battery.

1.3 The purpose

Although the word optophone appeared in pre-war

dictionaries, the only successful optophones available are devices to read aloud printed text. The requirement for a device that would describe intelligibly whatever it was pointed at, including text such as signs in the street, is self-evident, since that is exactly what sight does and what is missing from the blind person.

1.4 The beneficiaries

The blind form about 0.15% of the population, of which 25% are totally blind or have only light perception (Trouern-Trend & Bering Jr 1969). If we balance out on the one hand those totally blind but unable for whatever reason to use an optophone with on the other those with some sight that would sometimes like to see more clearly, we have 0.04% of the population as potential customers.

Unfortunately most blind people are poor. Although many may be prepared to pay dearly for a good optophone, few would be able to. The social services of some countries may contribute in varying degrees. It is possible that the signal processing requirements would make the electronics too expensive, but the price of similarly complex items such as video cameras is encouraging.

If we assume that three quarters of the world lives in countries too poor to afford optophones for their blind

population, then the number of potential customers comes to at most 2 million.

1.5 Some potential difficulties

The ability of subjects to learn the mapping of scenes to sounds implicit in their optophone will be one of the big unknowns of the whole project. Initially, of course, the sounds from the optophone will be completely meaningless. The bulk of the learning will be done in private, with the user finding out for himself what sounds are made by familiar objects in the home. In addition, there might be scope for some more formal training, where the sound of unreachable objects such as buildings would be explored by means of hands-on models. Third, with a particularly puzzling sound, there would sometimes be the opportunity to ask a friend "What's that over there?" There are plenty of examples of the human ability to learn, in time, to recognise effortlessly the meaning of completely arbitrary signals, such as learning a language or learning to read.

There is also the danger of an optophone acquiring a bad name by new users giving up through lack of support. However good the product, it is not anticipated that many users would be able to learn to use one just from braille instructions without some further encouragement. I have

several times bought a language course without subsequently learning the language. Most blind people know no braille anyway.

There is a chance that users might find any optophone of the type described too uncomfortable in some way. Given on the one hand that many nowadays like to wear personal radios in the street, and on the other that the optophone could be instantly turned off to enable proper hearing, this is considered unlikely.

1.6 Earlier work

There has never been an optophone of the type proposed here, although there have been applications for patents.

One reason has been the attention paid by researchers to echolocation, where the user detects his surroundings by listening to the echos, suitably transformed into sound by the electronics, from ultrasonic noises the device sends out. The reasoning behind this research preference is that echolocation is used by animals that can't see. The mistake is not to have realised why the animals can't see. They can't see because they go about in the dark (bats) or in murky water (dolphins), not because they don't have eyes. Every animal that goes about in daylight prefers to have eyes. The advantages of an

optophone over an echolocator are that the echolocator cannot see anything far, or through glass, or on paper. One hears sporadically of new echolocators (see references under **HARDWARE - BLIND AIDS - ECHOLOCATION**), but I have never seen a blind person wearing one.

Philips of the Netherlands have recently applied for an optophone patent (Meijer 1992). Both the general idea of a high-resolution optophone and the particular scene-to-sound mapping claimed were described in my 1987 MSc dissertation (O'Hea 1987), which concluded that the mapping was not very good.

To my knowledge the only other recent work on optophonics was Carver Mead's (Mead 1989, 207-227).

Both Mead and Philips appear to have been too keen to get into the implementation (hardware design) without thinking enough about what mapping they wanted to implement. This amounts to solving problem b before problem a (Section 1.7 below).

Because of the failure of electronic mobility devices in general, a successful optophone would only displace guide dogs and white canes.

1.7 The technical approach

In hardware terms the technical approach is easily stated and, apart perhaps from some specialised chips, already solved and available off the shelf: electronic cameras, signal-processing electronics, earphones and batteries. These are becoming smaller, cheaper, lighter and better all the time, and just crying out for someone to fit them together into this type of application.

What is not so self-evident is how to design the software for such an optophone, or whether one is possible even in theory (the present research concludes that it is). The two major problems to be overcome are therefore

a the derivation in mathematical terms of the mapping from images to sounds most suitable to the needs of a blind person

b the derivation of an algorithm to perform the mapping, or a good enough approximation to it, in real time on existing electronic components.

<u>The first of these two problems is the subject of the research covered in this thesis.</u>

The solution to problem b, the real-time mapping algorithm, really has to wait until the theoretical

mapping, the answer to problem a, is known. However, some attention must be paid to speed even in developing the theoretical mapping, since otherwise it will be impossible to test on an ordinary computer.

1.8 The present research

The present research addresses problem a above, and therefore amounts to an attempt to answer the question "What do you want things to sound like?" Such things must include not only all the things the optophone designer can think of but also anything else the blind person might point the optophone at, including objects not yet invented. The approach in this project to solving this problem is based on the following four requirements, all based on common sense.

1 Previous work. It is of course necessary to take into account all that is known about the psychophysics of seeing and hearing. The research already undertaken in this respect can be judged by the list of over 600 references compiled since the project started.

2 Continuity. Small changes in the scene should result in small changes in the sound. No two scenes are ever identical, but one would not want a new

scene to sound completely different from a similar scene that was already familiar.

3 Completeness. No scene should be unmappable nor any sound unused. On the one hand, every scene should be mappable since there is no knowing what the optophone will be pointed at. On the other hand, if some sounds were unused, then this would squeeze all possible scenes on to a smaller range of sounds, resulting in more loss of detail than necessary.

4 Subjectivity. Undetectable differences in scenes or sounds don't count, even if detectable by the hardware. That is to say, proper account must be taken of the resolution of human sight and hearing.

1.9 Results

Having only the above four self-imposed requirements to go on, the work proceeded according to no fixed plan. Many different schemes (that is, scene-to-sound mappings) were investigated before being dropped. Sometimes they were abandoned for some inherent defect, sometimes because a more promising scheme came to mind. Not all of the schemes dropped were blind alleys; some were later revived in some modified form before being again abandoned or left open.

Chapter 2 goes into the general scene-to-sound mapping problem in some detail and presents the work of earlier investigators.

Chapter 3 discusses the design of tests to evaluate the performance of candidate mappings. Most favoured are tests of mappings in functioning optophones, requiring users to perform some well defined task, such as reading, against the clock. The need for sufficient training is stressed.

Testing at an earlier stage of the development of a new mapping, without the need for a fully functioning optophone, or even of a user, is possible by the following sequence of calculations:

1 Obtain a digital still scene using a television camera.

2 Calculate the corresponding sound using the mapping under trial.

3 Calculate an almost perceptibly different sound using the known inaccuracies of human hearing.

4 Recalculate the digital scene using a suitable inverse version of the mapping.

5 Criticise the recalculated scene visually, by comparison with the original or otherwise.

6 See if there emerge any clues to a better mapping.

This is called the theoretical performance test (TPT). Note that it is only possible if the mapping has an inverse.

Chapter 4 applies the TPT to the most obvious mapping, the cartesian piano transform. A more elaborate version of the piano transform is taken up again in Chapter 7.

Chapter 5 presents a scheme based on the cosine transform, and attempts to evaluate the scheme by the theoretical performance test of Chapter 3. Due to eagerness to press on with other ideas, audible sounds for this scheme were never produced and a proper subjective assessment was therefore not possible. Neither, having regard to the qualities of the scheme, is one recommended.

Chapter 6 considers how it might be possible to derive the most appropriate scheme from first principles, using on the one hand knowledge of the inaccuracies of human sight and hearing, and on the other knowledge of the statistical properties of real scenes and sounds.

Difficulties with this approach led to its abandonment until its reuse in the scheme of Chapter 9.

Chapter 7 presents a complete scheme, implemented in software, for representing digitised colour scenes by audible digitised stereo sound. Luckily, despite being based on the piano transform, the mapping in this scheme is not invertible, so application of the theoretical performance test described in Chapter 3 was not possible. This forced attention onto the design of proper subjective tests, also discussed in Chapter 3.

Chapter 8 tries to decide how many numbers are required to specify a steady spectrum for human consumption, with no noticeable degradation. This information is required in Chapter 9.

Chapter 9 looks at a scheme designed to produce more natural-sounding sounds related to more meaningful portions of the scene. This scheme maps the contents of a window in the scene to a steady spectral pattern of short duration, the location of the window being conveyed by simulated free-field presentation of the sound.

Conclusions and recommendations form Chapter 10.

1.10 Style

Optophonics touches on many different specialised fields of study. I have tried to treat each one in elementary fashion, which is the level I started at in all of them. Hence the chatty style. Those familiar with a subject will inevitably find the corresponding sections laboured. The style was not chosen for them.

The continental we is used quite a bit, a result of me having grown up in France. It refers in a general way to me and the reader facing a problem together, and is a convenient way of avoiding constant use of the passive.

I try in general to use words as in ordinary English, and to resist where not helpful the theft and devaluation of words practised by some professionals. For instance, brightness and loudness have their ordinary meaning and don't generally refer to any particular scale of measurement.

1.11 References

All references consulted during the course of this work are listed, though not all referred to directly in the text. In order to form a more useful guide to further study, the references are ordered by subject, and within

subject by date. However, many of the subjects included have only a few references, and the list is not at all comprehensive in this respect.

1.12 Figures

The figures are all closely connected to the text. In general, the text will not be clear without looking at the current figure. In addition, figures are if possible annotated, the intention being to make each as self-explanatory as possible. Where this has not been possible, a compensatory level of explanation will be found in the text. Unannotated figures have the same orientation as the rest: the top of the figure is the left of the paper.

CHAPTER 2 GENERAL DISCUSSION AND EARLIER WORK

2.1 Comparison of scenes and sounds, sight and hearing

In looking for a mapping from scenes to sounds, it is natural to look at what attributes describe scenes and what attributes describe sounds, and to try to match the attributes two by two, one scene attribute against one sound attribute.

People have looked for analogies between seeing and hearing for a variety of different reasons, ranging from the most primitive to the most complex. For instance, one can ask on the one hand whether brightness is more analogous to loudness or to pitch, and on the other hand whether there is any connection between the evolution in music from Brahms to Schoenberg and the evolution in painting from Renoir to Picasso.

Comparison of symphonies and paintings is instructive. The salient difference is that, even though both take time to take in, in a symphony order is everything, in a painting nothing. The different bits of a painting can be looked at in any order, and are in fact never looked at in the same order twice. Some studies (see under PSYCHOPHYSICS - SIGHT - EYE MOVEMENTS) show general tendencies in the order in which people look at things,

but no-one would claim that the painting is changed by the order.

The references headed PSYCHOPHYSICS - CROSSMODAL STUDIES make fascinating reading, but no consensus emerges that might be useful in optophonics. In general the authors are obliged either to discuss the subject discursively and anecdotally, or, if performing experiments, to limit their scope to comparing just one or two of the variables in each domain.

Handel (1988) concludes that no one analogy is sufficient, the best depending on context. However; a mapping depending on context, in addition to requiring artifical intelligence to implement, would violate our continuity requirement.

The problem with trying to match such variables as brightness and pitch two by two is that there are more perceived dimensions in scenes than in sounds. Even if we leave out the third spatial dimension (distance from the viewer) as being generated in the viewer, and colour as being of minor importance, we are still left with a scene described by a brightness function of (x, y, time) and a sound described by a loudness function of (pitch, time).

One way forward could well be to sample the scene in

black and white at say one-second intervals, and then map each frozen scene to a one-second sound sequence. Evolution of the scene would then be derived by the user detecting differences between successive sound sequences. Much of the work reported here is based on this scheme.

For a more detailed discussion of different possible mappings, see O'Hea (1987).

2.2 Some past mappings

A good overview of electronic mobility aids for the blind was provided by Kay (1984). The other main work to recommend to the newcomer is Warren & Strelow (1984). There is very little in either about optophonics, most workers having been attracted either to other inputs such as echolocation or to other outputs such as vibrotactile displays.

The eight known attempts at optophone mappings are by

Fish (1976)

Dallas (1980)

Kurcz (1981)

Deering (1984)

Tou & Adjouadi (1984)

O'Hea (1987)

Nielsen, Mahowald & Mead (1989)

and Meijer (1992).

2.2.1 Fish (1976)

Fish (1976) mapped vertical position to tone frequency and horizontal position to binaural loudness difference in several systems where the sound at any instant depended on the brightness gradient at one point of the scene only, the scene being scanned by the point in raster fashion. The mapping, together with the heliotrope of Kurcz (below), is thus an example of a point mapping.

The scanning rate was variable, being faster when no edges were being crossed. In this way more time was spent on interesting parts of the scene than on plain areas.

Subjects were able to identify 18 test patterns with at most four hours training. They could also describe new patterns not in the training set, indicating that they had understood the mapping. Minimum presentation time was from 0.8 to 8 seconds depending on the complexity of the pattern.

2.2.2 Dallas (1980)

Dallas (1980), in a patent application, mapped vertical position in the two-dimensional visual field to sound frequency, horizontal position to time, and brightness to loudness. Thus a horizontal white line would sound like a continuous tone, the higher the line the higher the tone, and a vertical white line would sound like a click or thud, the further to the left the earlier the click. Permutations and reversals of this mapping are also covered in the patent application.

Dallas's mapping is an example of the piano transform, rediscovered independently by O'Hea (1987) and Meijer (1992) and so named because one can imagine the scene being scanned from left to right by a vertically oriented piano keyboard having the high notes at the top of the picture.

The piano transform is an example of a slot mapping. This simply means that the scene may be considered masked by a template containing a slot (long thin hole). The template is drawn across the scene at a steady speed and perpendicularly to the slot in direction. The sound at a given time depends only on the part of the scene showing through the slot at that time. In the piano transform, the slot is the piano keyboard.

2.2.3 Kurcz (1981)

Kurcz (1981) describes a hand-held device called a heliotrope which senses the light output from only one point in the scene. The heliotrope is used to scan the scene manually at will in any direction and outputs a sound related to the light intensity at the point.

2.2.4 Deering and Tou & Adjouadi (1984)

Deering and Tou & Adjouadi, both in Warren & Strelow (1984), use verbal description as output, a line independently discovered by my daughter Shanti: "Easy, Daddy. If it sees a dog, why doesn't it just say "Dog!"?"

My instinctive revulsion against such a device needs explaining. Computers, compared to people, are notoriously bad at visual recognition. To build recognition into an optophone is to forget that optophones are to be worn by real people, potentially far better recognisers of objects than any computer. In addition, to build recognition into an optophone inevitably involves censorship of the scene, which I also find abhorrent.

2.2.5 O'Hea (1987)

O'Hea (1987) discussed the general problem of optophone mapping and considered a number of desirable properties that a mapping should have, arguing strongly for the presence of a fovea. Two mappings were simulated on computer.

One of the mappings simulated was the same as in Dallas's work, although O'Hea was unaware of this. He called this mapping the piano transform, and found it unsatisfactory, especially for conveying a wide light shape on a dark background, where the mapping is equivalent to trying to convey two notes on the piano (the edges of the shape) by playing all the notes in between.

The second mapping simulated, though only partially, was again a slot mapping, this time from edge orientation to musical (circular) pitch, and from position along the slot to interaural intensity difference. The edges were analysed at different spatial scales each separated by a factor of 2, with the corresponding sound two octaves higher or lower (a separation of one octave being a 180° edge rotation or sign reversal).

2.2.6 Nielsen, Mahowald and Mead (1989)

Nielsen, Mahowald and Mead (in Mead 1989) mapped the time derivative of light log-intensity at any place in a two-dimensional visual field to an auditory transient (click) filtered so as to appear to come from the same place in a two-dimensional auditory field (using simulated free-field listening). This mapping has the practical advantage of being uninterrupted in time.

It is claimed that the selection of time derivatives as the information to transmit enables the perception of motion and thus in theory the reconstruction of the third spatial dimension. While this is so for parallax motion of the camera, it is not clear how or whether the effect is suppressed during panning motion, and if so what is used instead. Presumably, steadily fixated scenes would produce silence in the same way as steadily fixated test objects disappear (Riggs et al, 1953).

It is not clear that the best possible mapping should turn a normally unnoticeable effect of human vision (the disappearance of steadily-fixated test objects) into an overwhelmingly present characteristic of the optophone.

2.2.7 Meijer (1992)

Meijer (1992), in a patent application by Philips of the

Netherlands, used the piano transform, in a way not obviously different from Dallas (1980) but in a more modern electronic implementation.

2.3 Spaces for sounds

2.3.1 Multidimensional scaling

In looking for a mapping from scenes to sounds, it is natural to ask if there is such a thing as a multidimensional scene or sound space, in which any scene or sound would be represented by a point. If so, and the two spaces for scenes and sounds were sufficiently similar, then simply equating the two spaces would produce a mapping.

Multidimensional scaling is an automatic technique designed to place a sensation into a multidimensional space in such a position that it is closest to sensations that appear most similar to it and farthest from those that appear most different (see references under PSYCHOPHYSICS - MULTIDIMENSIONAL SCALING). Famous examples of its use are the horseshoe shape of the colours of single-wavelength light and the circular arrangement of pure tones.

The technique has several variants, but they all involve asking subjects how different sample sensations are from one another. While it would be unsafe to ask, "How many times bigger is the difference between sensations C and A than between B and A?", it is reasonable to ask, "Is C more different from A than B is?". From the resulting ranking, the multidimensional space is derived.

Theoretically, it would be possible to represent every sound as a point in a subjective multidimensional sound space and every scene in a similar scene space, using these techniques. It would then suffice to equate the two spaces, or the N most important dimensions of each, to obtain the required mapping from scenes to sounds.

Unfortunately, not only would a sufficiently thorough sampling of all possible scenes produce a huge number of sample scenes, but subjects would be required to compare each of these samples with every other, making an astronomical number of comparisons. The same of course applies to sounds. The method is rapidly defeated by combinatorial explosion.

2.3.2 Trial and error

An attempt was made, in the case of steady sounds, to derive a subjective multidimensional space by reasoning

from what is already known.

Consider two sounds A and B consisting of a single pure tone each. Suppose we plot them as points on a graph of loudness against pitch (Figure 2.1). As shown, A and B are of the same loudness but different pitch. We then turn A and B down so that they are inaudible, obtaining sounds C and D. Whereas A and B sound different, C and D sound the same (silence), and yet there is still a distance between C and D on the graph. Thus distance on this graph cannot be made to represent difference in sensation.

An (r, θ) representation is more suited (Figure 2.2). If pitch is related to angle θ from the x axis, and loudness to distance r from the origin, then silence has only one position (the origin). If θ is so scaled that the audible frequency spectrum fits into 360°, then the space is largely used up by all possible pure tones. However, the close resemblance of tones one octave apart is not reproduced in this scheme.

θ can be rescaled so that 360° corresponds to one octave, and a third dimension introduced, also related to pitch (Figure 2.3). This new dimension z represents monotonic or straight pitch f, as opposed to θ which represents cyclical or circular pitch or pitch-in-the-octave p. The terms straight and circular will be used here. The space

for pure tones is now a helical surface. A previous defect is reintroduded, however, since all points along the new axis now represent silence.

This defect is overcome by collapsing all points on the new axis z on to the origin, or, equivalently, representing straight pitch not as distance along the z axis but by angle ϕ from it, and loudness as distance not from the z axis but from the origin (Figure 2.4).

What sounds can be assigned to the space between the turns of the helix? Two-tone sounds, with the tones one octave apart, fill the space nicely, with overall loudness as distance from the origin, and the relative loudness of the two tones proportional to the relative closeness of the two adjacent turns of the pure-tone surface.

Encouraged by the apparently successful derivation of this space, much thought went into extending it to encompass more complex steady sounds, with no success at all. One reason for suspecting that extension of any pure-tone space to more complex sounds would be inappropriate is that while a sound can be composed of many pure tones, it can only have one pitch. It is true that a sound containing only a few dominant pure tones can have a different pitch according to which tone is being attended to, but it can only have one pitch at a

time [Plomp, 1976].

2.4 Automated derivation of mapping

An apparently less restrictive approach is to ignore the dimensional structure of scenes and sounds and represent each by a one-dimensional list of numbers (a vector). Standard ways of doing this are the raster scan for scenes and the time sampling of air pressure for sounds, but there may be better ways for our purpose.

In contrast to the conscious pairing off of attributes described above, the idea here is to derive a suitable mapping blindly, using only the known statistical properties of the numerical vectors describing scenes and sounds, the known discriminatory properties of human sight and hearing, and some automatic procedure to do the derivation.

The approach adopted is based on the premise that if the following two requirements are met then the mapping will be satisfactory. First, that in order to "sound right" the sounds should be generated from vectors having the same mean, variance and covariance as vectors representing real sounds. Second, given that both sight and hearing are imprecise, that the imprecision in hearing the sound should correspond, via the mapping, to

the imprecision of human vision. In other words, a mapping should not do better than human vision at conveying one aspect of the scene if it means doing worse at conveying another.

2.4.1 Producing sounds with the "right" statistics

The first requirement may be achieved as follows. Suppose a large number of real sounds are represented as vectors s and the mean, variance and covariance of the vectors are calculated. Call the mean vector n. From the resulting covariance matrix, a matrix S can be derived such that premultiplying sound vectors s-n by S produces vectors w whose elements are totally uncorrelated.

$$w = S(s - n) \qquad (1)$$

This is a standard decorrelating procedure called principal component analysis and the matrix S is called a Hotelling or Karhunen-Loève (KL) transform (eg Gonzales & Wintz, 1987, p122ff). It is usual to arrange the rows of a KL transform so that the elements of the resulting uncorrelated vectors appear in order of variance.

Now suppose some vectors w are generated from uncorrelated random elements having zero mean and the correct variance, are then premultiplied by the

inverse S^{-1} of S, and have the mean vector n added back on.

$$s = S^{-1} w + n \qquad (2)$$

Then the resulting sound vectors s have the same mean, variance and covariance as those of real sounds.

Now suppose the same were done for scenes, obtaining scene vectors p (for "picture"), a mean scene m, a KL transform P, and uncorrelated vectors v from

$$v = P (p - m) \qquad (3)$$

If the scene and sound vectors p and s are chosen, by adjusting the amount of detail in one of them, to be of the same length, and if the elements of v and w are in order of variance, and if the variances of v and w are not too different, then v may be a good candidate for w, and sounds may be generated from scenes by

$$s = S^{-1} P (p - m) + n \qquad (4)$$

2.4.2 Proper distribution of imprecision

The second requirement depends on the elements of p and s being correctly scaled. Ozeki (1979) pointed out that unless proper attention was paid to the relative scaling of the different variables, principal component analysis

could be made to prove anything. For instance, one component could be made vastly more important by expressing it in millimetres instead of metres. Ozeki's solution was to scale the variables according to the amount of noise present in their measurement.

What seems to correspond to noise for our purpose is the difference limen (DL), also called the just noticeable difference (JND). In hearing, for instance, the intensity difference limen is in the region of 1 or 2 dB (Moore, 1989). If the sound vector s were chosen as a list of intensity values (say a raster scan of a spectrogram), then the numbers would have to be in say 2-dB units, and a change of 1 in the value of any of the numbers would be taken to correspond to a just detectable change in the sound. Let such a vector be called a psychophysical representation (PR).

Similarly, a vector representation for scenes would be required in which a change of 1 in the value of any of the numbers would correspond to a just noticeable change in the scene.

2.4.3 Limitations

There are two main difficulties in this approach. First, the scenes and sounds must be described, by initial and

probably nonlinear transformations from the raw data, in terms of PRs, that is, in terms of variables with values expressed in units of a difference limen. Difference limens, to be reproducible, are generally derived using very contrived sounds. It is not immediately clear that such results can be usefully extended to describe general sounds.

In many studies (**PSYCHOPHYSICS - HEARING - AUDITORY PROFILE ANALYSIS**), Green in particular has shown that the intensity difference limen of one tone in a complex sound depends on the intensity and frequency of the other tones (quite apart from the masking effect). In general, the more tones and the more equal their intensities, the smaller the intensity difference limen of any one of them.

Second, KL transforms only remove linear correlation. Any strong nonlinear correlations in either the scene or sound variables would invalidate the method.

A subjective method for testing the suitability of a PR for the present purpose is as follows. Having derived S or P, generate random sounds or scenes by using uncorrelated random numbers of the correct variance and applying S^{-1} or P^{-1}. Suitable PRs will produce sounds or scenes of every description and with no preponderance of any particular type. If this can be achieved to a

certain degree for both scenes and sounds, then the object of mapping all scenes to all sounds is also achieved to the same degree.

The progress made along these lines is reported in Chapter 6.

2.5 Invariances

Unfortunately, there are more invariances in vision than in hearing. Things undeniably look "the same" when moved sideways, upwards or further away. Sounds sound "the same" when slowed down (but with no frequency change) or delayed. Three against two. Interestingly, things don't look the same when rotated more than a certain amount. A square turned through 45° is called "a diamond". Reading upside-down is difficult. Upside-down faces are often impossible to recognise.
A similarly limited invariance on the sound side is invariance to frequency shift.

Of the three visual invariances listed, the strongest seems to be invariance to size. The strength of invariance to translation is a difficult one, since one doesn't usually try to recognise something without looking at it. Certainly the invariance seems strong with the fovea still inside the object boundary.

It is interesting to speculate what invariance in the scene could be made to correspond to speed invariance in the sound. Take two otherwise identical time-varying sounds, sound B lasting twice as long as sound A. Twice as much information can be extracted from B than from A, so a mapping from scene size to sound duration suggests itself. However, in the case of general complex sounds, the increase in detail is all in one direction, so there would be an increase in resolution in whatever in the scene maps to time in the sound, and no increase in the orthogonal direction.

There is a class of sounds where this is not true, and where a slowing down of the sound (still not altering the frequencies) involves greater resolution not in the time but in the frequency direction. This is the case with sounds made only of short pure tones, since it is known that, up to a duration of about 0.1 s, the frequency difference limen decreases with increasing tone duration (Moore 1989). It might be interesting to design some intermediate class of sound which would increase resolution in both directions when played slower. Three words of caution, however. First, restricting the sounds produced by an optophone to a certain class would violate our requirement n° 3 - completeness (see Chapter 1). Second, real sounds already contain spectral peaks (in the way that real scenes contain edges), and it is uncertain to what extent such a process happens already.

Third, we have to distinguish carefully between the difference limens of pure tones and the difference limens of spectral resonance peaks, which are much coarser.

It is important to realise that other invariances than those listed above come into play in human vision, notably invariance to affine (shear) distortions and distortions associated with rotation of three-dimensional surfaces about a vertical axis. Affine distortions, modified for perspective at close range, are the distortions that happen to, say, the letter R when it is painted on the three visible faces of a cube and then photographed. These invariances are comparable in strength to translation and size invariance, in that they do not hinder instant recognition.

Whatever the mapping, therefore, there will unfortunately be strong unmapped invariances. Is this a disaster? I think not. Invariances can be learnt. Suppose we have no built-in invariance to vertical translation, as with the cartesian piano transform with the keyboard vertical. An object shifted up four octaves (by the user tilting his head downward from looking say 30° above the object to 30° below it), and also stretched a bit by the action of the erb scale, will initially be unrecognisable (the Donald Duck effect). But the transformation to the sound will be exactly the same every time the head tilts downward through 60°, and similar to when it tilts 50°

or 70 . In time, one should be able to predict this transformation accurately, in the same way that one can learn in time to predict the path of falling objects and catch them, and to carry out an amazing variety of skilled tasks.

2.6 Colour

2.6.1 General

Colour is an immensely complicated and deceptive subject. We will only go into it here as little as necessary. A good starting point for further study is Pratt (1978).

It is known that there are three types of colour receptor in the human eye, and that the sensation of colour is due to differential excitation of the three types of receptor. This accounts for the representation of colour in digital systems by three numbers, usually either the amounts of red, green and blue or the amounts of brightness, saturation and hue.

Hue is a pure measure of colour excluding brightness or saturation. Saturation is a measure of the strength or paleness of the colour. Pratt (1978, p28) and Gonzalez & Wintz (1987, p192) define saturation in opposite

directions. While Pratt has "saturation describes the whiteness of a light source", Gonzalez & Wintz have "The pure spectrum colors are fully saturated" and "the degree of saturation being inversely proportional to the amount of white light added". Thus white is either totally saturated or totally unsaturated, and it's a good idea to check which way the word is being used. Here, white is unsaturated.

Consider a set of three cartesian coodinate axes. They form the edges starting at one corner of a cube, the corner being the origin. Let these three axes represent amounts of red, green and blue respectively. The origin therefore represents black. One line leaving the origin and going off into the cube joins progressively lighter greys.

In television systems, the numbers are so scaled that this grey line is straight and ends up at white at the far corner of the cube. Now place the cube with the grey line vertical and the origin at the bottom. Take a cross section a short way above the origin, roughly horizontal but with some tilt. A triangular figure results, with one corner on each of the three axes. It is possible to choose the tilt so that all the colours in the plane of the section have much the same brightness and differ only in saturation and hue.

Unfortunately, not all natural colours fall within the triangle, some requiring negative coefficients. In 1964 the CIE (Commission internationale de l'éclairage) specified chromaticity coordinates in which the three numbers for red, green and blue were never negative. Although the system had not much else going for it, it came into common use as a basis for the presentation of colours regardless of brightness.

Figure 2.5 is based on the 1964 CIE coordinates. The triangle shown is the triangle described above. The horseshoe is the pure colours (colours of one wavelength only), outside which no colours exist.

MacAdam (Pratt 1978) investigated colour difference limens and expressed the results in the CIE plane. Surrounding any point in the plane is a ring of other points representing just noticeably different colours. These rings became known as MacAdam's ellipses. Unfortunately, they vary manyfold in both size and ellipticality over the CIE plane, while in a psychophysically satisfactory plane they would be of constant size and circular.

Inspired by experiments on the perception of colour by partially colourblind people, Oleari (1991, 1993) produced a new and computationally tractable set of colour coordinates (Figure 2.5) in which difference

limens are constant. Where colours are used in the present research (Chapter 7), they are first expressed in terms of Oleari's saturation and hue.

Note that in Oleari's system the spectral colours do not all have the same saturation. This is reasonable, since the system is psychophysically based and there is no reason to expect the spectral colours all to appear equally saturated. However, it does pose a problem if it is desired to scale the saturation to the range 0 to 1 or 0 to 100%. A 100% value must then be chosen which most colours will never attain.

2.6.2 Should colour be included in an optophone?

The answer depends on the balance between how useful it is to do and how easy it is to do. As long as it was difficult, colour was omitted from television. Later, it was universally included.

My own feeling is that as long as it is not too difficult, it should be included. First, all the hardware for capturing colour scenes is there. Second, colour is useful, specially for looking for things and for living in a man-made environment. Third, optophones are bound to be difficult to use at the beginning, and the more clues as to the nature of an object the better.

Fourth, if colour is omitted initially as a matter of policy, there is the danger of the resulting black-and-white mapping being incompatible with a future colour mapping, requiring users to start the learning process again from scratch.

Refer to Figure 6.2 (or 6.3) giving spatial frequency sensitivity of human vision. The curve concerns grey scenes or the brightness of colour scenes. There is an equivalent curve for sensitivity to the spatial frequency of variation in chromaticity (colours of the same brightness). Starting at the fine end of the scale, on the "best the eye can do" side, the chromatic response curve rises from zero in a similar way to the achromatic curve but at resolutions more than three times coarser (Pratt 1978). Having reached its peak, however, the chromatic curve carries on forever at the peak response of 1. Thus the whole of the "best the eye wants to do" side is horizontal and equal to 1.

The conclusions for optophone design are twofold. First, it is not necessary or desirable to model colour in as fine a spatial detail as brightness. Because it only needs to be one tenth as finely specified on a solid-angle or pixel-count basis, it should demand relatively little computation and be well worth doing. Second, it is not necessary or desirable to reduce the response to colour as the size of the coloured area increases.

2.7 Music

2.7.1 General

Music is a subject that can do with a lifetime's study, and I am no musician. Should I therefore steer clear of it? The psychophysics of hearing is a subject that can do with a lifetime's study, and I am no psychophysicist. Reasoning along those lines, I would steer clear of most subjects and not make an optophone. It could well be that so little work on optophones has ever been attempted because there are no specialists in all the subjects involved. The only course in such circumstances is to have a go but to limit one's delving strictly to the exigencies of the task at hand. It is for this reason, for instance, that I have managed to avoid almost all reference to physiology, however interesting rods and cones and phase locking of spike trains might be.

The argument against music in an optophone is that there are nonmusical sounds. If nonmusical sounds are excluded, then our completeness requirement (Chapter 1) is violated. Thus there can be musical sounds but not only musical sounds. In fact, since musical sounds exist, the completeness requirement requires that they be

included. The question is whether they should just happen randomly every now and again as an unintentional result of the mapping, or whether they should be deliberately made to correspond to some subjectively separate aspect of scenes. Since we have a shortage of dimensions in sounds as compared to scenes, subjectively recognisable classes of sound, such as music, should indeed be brought into use in this way (unless we're deriving a mapping blind by the KL method).

2.7.2 Musical key space

One of the foundations of polyphonic music is the concept of key (Karolyi 1965). A key is formed by three notes in harmony, either spaced 3, 4 and 5 semitones apart to form a minor key, or 4, 3 and 5 semitones apart to form a major key. These notes are circular notes (see section 2.3.2 above) which is why three notes specify three intervals and not two. Looked at another way, three intervals in the normal way require four notes, but the first is the same as the fourth.

The keys are named by the lower note of interval 3 in the case of minor keys and of interval 4 in the case of major keys. Leapfrogging does not change the key: intervals (3, 4, 5), (4, 5, 3) and (5, 3, 4) all form the same key provided the lower note of interval 3 has the same name

in each case. There are twelve minor keys and twelve major keys.

The twenty-four keys can be pictured more clearly by examining the kind of space they occupy. Suppose all twenty-four are named on paper and joined by a line if they differ by only one note. The result is Figure 2.6. The keys lie on a two-dimensional surface. On close inspection, however, it is seen that the surface can be folded and rejoined so as to form a torus.

It is possible to move continuously between keys by moving along the lines shown. First the note to be changed is softened and disappears when the middle of the line is reached. Then its replacement gradually appears and reaches full strength at the end of the line. At the middle of the line only two notes are present. However, no two lines are centred on the same two notes, so there is no ambiguity in doing this.

Figure 2.1

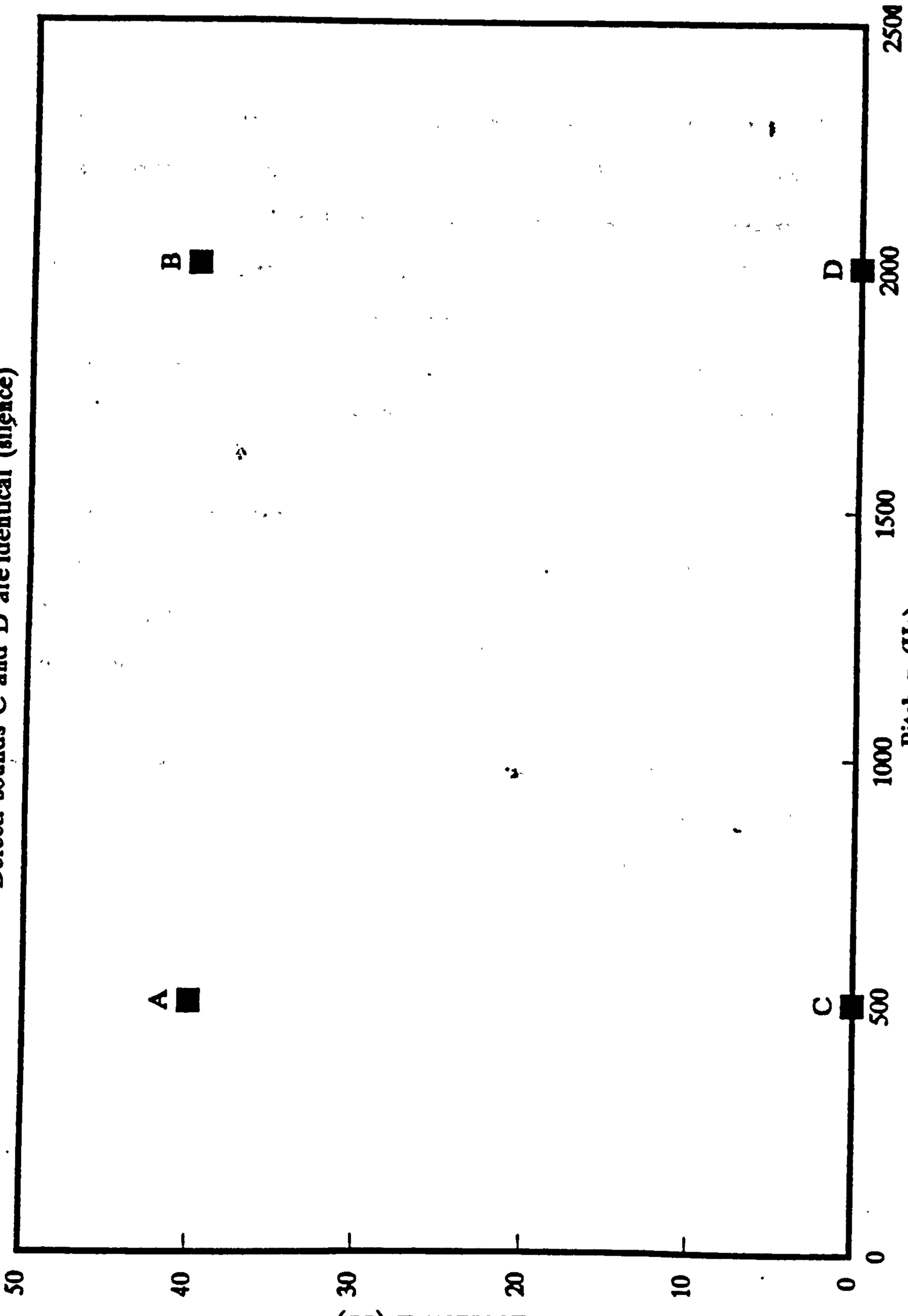
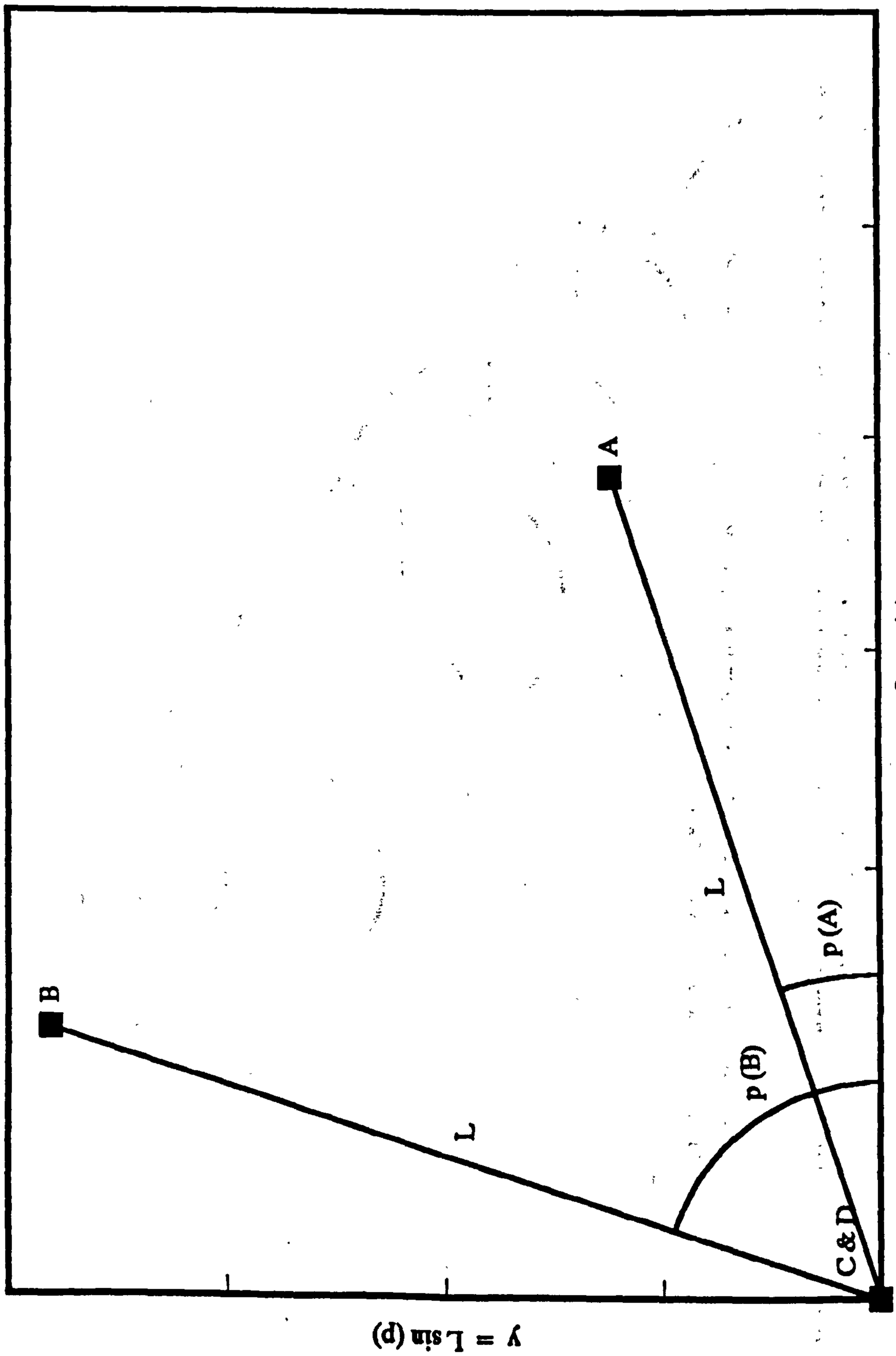

Pure-tone space: cartesian plane
Defect: sounds C and D are identical (silence)
A
B
C
D
500
1000
1500
2000
2500
Pitch p (Hz)
0
10
20
30
40
50
Loudness L (dB)

Figure 2.2


Pure-tone space: polar plane
Loudness L = distance from origin
Pitch p = angle from x axis
A
B
L
p(A)
p(B)
C&D
x = L cos (p)
y = L sin (p)

Pure-tone space: helical surface

Defect: all sounds on z axis are identically silence

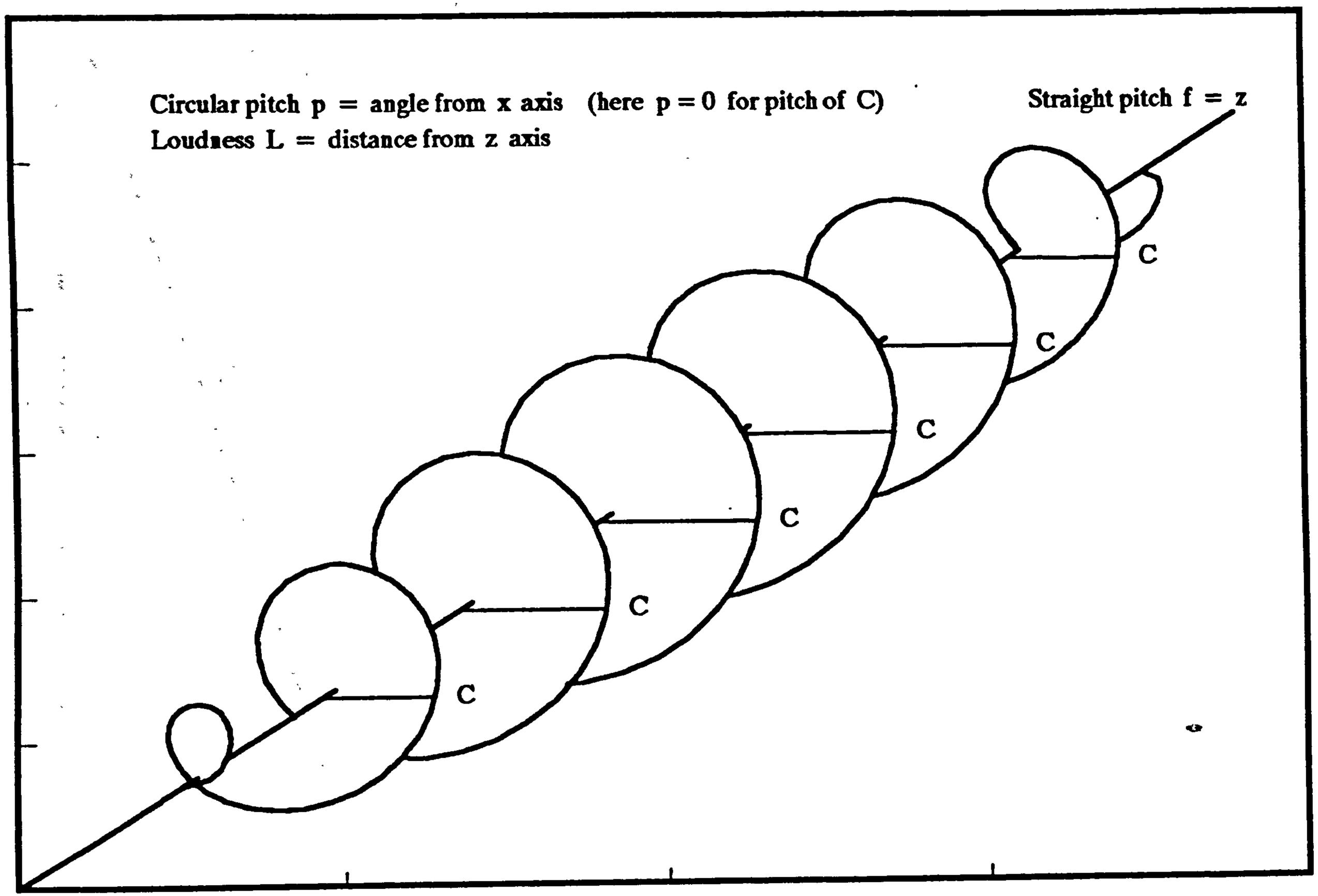
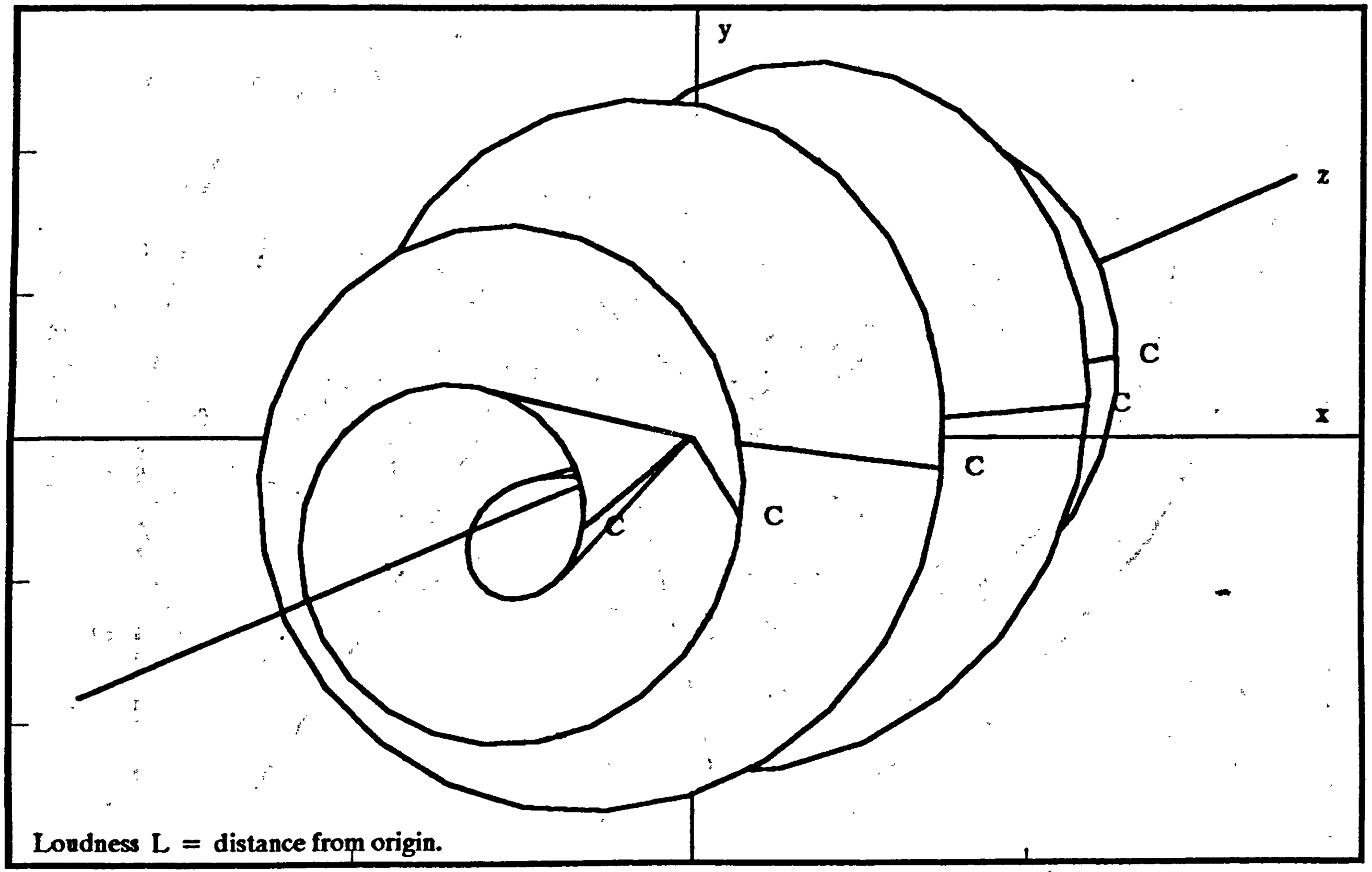


Figure 2.3

Pure-tone space: conical helix

Only the origin represents silence

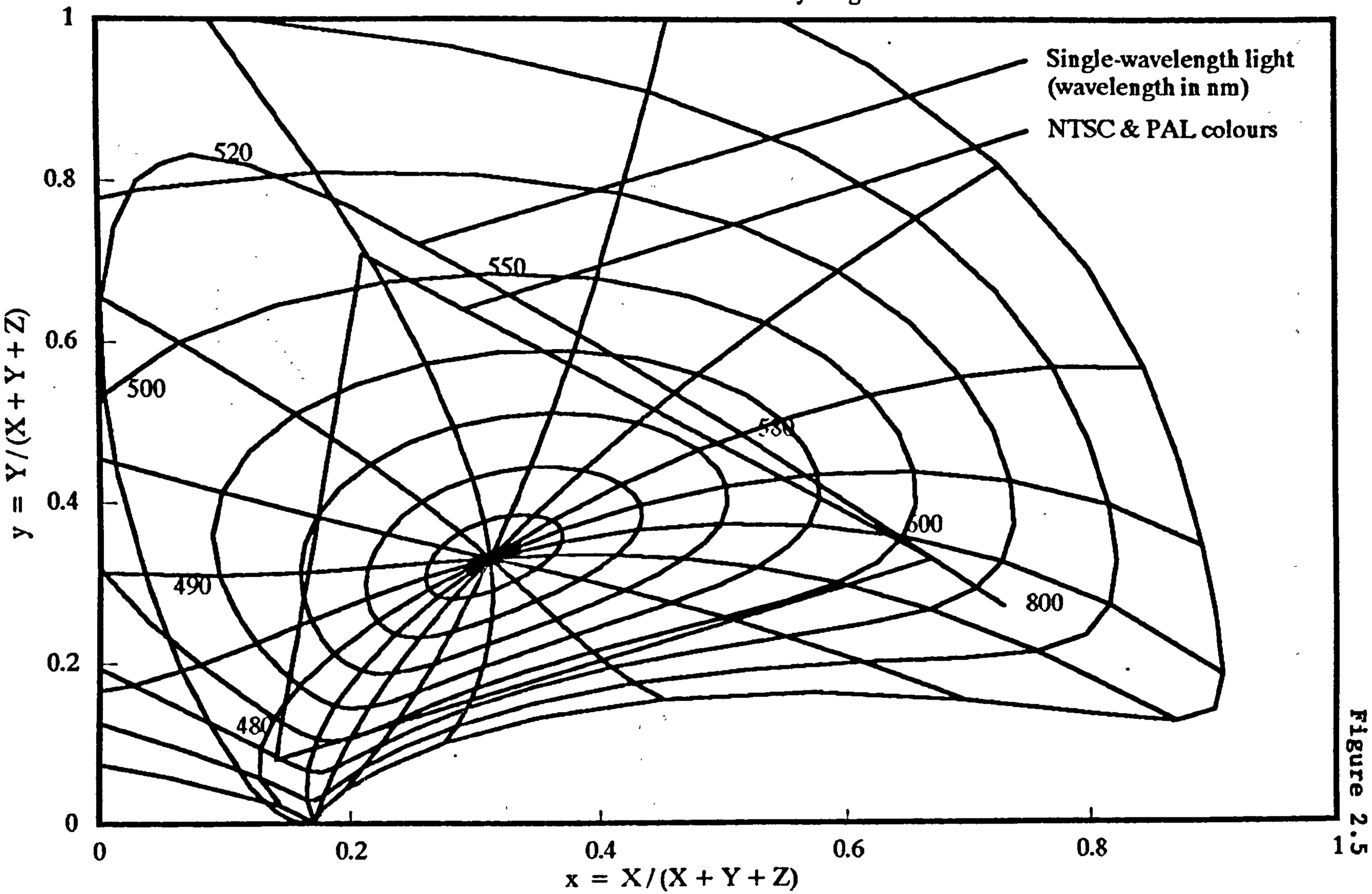


Straight pitch f = angle from negative z axis.
Circular pitch p = angle round z axis from xz plane (here p = 0 for pitch of C).

Figure 2.4

Oleari hue and saturation

on 1964 CIE chromaticity diagram



Sources: Oleari (1993), Pratt (1978).

Figure 2.5

Adjacent musical keys

Each linked pair of keys differs by only one note according to one of the three progressions shown.

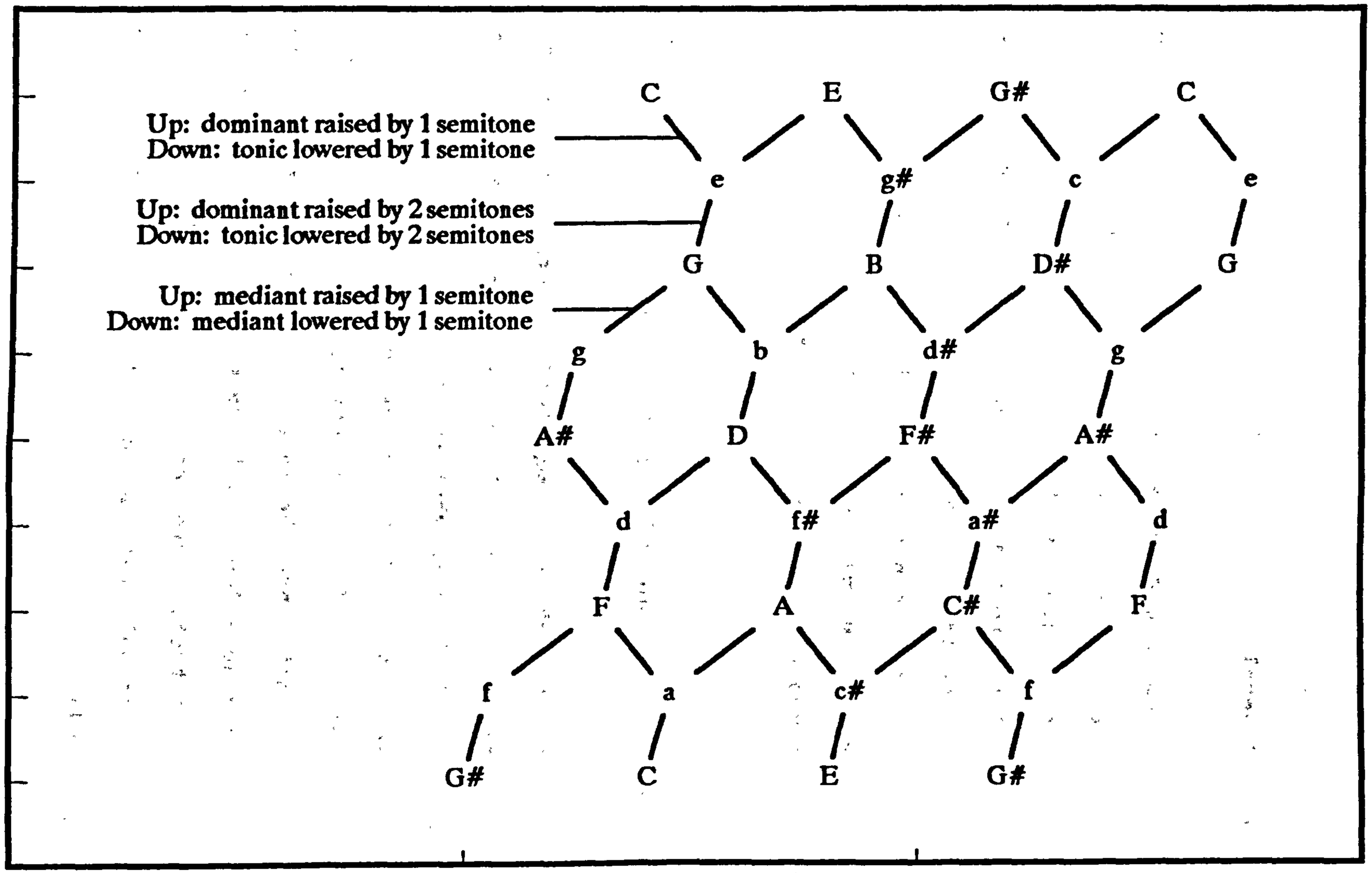


Figure 2.6

Minor keys are called by small letters.
Top row is same as bottom row, left columns are same as right columns. Keys therefore lie on surface of torus.

CHAPTER 3 TESTING OF MAPPINGS

3.1 General

It is very difficult to think of a test that would rank scene-to-sound mappings in order of desirability. Suppose two optophones have been invented and it is desired to compare them. In the same way as magazines review cars or computers, it would be possible to allocate points to such attributes as battery life, weight, price, reliability in the rain, and so on, and compare a weighted sum of the results.

3.1.1 Categorical testing

But how about the mapping? In comparing two mappings, the two things to be compared are themselves comparisons, namely between what is there and what is seen. (Arguments as to whether an optophone would allow a blind person to "see" are futile. The convention adopted here is that since the input to the optophone is light, the optophone does enable, however badly, the user to see.) In a test, therefore, a user must be able to report accurately what he sees, or no proper comparison can be made.

Unfortunately, a picture being worth I forget how many thousand words, reporting accurately what you see is only possible if it involves naming items in the picture the listener already knows about.

One field in which it is easy for the user to report accurately what he sees, in the sense of naming items in the scene, is reading out loud. Comparison of reading speeds, as a function of hours of training, would give useful results.

In another test, a user might be asked to name as quickly as possible all the objects on the table in front of him.

3.1.2 Noncatgorical and objective testing

General seeing, of a noncategorical nature, is more difficult, and few tests come to mind. Some might find the following test impressive: blindfolded or recently blinded realistic painters might be asked to paint the scene in front of them.

On the other hand, it can be argued that all useful seeing is categorical, in the sense implied by the sentence "I can't see what that is." In that sense, worrying about testing noncategorical seeing is a red herring.

I am particularly concerned not to get bogged down in an endless series of tests using artificial patterns designed to describe objectively the performance of a particular mapping, in the way that human vision is tested in psychophysical experiments. The key word here is "objectively", and I must say that for me it is a dirty word. All tests, in all fields of enquiry, are ultimately subjective. "Objective" tests are only locally objective, and are in fact subjective by reason of the subjective choice of the test criterion.

An example of a nominally objective test is the theoretical performance test described below. The performance criterion is carefully spelled out and the mappings tested against it. Unfortunately there are subjective reasons for thinking that the criterion is not very good.

In case the distinction between the objective and subjective approaches isn't yet clear, I'll have one last go. The objective approach is to carefully specify a test criterion and to test performance against it. The subjective approach is to carefully choose a test criterion, then carefully specify it and test performance against it. The first is good *science*, the second *good* science.

3.1.3 Importance of training

Care will be needed to test the performance of the mappings and not of the users. Have the users all had sufficient training? It is important to bear in mind the amount of training expected to be needed for even a good mapping to prove its worth: several hours a day for several months.

Note that the word "training" is not intended here to mean a predetermined or directed activity, merely time spent using one's optophone.

To be convinced of this, consider another crossmodal activity: representing the sounds of a language by arbitrary shapes on paper; otherwise known as writing. How long does it take to learn to read? The answer depends almost entirely on how fast. Alphabets (used in Europe) and syllabaries (in India and south-east Asia) can be learnt in a day (and, if not used, forgotten almost as quickly) in the sense that any word can then be deciphered without reference to a key.

Suppose on the other hand one wanted to know whether Hindi writing (the Devnagri script) was better than English writing (the Roman script) for the English language. Suppose a test were devised involving reading speed, English in Roman against English in Devnagri.

There is no difficulty in doing this: many signs in India are in English in Devnagri script (Bank of India, Indian Airlines, and so on). Now let's ask the question again (how long does it take to learn to read?) in the following sense: how long would one have to practise reading English in Devnagri in order for the result of the test not to be biased if favour of Roman? Quite a long time.

3.1.4 Previous work

Nothing has been found in the literature on the subject of testing scene-to-sound mappings. Mann (1965) and Tachi et al (1983) have devised an automated procedure for evaluating the navigational skills of users of mobility devices for the blind. However, the mobility devices they test are designed to impart a predetermined course for the blind person to follow.

3.2 Theoretical performance test (TPT)

3.2.1 Motivation

Hearing is imperfect in that slightly different sounds are indistinguishable. If that were not the case, then

the ears would have unlimited bandwidth (information carrying capacity) and many mappings would be perfect. One way to evaluate a mapping from scenes to sounds is as follows.

1 Obtain a digital still scene using an electronic camera.

2 Calculate the corresponding sound using the mapping under trial.

3 Calculate an almost perceptibly different sound using the known inaccuracies of human hearing.

4 Recalculate the digital scene using a suitable inverse version of the mapping.

5 Criticise the recalculated scene visually, by comparison with the original or otherwise.

6 See if there emerge any clues to a better mapping.

This section presents a method for corrupting sounds (Step 3) for this purpose.

All sounds produce an excitation pattern along the basilar membrane of the ear (Moore, 1989). The

excitation pattern resulting from a simple sound consisting of five pure sinusoids is shown in Figure 3.1. The excitation pattern is a graph where the ordinate, representing the strength of the excitation, is a function of the abscissa, representing either frequency or position along the membrane, the two being monotonically related. The value of the ordinate at a particular frequency is derived by weighting the powers of all the sinusoids or narrow noise bands in the sound according to their distance from the frequency in question, and summing (Moore & Glasberg, 1983). (It is assumed that the sound is sufficiently steady for temporal masking to be ignored.) The ordinate is therefore in units of sound power (W/m^2) or in decibels thereof, while the abscissa is in Hz or some monotonic function of Hz such as octaves, erbs (Figure 3.2) or critical bands (Moore, 1989).

There are good reasons for taking the parameters of the excitation pattern, and not the usual physical parameters of the sound, as our mathematical description of what is to be corrupted. If the excitation pattern is taken as an exact function of the sound, it follows that small changes in the pattern, just like small changes in the sound, are inaudible. When the sound consists of a single sinusoid (or narrow noise band), it is possible to determine by what margin its power can be imperceptibly varied, and if the answer is ± 1 dB, you can either say

that the difference limen of the power of the sinusoid is 1 dB, or that the difference limen of the excitation level is 1 dB. However, if a second, louder sinusoid is now added to the sound, the original sinusoid may be completely masked, in which case it may be increased in power severalfold or removed completely without noticeable effect, and its difference limen is no longer 1 dB.

A procedure which faithfully reproduces this effect is to apply the difference limen concept to the excitation pattern, and say that two sounds are noticeably different if their excitation patterns differ anywhere by more than the relevant difference limen. There is much in the literature on exactly how true this is (see under **PSYCHOPHYSICS - HEARING - MASKING - Simultaneous**, in particular Lutfi (1983) and Moore (1985)), but it is taken to be sufficiently true for the present purpose.

3.2.2 Bounds on corrupted intensities - Method 1

With the above in mind, what would it mean to "calculate an almost perceptibly different sound using the known inaccuracies of human hearing"? Clearly, the corruption must not be such as to produce a noticeably different excitation pattern, so one approach might be "to calculate the excitation pattern, introduce random errors

into it (random but smaller than a difference limen), and recalculate the sound". The difficulty here is how to prevent the random errors from producing an impossible excitation pattern, that is, one that can only be produced, mathematically, by some sinusoids having negative powers.

If P is a vector of the powers of a set of n sinusoids in order of increasing frequency, and E is the vector of the excitation levels at the same n frequencies, then

$$E = AP \qquad (1)$$

where A is an n by n attenuation matrix with unit principal diagonal and values tailing off towards zero in the other two corners. If the sinusoids are few and well separated in frequency, A is little different from the unit matrix I.

Corrupting E and reversing (1) we have

$$P' = A^{-1}(E + R) \qquad (2)$$

where R is a vector of the n random errors and P' is the reconstituted sound.

To appreciate the difficulty, say

$$n = 3 \qquad P = \begin{matrix} 0 \\ 1 \\ 0 \end{matrix} \qquad A = \begin{matrix} 1 & .39 & .08 \\ .42 & 1 & .39 \\ .11 & .42 & 1 \end{matrix}$$

giving, from (1),

$$E = \begin{matrix} .39 \\ 1 \\ .42 \end{matrix}$$

Now say

$$R = \begin{matrix} -.08 \\ .25 \\ -.08 \end{matrix}$$

with each element less than 1 dB (26%) of E.

Inverting A,

$$A^{-1} = \begin{matrix} 1.2 & -.51 & .1 \\ -.54 & 1.42 & -.51 \\ .1 & -.54 & 1.2 \end{matrix}$$

and, from (2),

$$P' = \begin{matrix} -.23 \\ 1.45 \\ -.24 \end{matrix}$$

Not only are the first and third elements of P' negative, but the second is more than 1 dB (our assumed difference limen) from the second element of P. So although E' (= E + R) is within a difference limen of E, it is not

a possible nor even a useful excitation pattern.

3.2.3 Bounds on corrupted intensities - Method 2

A different approach is to ask what bounds the perception of E puts on the values of P'. That is, what are the minimum and maximum values of p_i' that would not produce an audible change in E?

By the nature of the attenuation matrix A, the element of E most affected by changes in p_i is e_i, so the minimum and maximum values of p_i' are those that decrease and increase e_i by the relevant difference limen, or zero if such a decrease is not possible. Taking A and P and the difference limen as before, and using superscripts - and + for minimum and maximum, we have

$$E^- = \frac{E}{1.26} = \begin{matrix} .31 \\ .79 \\ .33 \end{matrix} \qquad E^+ = 1.26E = \begin{matrix} .49 \\ 1.26 \\ .53 \end{matrix}$$

Taking p_1 as an example, we have, from (1),

$$p_1^+ + a_{12}p_2 + a_{13}p_3 = e_1^+$$

$$
\begin{aligned}
p_1^+ &= e_1^+ - (a_{12}p_2 + a_{13}p_3) \\
&= e_1^+ - (e_1 - p_1) \\
&= p_1 + e_1^+ - e_1
\end{aligned}
$$

and, in general,

$$p_i^+ = p_i + e_i^+ - e_i$$

or

$$P^+ = P + E^+ - E \qquad (3)$$

Similarly,

$$P^- = \max\,(0,\ P + E^- - E) \qquad (4)$$

So here,

$$P^- = \begin{matrix} 0 \\ .79 \\ 0 \end{matrix} \qquad P^+ = \begin{matrix} .1 \\ 1.26 \\ .11 \end{matrix}$$

Notice that each element of P^- and P^+ is derived here without regard to its neighbours. That is, P^- and P^+ contain the lower and upper bounds to the power of any sinusoid provided the others are unchanged. The question arises whether any sound P' lying between P^- and P^+ is an acceptable corruption of P. Take $P' = P^+$. The corresponding excitation pattern, from (1), is

$$E' = AP' = \begin{matrix} .6 \\ 1.34 \\ .65 \end{matrix}$$

which is clearly greater than E^+ and therefore out of bounds.

3.2.4 Bounds on corrupted intensities - Method 3

A solution to this difficulty is to delay the derivation of bounds p_{l+1}^- and p_{l+1}^+ until some corrupted value p_l' lying between p_l^- and p_l^+ has been chosen and substituted for p_l.

Define P' as an updatable vector containing p_l for those sinusoids not yet corrupted and p_l' for those already corrupted. Define $E' = AP'$. Take P, A, E, E_- and E^+ as before, and suppose p_1' is chosen as $p_1^+ = .1$, also as before. P' is now $[.1\ 1\ 0]^T$, and E' is $[.49\ 1.04\ .43]^T$. Note that $e_1' = e_1^+$, as expected. Now to choose p_2'. Since e_1' depends to some extent on p_2' and is already at its upper limit when calculated using $p_2' = p_2$, p_2' can be chosen no greater than p_2. This shows that, in choosing how to corrupt the power of one sinusoid, we must take care that the new excitation pattern remains within bounds at all other locations as well.

In our example with three sinusoids, consider the bounds on p_2' with p_1' already chosen. P' is $[p_1' \; p_2 \; p_3]^T$, and E' is AP', or, in full,

$$\begin{aligned} p_1' + a_{12}p_2 + a_{13}p_3 &= e_1' \\ a_{21}p_1' + p_2 + a_{23}p_3 &= e_2' \\ a_{31}p_1' + a_{32}p_2 + p_3 &= e_3' \end{aligned}$$

The upper bound p_2^+ to p_2' is given by

$$\begin{aligned} p_1' + a_{12}p_2^+ + a_{13}p_3 &\le e_1^+ \\ a_{21}p_1' + p_2^+ + a_{23}p_3 &\le e_2^+ \\ a_{31}p_1' + a_{32}p_2^+ + p_3 &\le e_3^+ \end{aligned}$$

With some rearrangement and substitution these three equations give

$$\begin{aligned} p_2^+ &\le \frac{e_1^+ - p_1' - a_{13}p_3}{a_{12}} \\ &\le \frac{e_1^+ - (e_1' - a_{12}p_2)}{a_{12}} \\ &\le p_2 + \frac{e_1^+ - e_1'}{a_{12}} \end{aligned}$$

$$\begin{aligned} p_2^+ &\le e_2^+ - a_{21}p_1' - a_{23}p_3 \\ &\le e_2^+ - (e_2' - p_2) \\ &\le p_2 + e_2^+ - e_2' \end{aligned}$$

$$\begin{aligned} p_2^+ &\le \frac{e_3^+ - a_{31}p_1' - p_3}{a_{32}} \\ &\le \frac{e_3^+ - (e_2' - a_{32}p_2)}{a_{32}} \\ &\le p_2 + \frac{e_3^+ - e_3'}{a_{32}} \end{aligned}$$

and, in general,

$$p_j^+ = p_j + \min_i \frac{e_i^+ - e_i'}{a_{ij}} \qquad (5)$$

Similarly,

$$p_j^- = \max\left(0,\ p_j + \max_i \frac{e_i^- - e_i'}{a_{ij}}\right) \qquad (6)$$

Note that there is no requirement in this method to corrupt the sinusoids in increasing order of frequency or in any other particular order.

In the above discussion the intensity difference limen has been considered to be a fixed number of decibels regardless of level. For very quiet sounds this isn't true, as illustrated in Figures 3.3 and 3.4, which give the slightly more complicated equations for e^- and e^+ necessary to deal sensibly with sounds near threshold (Figure 3.5). The behaviour of the two bounds near threshold is shown in a complete excitation pattern in Figure 3.1.

3.2.5 Choice of corrupted intensities

To complete the corruption of P into P', we have to decide how to choose each p_i' knowing the bounds p_i^-

and p_i^+. Now P' is the user's best stab at P, so it seems reasonable to choose for p_i' the most likely value of p_i according to its probability distribution between the given bounds. However, the statistics of P will depend on the mapping producing P from the scene, so we have to look at the statistics of the scene.

The necessity for the best stab at P was demonstrated by early attempts to use, instead, the worst stab at P that was still inaudibly different from P. In the scene reconstructed from the corrupted sound, this worst-stab strategy produced obvious artefacts such as pronounced striping in some direction. A user would obviously not be fooled into thinking that everything he looked at was striped, and would instead try to get the best out of his optophone.

Let X be a suitable set of parameters describing the scene, chosen so that each p_i is a function of only one x_r, and not of (x_r, x_s, ...). This is in general possible because such a tuple would only be generating one p_i and can therefore be replaced by a single x_r which is a function of the tuple.

However, in the case of an unspecified mapping, (p_j, p_k, ...) may also be functions of x_r. This can arise for instance if it is decided, for clarity, to sound some feature of the scene in two different ways in case one is

masked.

Let

$$p_i = f_i(x_r) \tag{7}$$

To avoid ambiguous inverses, let each f_i be monotonic and either rising or falling. Then

$$x_{ri}^- = f_i^{-1}(p_i^-) \tag{8}$$

is either a lower or an upper bound on x_r depending on the sign of the slope of f. Similarly,

$$x_{ri}^+ = f_i^{-1}(p_i^+) \tag{9}$$

Combining the evidence from each of the relevant sinusoids, we have

$$x_r^- = \max_i \min(x_{ri}^-, x_{ri}^+) \tag{10}$$

and

$$x_r^+ = \min_i \max(x_{ri}^-, x_{ri}^+) \tag{11}$$

as the lower and upper bounds on x_r, where i refers to those sinusoids whose powers are functions of x_r.

Let X' be the corrupted scene, and choose x_r' between x_r^- and x_r^+ as the centroid of the probability distribution of x_r between those bounds. Again, this will depend on

what set of parameters X actually is. Having chosen x_1', update P' with

$$p_1' = f_1(x_1') \tag{12}$$

Note that in the general case the elements of P' are updated in dependent groups rather than singly as described.

3.2.6 Limitations

The first limitation is that the TPT is only possible if the mapping has an inverse with which to reconstitute the scene from the corrupted sound. Chapter 7 presents such a scheme with no inverse.

The second limitation is that the additivity of masking is only approximate, as mentioned above.

The third limitation is that the test, as it stands, only applies to sufficiently steady sounds. When trying to convey large amounts of information as sound, it is natural to do so as fast as possible. It is known that the perception of a spectral profile degrades as the presentation time is reduced. The TPT is of no help in estimating the presentation speed giving best performance.

It may be possible to adapt the kernel or point-spread function derived in Chapter 6 to produce a two-dimensional TPT along the lines of the one-dimensional TPT described here. On the other hand, it may be thought not worth while, since the best presentation speed may also be expected to increase with user training.

The fourth limitation applies when the TPT is used for each ear separately. The method in effect assumes that two independent signals can be received in each ear without difficulty, whereas in fact both signals can be used only if closely related, when the differences between the two become significant. If the differences between the signals at each ear are not of the type caused by the position of a single sound source, then the signals are incompatible and one or other must be ignored.

The fifth limitation is that the information is considered to be contained in the intensity of the sound at predetermined frequencies. A scheme containing say 10 spectral peaks, with the information contained in both the intensity and frequency of the peaks, is not covered. Actually, this is an instance of an earlier limitation: a mapping with no inverse. In order to apply the theoretical performance test to such a mapping, an inverse would have to be devised in the form of a peak-picking algorithm working on the corrupted excitation

pattern and giving as output corrupted versions of the peak intensities and frequencies.

The sixth limitation is that however many frequencies (sinusoids) in the audible range are used, and however closely packed they are, it is assumed that there can never be any confusion between them. For instance, where the TPT is used in the following chapters, the number of rows of pixels and so the number of sinusoids is 175. Given that the audible range is about 30 erbs wide, that makes 0.17 erbs per sinusoid. While this is wider than a frequency difference limen, whether for sinusoids or narrow noise bands (Moore 1973a&b, Gagné & Zurek 1988), confusions in the heat of the moment still seem probable. Suppose one sinusoid completely masks its neighbour, a common occurrence at such a close spacing, and that the dominant sinusoid is mistakenly identified as the masked neighbour. Now consider two mappings, one in which adjacent sinusoids are closely correlated, one not. The identity mistake would be more serious in the second mapping than in the first, but this difference in performance between the two mappings would not be revealed by the TPT.

The seventh limitation concerns interference between closely spaced sinusoids. The spacing being less than an erb, there will be interference (heard as beats if the sound stays steady long enough). The closer the spacing,

the longer the interference repetition period, and the slower the sound must be presented to allow a sufficiently long time average of the intensities to get the assumed accuracy. This is not properly modelled by the TPT, which assumes the excitation pattern to be derived from a time average of the intensities, whereas in fact the excitation pattern has its own time constant or equivalent rectanglar duration (ERD) of around 8 ms (Moore & Glasberg 1988). Note that it is not immediately clear whether this deficiency of the TPT underestimates or overestimates the performance of a mapping with closely spaced frequencies.

I consider that these limitations of the TPT are so severe and fundamental that further work to improve it, although interesting, would be a waste of time and money better spent on testing mappings in their proper environment, namely a functioning head-mounted prototype optophone, which as discussed above would be so much more revealing.

3.2.7 Implementation

Main program main in file \cwork\progs\hear.c

main first reads a data file heardata.t containing instructions as to the various options for the run. Some

questions have only one allowable answer and were intended for future use. Not all options available have even been tested.

main then calls hear.

Data file heardata.t in \cwork\progs

For completeness, an example of heardata.t is given here.

```
picture (with extension .bm or .q)                          gbike.q
add to output file name                                     c
start at line                                               0
scan                                                [hv]    v
brightness function                                  [r]    r
colour function (not for g files)                    [r]    r
colours treated (not for g files)                   [st]    s
three transforms (g files use only 1st)             [pc]    ccc
frequencies                                         [eh]    e
  e&h: bottom frequency                              (Hz)    50
  e:   frequency spacing                            (erbs)   .2
loudness function                                   [eps]    p
  e: exponential  poc = pow(rl, kx)                   rl     10
  p: power        poc = pow(max(0, kx), n)            n      1.1
  s: shift        poc = kx + s                        s      2
ear distribution                                    [msi]    s
```

Allowable options are in sqare brackets. Where only one option appears, others were foreseen but not implemented. The colour options have never been used and are not tested. All input files therefore must begin with g (grey). The important options are as follows.

The two transforms available are p for piano and c for cos, described in the following chapters.

The frequency options are e for equalerb and h for harmonic, and concern the distribution of frequencies in the audible range. Only equalerb has been tested.

The loudness function relates scene variable x to sinusoid power p in the form of a function from kx to p/c, called kxtopoc in the code, with inverse poctokx. Here k is either 1 or -1, as determined by the choice of ear distribution, and c is a reference power varying with frequency to give the proper "pre-emphasis". Only loudness function p has been tested, with n chosen for no reason at all as 1.1.

The ear distribution options are m for mono, s for stereo and i for independent. In the mono case, there is only considered to be one ear and the scene vector X is distributed along the audible range of the frequency scale. In the stereo case, each element of X may be sounded in either ear, in the left ear when negative and

in the right ear when positive. The intention here is that extreme values should be heard (not masked) whether positive or negative. In the independent case, the two ears are considered independent, and half the elements of X are assigned to the left ear and half to the right.

Function hear in file \cwork\lib\hearing2.c

hear first calls sethearing in file \cwork\lib\hearing.c to set various hearing constants.

hear then calls settransform in file \cwork\lib\trans.c and xstatistics in file \cwork\lib\pic.c to calculate the statistics of the variables in X necessary to calculate when the time comes the distribution of the current x knowing the adjacent previously corrupted values in X.

hear then calls eardistribution in file \cwork\lib\hearing2.c to allocate frequencies in each ear to each element of X.

hear then calls setfreq in file \cwork\lib\hearing2.c to calculate various constants at each frequency and in particular the attenuation matrix A between the list of frequencies fs at which sinusoids are present and the list of frequencies fe at which it is desired to calculate excitation levels.

For each row or column of the picture, hear then first calls fortransform in file \cwork\lib\hearing2.c to carry out the transform from picture column (or row) to X, then calls corrupt in file \cwork\lib\hearing2.c to corrupt X according to the TPT described above, and then calls backtransform in file \cwork\lib\hearing2.c to recalculate a corrupted column (or row) of the picture from the corrupted X.

From pixel value to element of X

Let the pixel value be y_{255}, with a range of 0 to 255. For compatibility between transforms, it is desired first that the input to the forward transform have elements of unit variance, and second that all transforms be unitary (that is, that the sum of the variances of the transform output equal the sum of the variances of its input).

Define a pixel variable y_1 with range 0 to 1 and related to y_{255} by

$$y_1 = \left(\frac{y_{255}}{255}\right)^{\frac{1}{n_1}} \qquad (27)$$

with n_1 chosen so that y_1 has a mean value of 0.5. Let the standard deviation of y_1 be sig_1. Values of

$$n_1 = 2 \qquad (28)$$

and

$$sig_1 = 0.2 \tag{29}$$

have been fixed as being reasonable.

Define a second pixel variable y related to y_1 by

$$y = \frac{y_1}{sig_1} \tag{30}$$

and designed to have unit variance. Combining the two, we have

$$y = \frac{1}{sig_1}\left(\frac{y_{max}}{255}\right)^{\frac{1}{n_1}} \tag{31}$$

as an element of the input vector Y to the forward transform.

The output from the forward transform is the vector X. The implemented options for fortransform and backtransform are only piano and cos.

In the case of the piano transform, elements of X are related pixel by pixel to the corresponding elements of Y by the equation

$$x = y - \sqrt{6} \tag{32}$$

I can't remember where the 6 came from, but it is very close to the mean ybar of y, given by

$$\bar{y} = \frac{\overline{y_1}}{sig_1} \qquad (33)$$

which comes to 2.5.

For additional details, see the relevant chapters below.

Function corrupt in file \cwork\lib\hearing2.c

corrupt carries out the TPT described above. Notable functions called are etoeb in \cwork\lib\hearing.c and ebtoxb in \cwork\lib\hearing2.c. etoeb ("excitation to excitation bounds") calculates the upper and lower bounds on imperceptible changes to a given excitation level. ebtoxb ("excitation bounds to x bounds") translates the bounds on the excitation level to the corresponding bounds on the value of x.

Figure 3.1

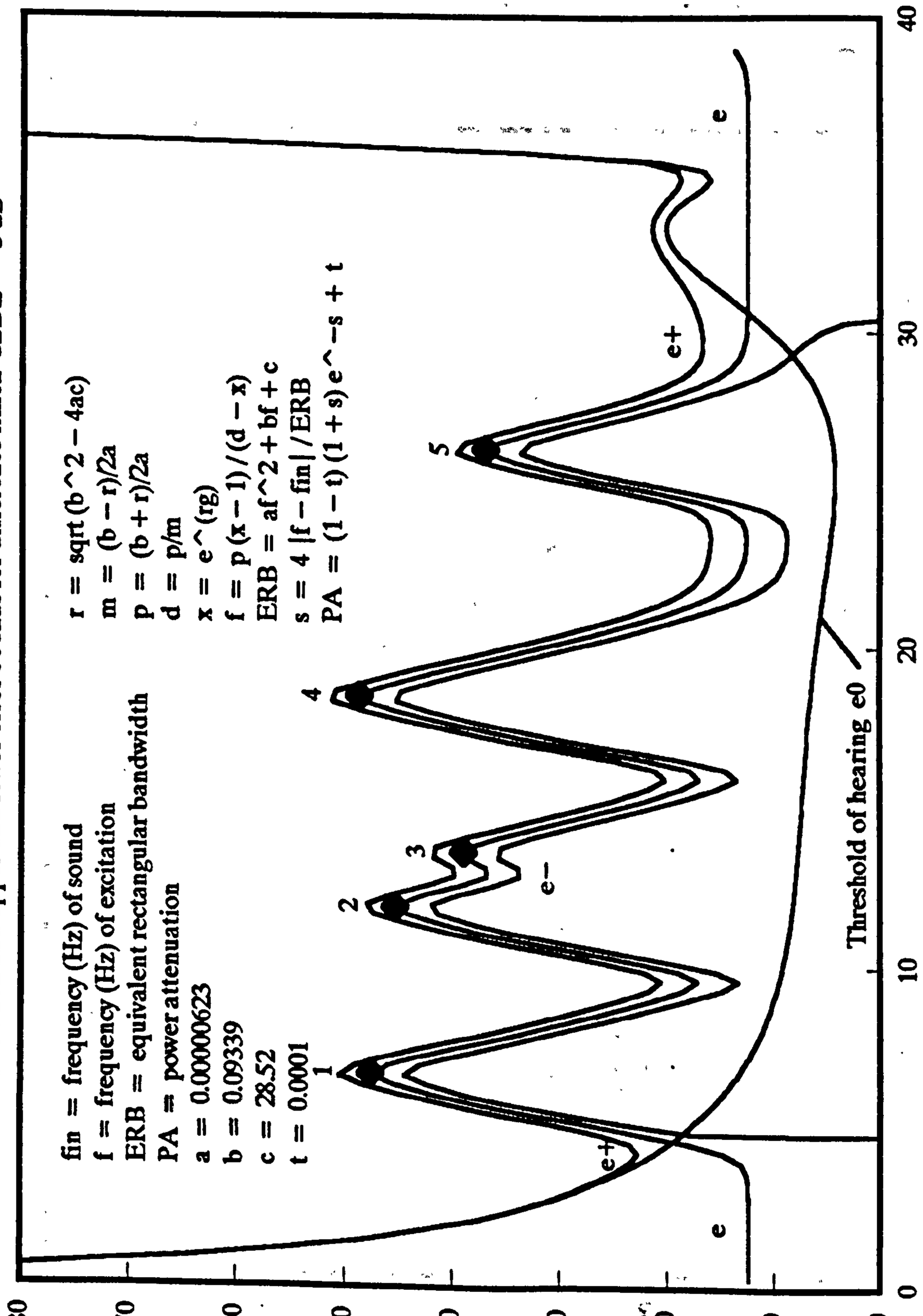
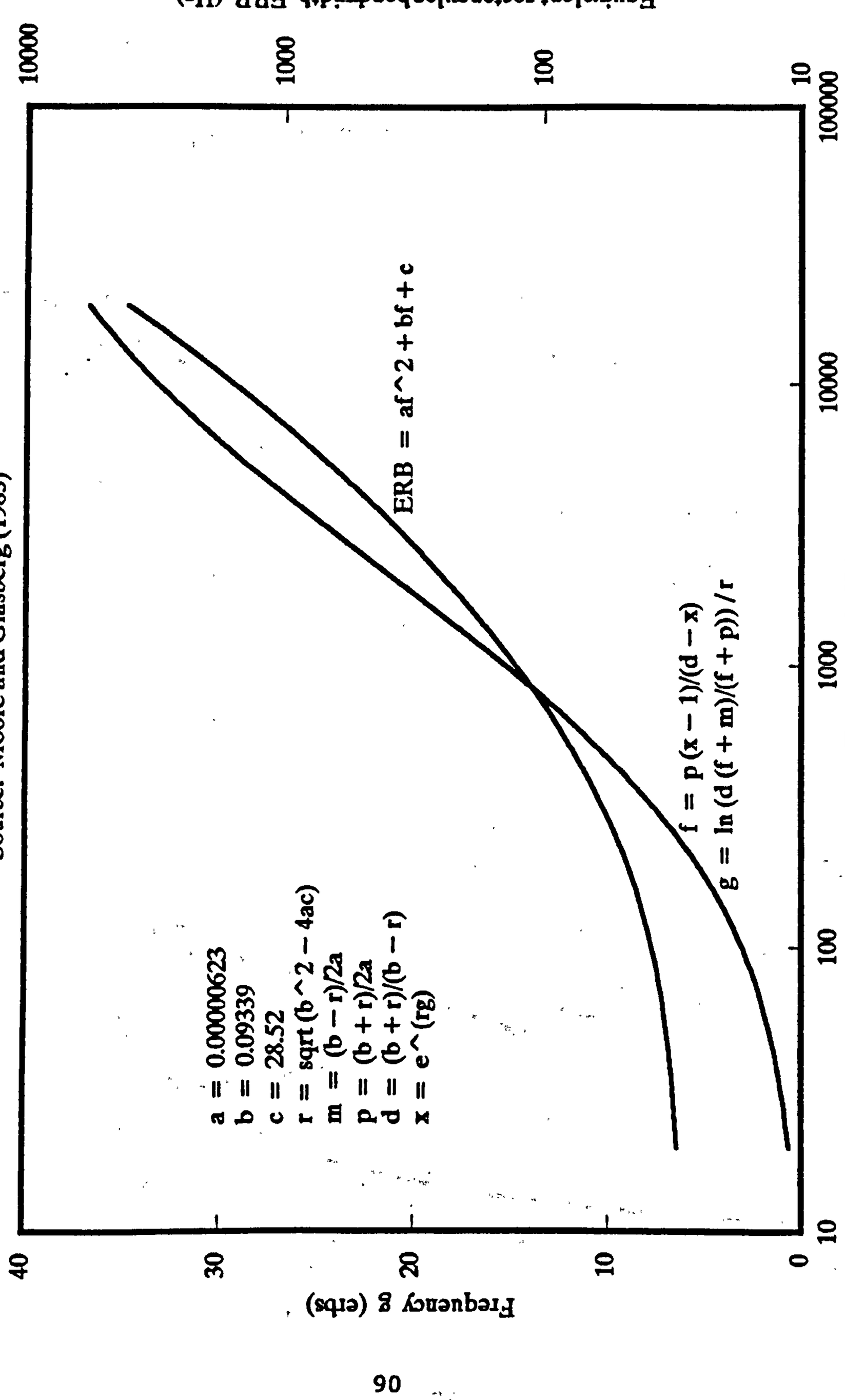
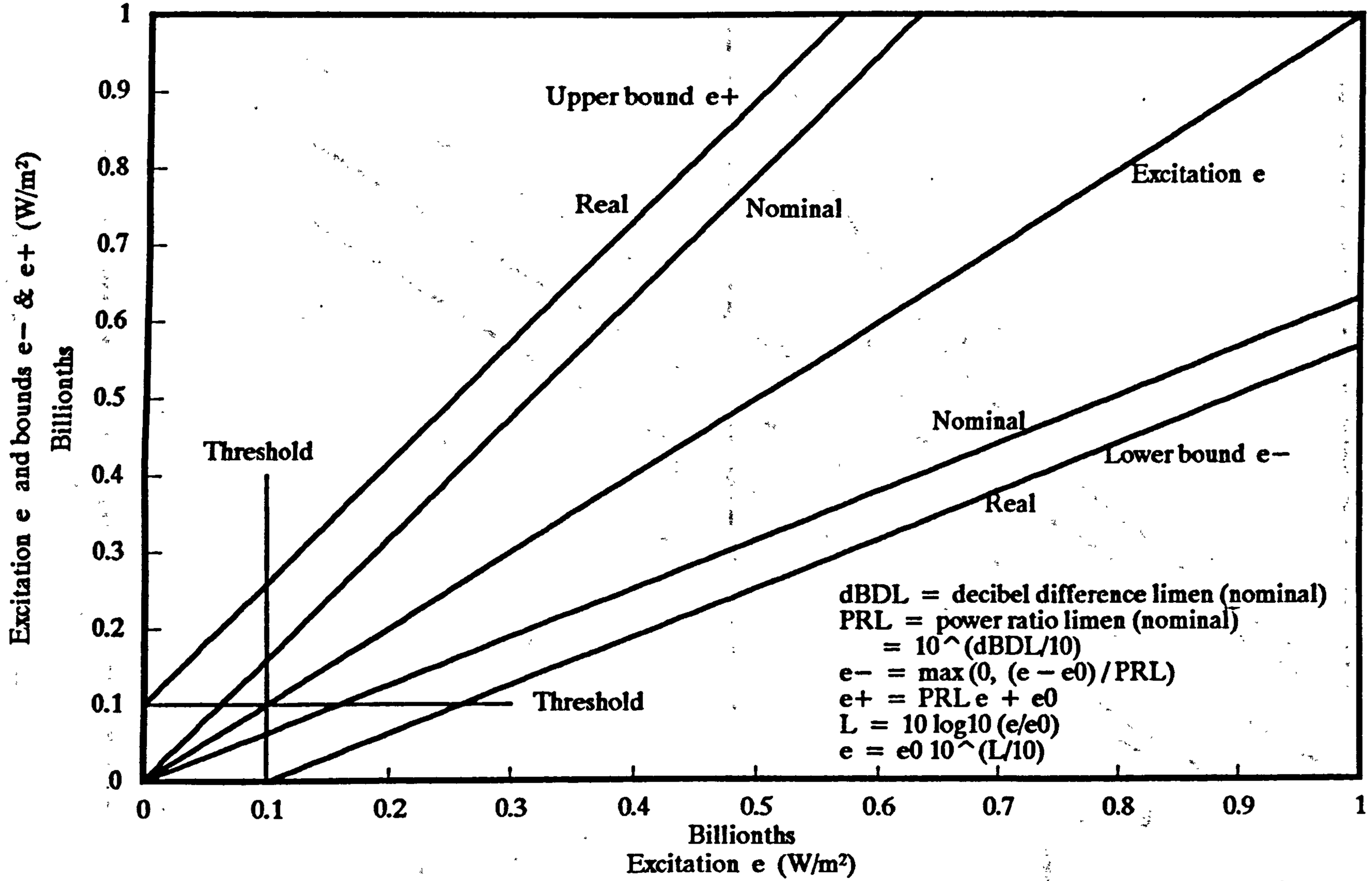

Excitation pattern from 5 pure tones
Excitation and upper and lower error bounds for difference limen dBDL = 3 dB
fin = frequency (Hz) of sound
f = frequency (Hz) of excitation
ERB = equivalent rectangular bandwidth
PA = power attenuation
a = 0.0000623
b = 0.09339
c = 28.52
t = 0.0001
r = sqrt (b^2 − 4ac)
m = (b − r)/2a
p = (b + r)/2a
d = p/m
x = e^(rg)
f = p (x − 1)/(d − x)
ERB = af^2 + bf + c
s = 4 |f − fin| /ERB
PA = (1 − t) (1 + s) e^−s + t
1
2
3
4
5
e+
e−
e
Threshold of hearing e0
Frequency g (erbs)
Excitation e (dB SPL)
0
10
20
30
40
50
60
70
80

Figure 3.2


Conversion between hertz and erbs
Source: Moore and Glasberg (1983)
ERB = af^2 + bf + c
a = 0.00000623
b = 0.09339
c = 28.52
r = sqrt(b^2 − 4ac)
m = (b − r)/2a
p = (b + r)/2a
d = (b + r)/(b − r)
x = e^(rg)
f = p(x − 1)/(d − x)
g = ln(d(f + m)/(f + p))/r
Equivalent rectangular bandwidth ERB (Hz)
Frequency f (Hz)
Frequency g (erbs)
10000
1000
100
10
40
30
20
10
0
The erb is used as a unit of frequency and is defined so that the ERB at any frequency is always one erb wide.
If f is a frequency expressed in Hz, then g is the same frequency expressed in erbs.

Error bounds near threshold

Threshold e0 = say 1 E−10 W/m² Difference limen dBDL = say 2 dB



Below threshold (e < e0, L < L0 = 0 dB SL) lower bound is e− = 0 since excitation can be unnoticeably removed.
With no excitation (e = 0 W/m²) upper bound is threshold (e+ = e0, L+ = L0 = 0 dB SL) since any more would be heard.

Figure 3.3

Figure 3.4

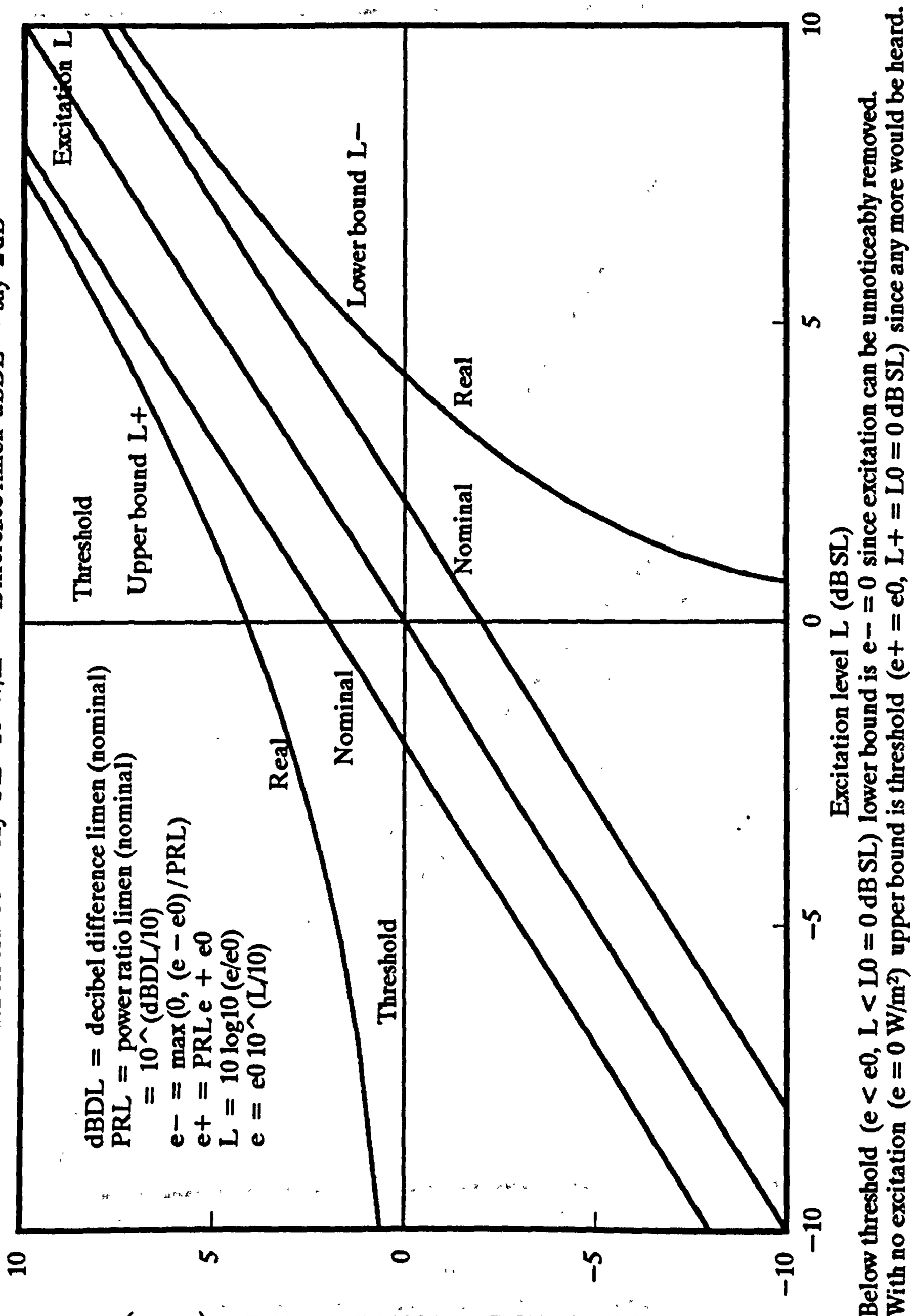
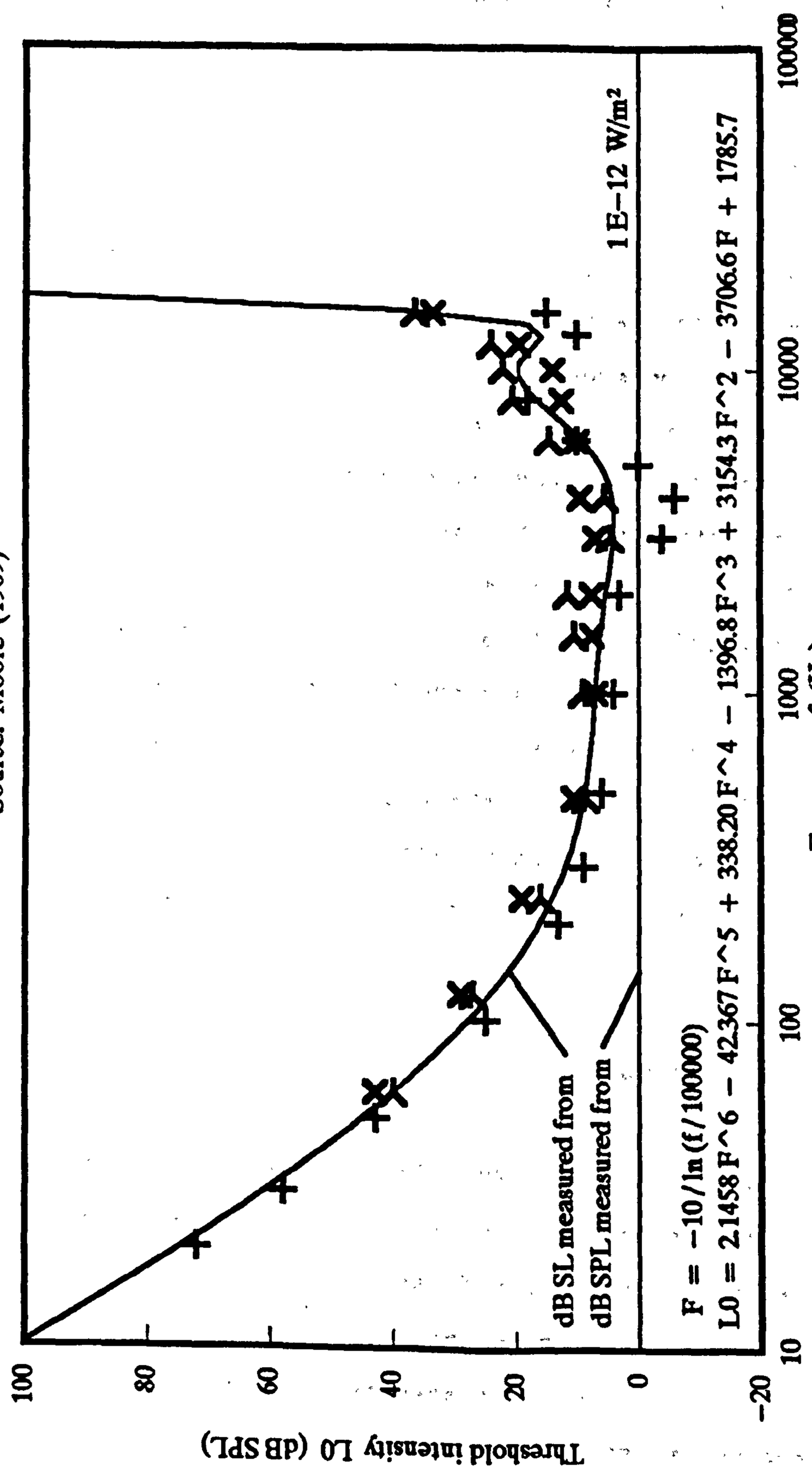

Error bounds near threshold
Threshold e0 = say 1 E−10 W/m²
Difference limen dBDL = say 2 dB
dBDL = decibel difference limen (nominal)
PRL = power ratio limen (nominal)
= 10^(dBDL/10)
e− = max(0, (e − e0) / PRL)
e+ = PRL e + e0
L = 10 log10 (e/e0)
e = e0 10^(L/10)
Excitation
Threshold
Upper bound L+
Lower bound L−
Real
Nominal
Real
Nominal
Threshold
10
5
0
−5
−10
Excitation level L (dB SL)
Excitation level L and bounds L− & L+ (dB SL)
Below threshold (e < e0, L < L0 = 0 dB SL) lower bound is e− = 0 since excitation can be unnoticeably removed.
With no excitation (e = 0 W/m²) upper bound is threshold (e+ = e0, L+ = L0 = 0 dB SL) since any more would be heard.

Figure 3.5


Hearing threshold and equation
Source: Moore (1989)
100000
1E−12 W/m²
10000
1000
100
Frequency f (Hz)
F = −10/ln(f/100000)
L0 = 21458 F^6 − 42367 F^5 + 338.20 F^4 − 1396.8 F^3 + 3154.3 F^2 − 3706.6 F + 1785.7
dB SL measured from
dB SPL measured from
100
80
60
40
20
0
−20
10
Threshold intensity L0 (dB SPL)
Fitted equation
p 48: 7 mm outside meatus
p 48: 3 mm inside meatus
p 53: position unspecified
Use of fewer significant figures gives surprising inaccuracy.

CHAPTER 4 SCHEME 1 - CARTESIAN PIANO TRANSFORM

4.1 Motivation

The cartesian piano transform (Dallas 1980, O'Hea 1987, Meyer 1992) is the simplest and problably the first mapping from scenes to sounds that comes to mind. The short definition of the piano transform is that the scene becomes the spectrogram, or some similar time-frequency representation, of the sound. The term piano transform arose because the scheme maps the scene y (or x or radial or circumferential) position to the piano keyboard (or some similar monotonic function of frequency), and the perpendicular scene position to time. If the transform maps the keyboard to the scene x or y direction then the transform is cartesian, and if to the radial or circumferential direction then it is polar.

4.2 Implementation

The cartesian piano transform was applied to two scenes represented by 175 rows and 320 columns of greyscale pixels, one natural and one artificial (Figures 4.1 & 4.2). The keyboard was taken to represent vertical position, so the sound spectrum was specified at 175 frequencies equally spaced (when expressed in erbs,

Figure 3.2) in the audible range.

320 sound spectrums were calculated, one for each column of pixels, and subjected to the TPT.

The program used was hear.c, described above.

4.3 Results

4.3.1 Effect of intensity difference limen

The three figures 4.3 to 4.5 are the result of applying the theoretical performance test to the cartesian piano transform, in mono mode, with the intensity difference limen taken as 1, 2 and 3 dB respectively. As expected, the performance deteriorates rapidly with increasing intensity difference limen. The intensity difference limen is of course not something that can in reality be chosen at will, and for that reason is not included in the data file heardata.t. For all subsequent work, unless otherwise stated, the intensity difference limen was set at 2 dB, as in Figure 4.6.

4.3.2 Effect of ear distribution

Stereo mode, in which negative and positive values of elements of X are sounded in opposite ears, gave mild improvement seen in Figure 4.7 as compared to Figure 4.4.

4.3.3 Conclusion

This feeble treatment of the cartesian piano transform doesn't do it justice. In particular, it was only subjected to the theoretical performance test and never actually sounded. Why? A similar scheme, using artificial shapes, was sounded by O'Hea (1987), who reported a mushy sound like trying to convey two notes on the piano by playing all the notes in between. Nevertheless, more interesting ideas came along as a result of the thought processes going on during the course of the feeble treatment.

The first idea that came along was from the field of image or data compression. It was naively thought that since the ears are an information bottleneck, it would be better to have a scheme that sounded uncorrelated variables instead of pixel brightnesses. This resulted in the work on the cosine transform reported below.

The second idea was to improve the piano transform by

spatial high-pass filtering intended to crispen the edges before further processing into sound. At the same time, however, many other improvements suggested themselves simply as a result of the decision to skip the TPT and think about mappings in their proper context, namely a functioning optophone. One of the results was a modified piano transform incorporating not only high-pass filtering but also foveation, colour discrimination and size invariance. This is reported in Chapter 7.

Figure 4.1 Bike.Q test image

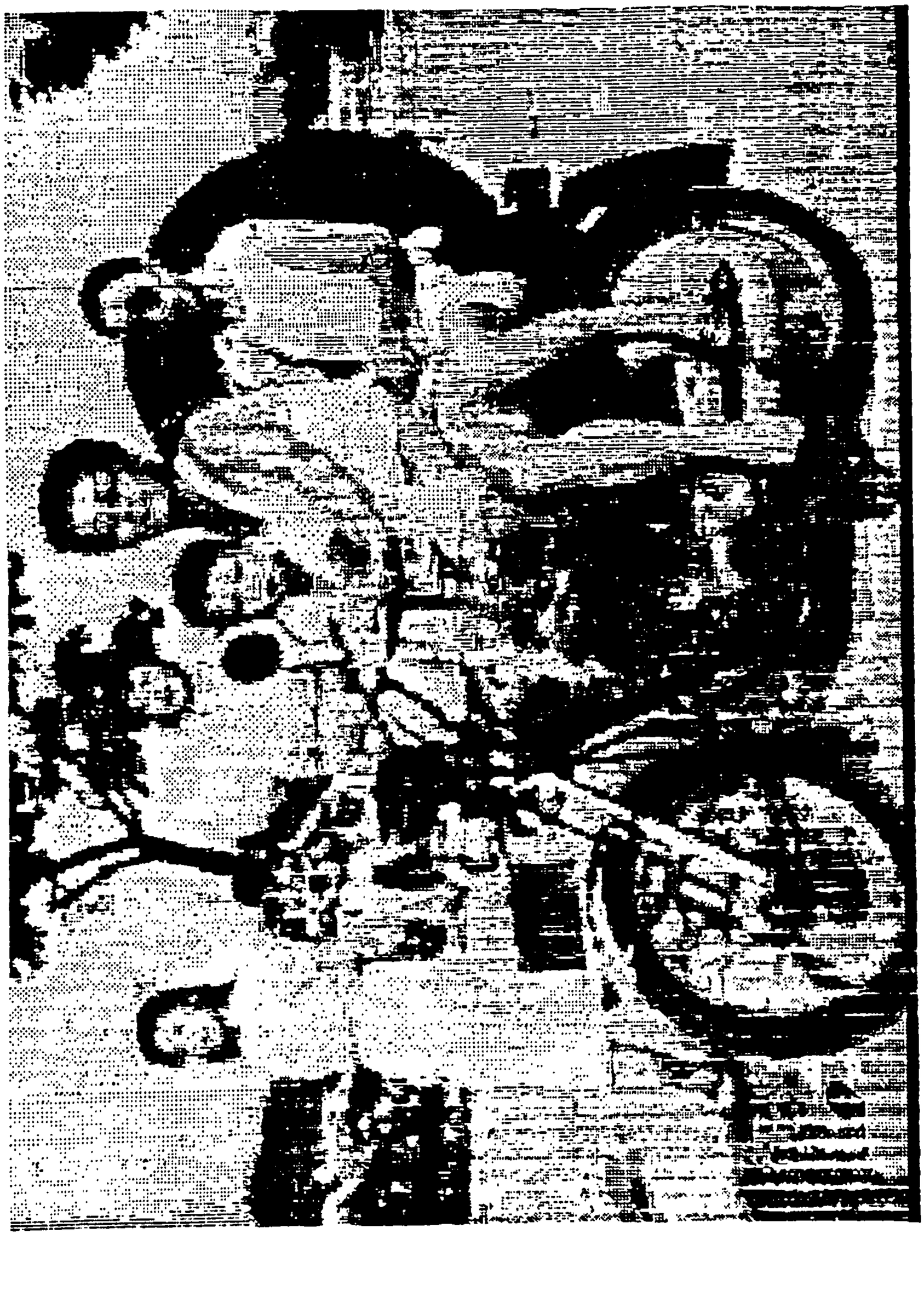


Figure 4.2
Stripe.Q testimage

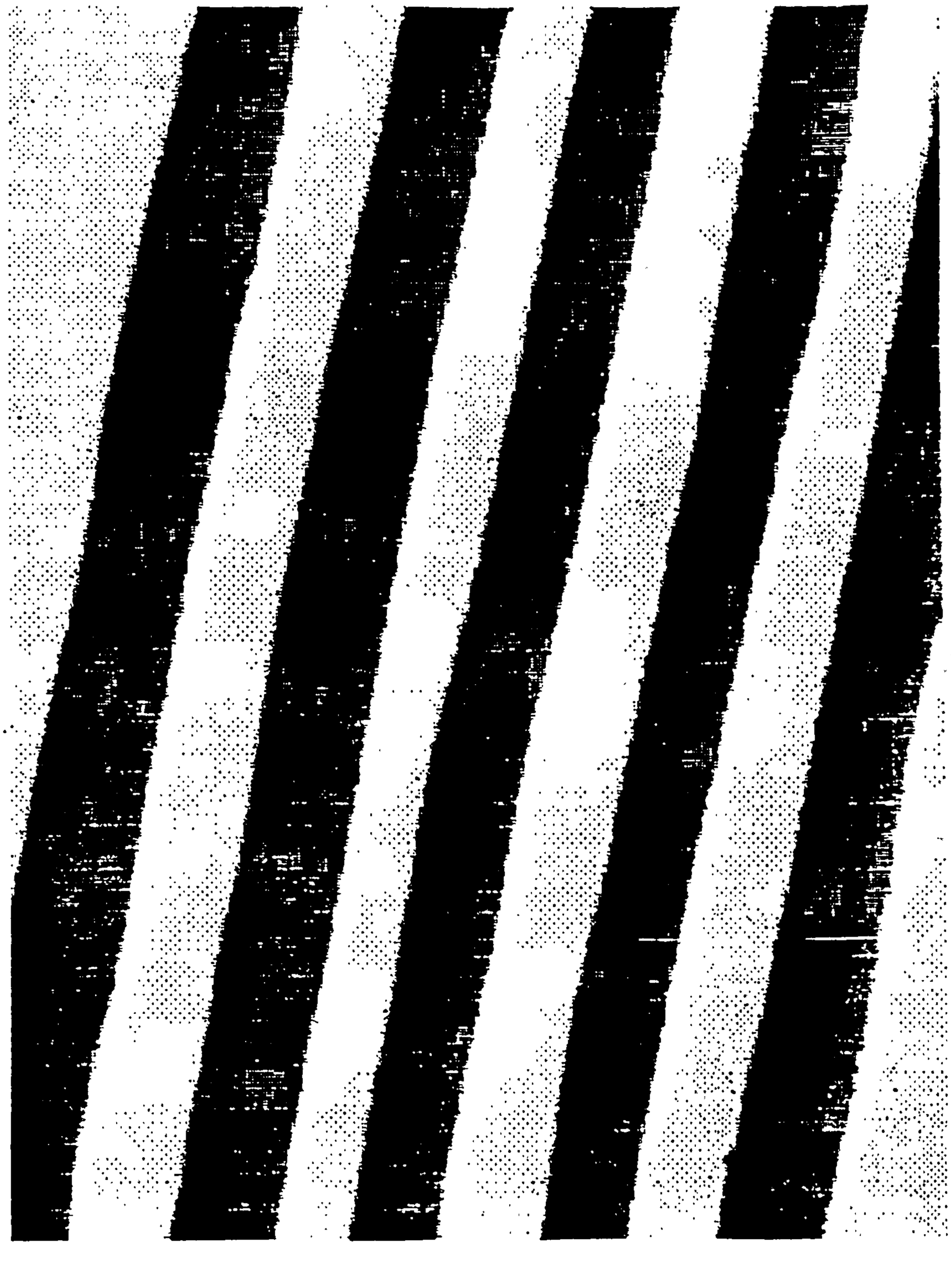
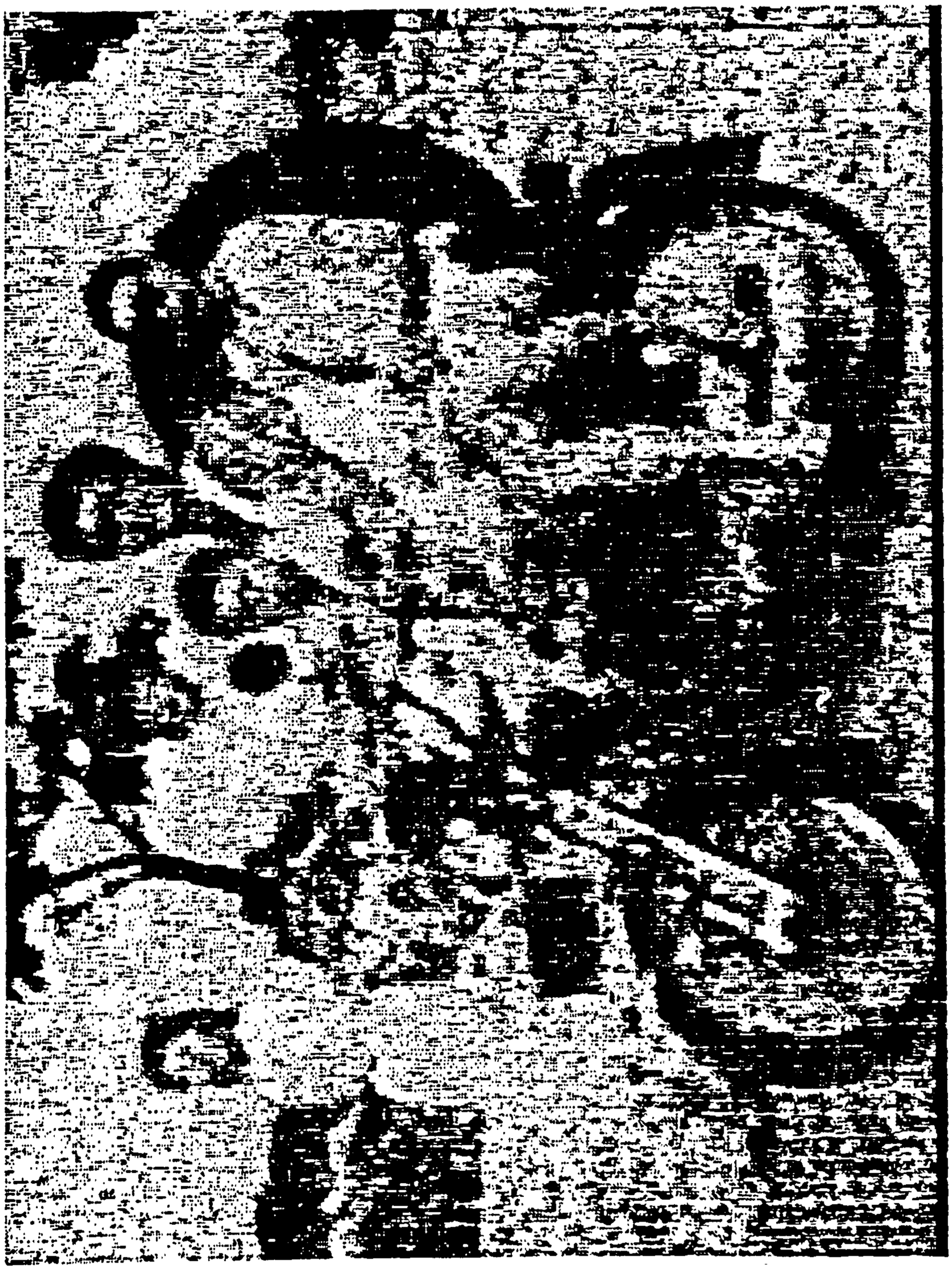


Figure 4.3



Minimum dBdL = 1 dB

Figure 4.4

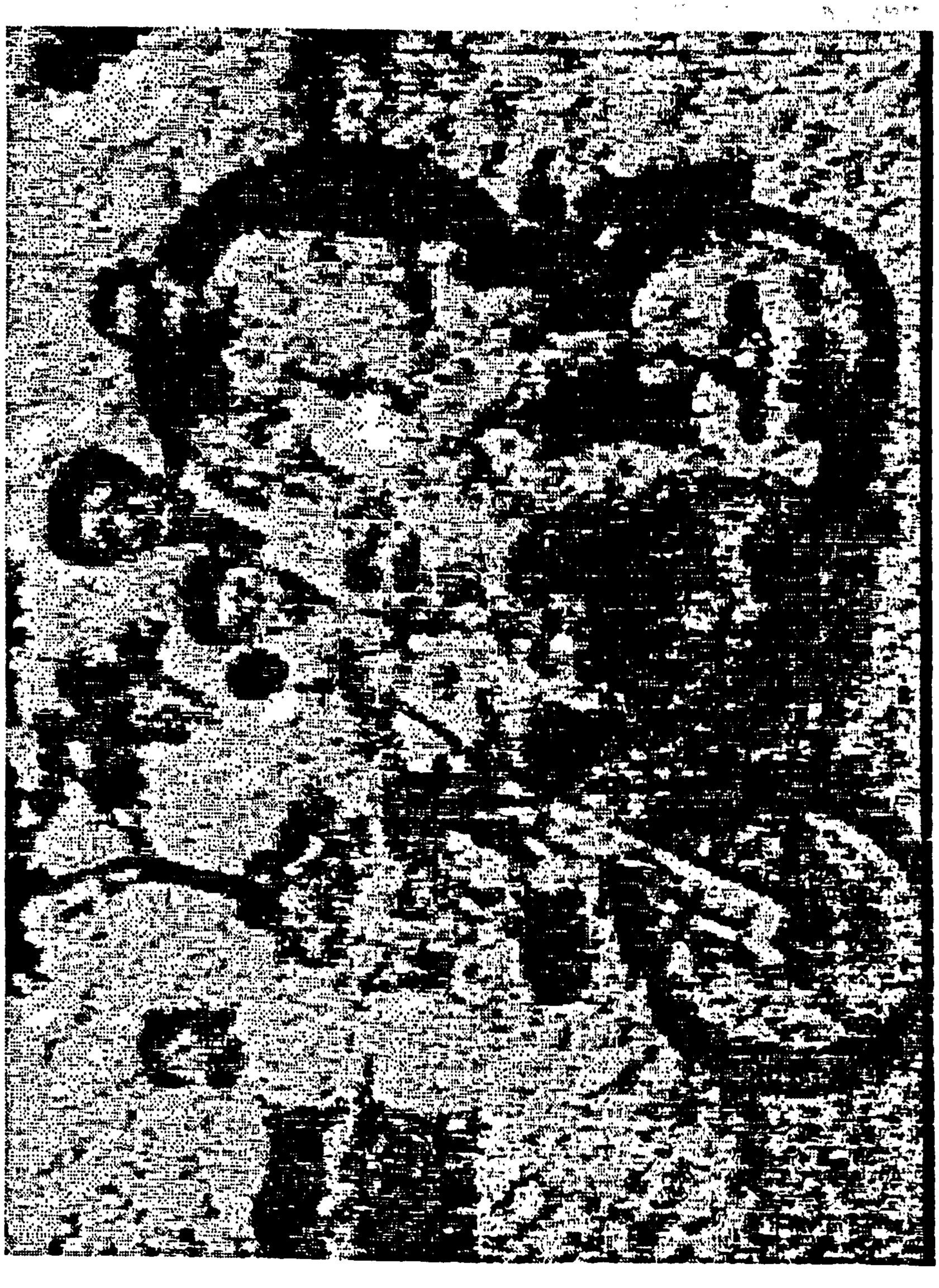


Figure 4.5

minimum detectable = 3dB

Figure 4.6

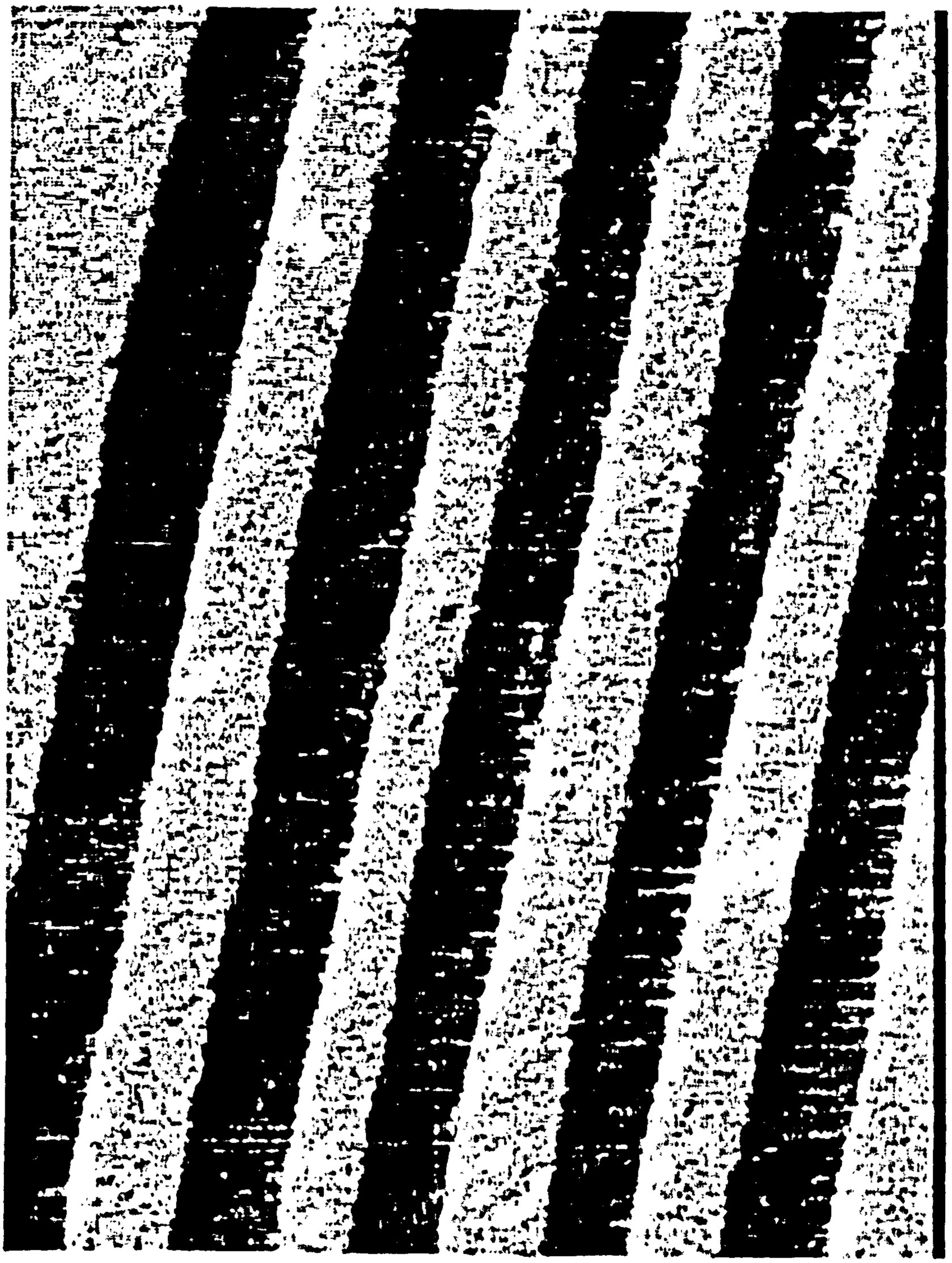
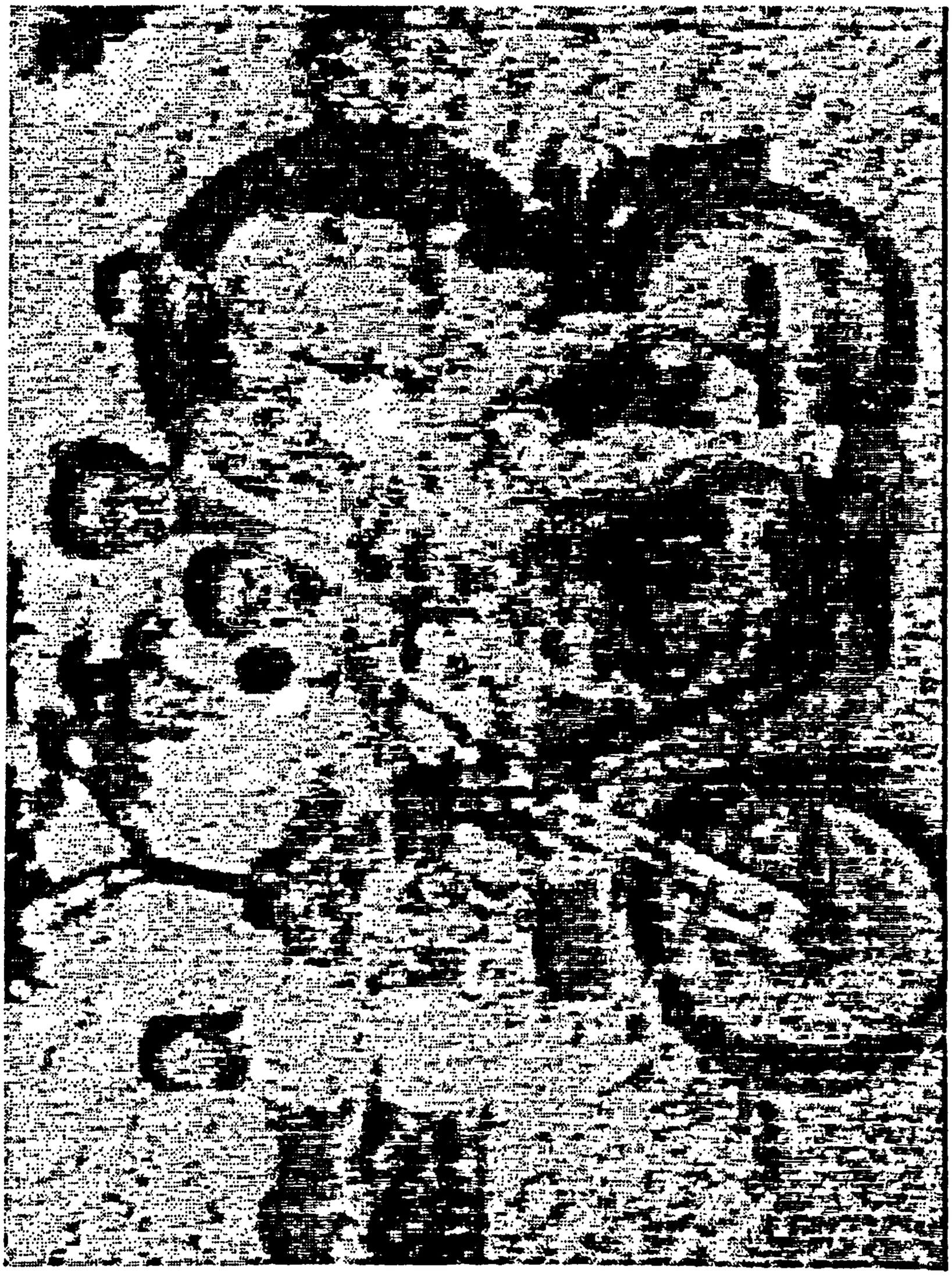


Figure 4.7



CHAPTER 5 SCHEME 2 - COSINE TRANSFORM (BOUSTROPHEDON)

5.1 Motivation

The notion of slowing down a television signal to auditory frequencies (without necessarily including the line and frame synchronising pulses) is tempting because of its simplicity.

In the normal method of scanning a scene, the sensor starts at the top left corner and scans the scene horizontally line by line, ending in the bottom right corner, as in reading English. In a typical scene, one line of the scan is very similar to the previous and following lines. The signal is therefore nearly periodic, at least locally. If slowed down to auditory frequencies, a scene would sound like a sound sequence of a few seconds, depending on the scan resolution. Unfortunately, many different scenes could produce the same sound, since the phase information in the signal is largely unrecognised by the ear.

This difficulty can be overcome by scanning one line from left to right and the next from right to left. (Before writing became very common, this is the way the ancient Greeks wrote, and they called it boustrophedon, after the way a bullock ploughs a field.) The resulting signal is

not only nearly periodic (with a period of two lines) but also symmetrical. It therefore contains no phase information and can be fully reconstructed from the amplitudes of the frequencies present. The ear detects squared amplitudes, meaning that positive and negative amplitudes are indistinguishable.

Two ways of overcoming this are as follows. Either a constant can be added to each amplitude, making them all positive, or the positive ones can be played to one ear and the negative ones to the other. Either method involves considerably more processing (a cosine transform followed by an inverse cosine transform of the rectified amplitudes) than promised by the idea at first sight.

These two schemes were assessed as follows.

1 Digitised scene submitted to line-by-line cosine transform.

2 Resulting amplitudes rectified according to scheme.

3 Amplitudes corrupted by normal hearing imperfections (masking, finite difference limens)

4 Scene reconstituted from corrupted amplitudes

by inverse of 2 and 1 and compared with original.

Step 1 is a standard operation. Step 2 is simple and described above. Step 3 is a complicated operation derived specially for this study, and is described in detail in Chapter 3.

5.2 Results

Original and reconstructed scenes using the scheme adding a constant to all amplitudes are shown in Figures 4.1 and 5.1 and in Figures 4.2 and 5.2. Bu contrast, Figure 5.3 shows a scene corrupted and reconstructed using the scheme sending amplitudes of different signs to each ear. As expected, the scheme sending amplitudes of different signs to each ear performs better than simply adding a constant to all amplitudes. However, several points must be borne in mind.

First, the assumption that completely different sounds can usefully be sent to each ear is invalid: slight differences in loudness and timing are useful in localising sounds (see references under **PSYCHOPHYSICS - HEARING - BINAURAL EFFECTS** and **PSYCHOPHYSICS - HEARING - LOCALISATION**), but if the differences are too great then attention is directed to only one ear, as in listening to

the telephone (PSYCHOPHYSICS - HEARING - ATTENTION).

Second, a sound wave consisting of only cosine waves of positive amplitude has a sharp peak at the start of every period. In practice, for reproduction through loudspeakers or earphones, the phase of each frequency would have to be shifted differently so as to remove this peak.

The third point concerns the method of assessment, the TPT, which although nominally objective is ultimately visual and subjective. This is quite proper, since it automatically deals with such things as selective sensitivity to errors at different spatial frequencies (Mannos & Sakrison 1974). The one thing missing is foveation. As argued in O'Hea (1987), because hearing is much slower than sight in terms of information rate, in order to provide a useful resolution without resorting to tunnel vision, foveation is needed even more in an optophone than in natural vision. Visual assessment of a foveated scene is difficult, however, since the eye naturally wanders off the fovea and finds the area of interest to be blurred.

Note that the boustrophedon could be made radial, thus producing foveation of a sort.

Fourth comes the time dimension. In corrupting the

scene, it is assumed that the sound is changing slowly enough for transient effects such as forward and backward masking to be ignored. Thus the scene is not properly corrupted perpendicularly to the scan lines, and it is not possible by this method to determine a suitable number of scan lines or, equivalently, the time required to sound a whole scene.

Last, and most fundamental, is the explanation of the poor performance of the cosine transform in this context. It is known (Pratt 1978) that for natural scenes the cosine is a highly decorrelating transform. This means that there is not much connection between the amplitudes of adjacent frequencies. Because of simultaneous masking (Moore 1989), only the locally dominant frequency can be heard, and the others must be assumed either to be zero at startup or unchanged if newly masked.

Figure 5.1

Figure 5.2

Figure 5.3

CHAPTER 6 PRINCIPAL COMPONENT ANALYSIS

6.1 Motivation

The statistical analysis of scenes and sounds was prompted by the prospect of an automatically generated mapping as described in section 2.4. To recapitulate that section, what is required is that a scene, or a sound, be represented by a vector of numbers expressed in units of one difference limen and of known mean, variance and covariance.

6.2 Statistics of scenes

6.2.1 Scenes as pixel brightnesses

The standard way of expressing a scene is an array of pixel brightnesses. With the rows or columns placed end to end, these become a vector. It is usual to model a scene as a two-dimensional Markov process (Pratt 1978), with adjacent pixel correlation ρ of something like 0.95, of pixels two pixels apart ρ^2, and so on. The more detailed the scene, the lower the value of ρ, but a representative value is sufficient for our purpose.

Brightness, grey-scale value, illuminance, illumination, intensity, irradiance, irradiation, lightness, luminance, luminosity, luminous flux, radiance, radiant energy, radiation, are all terms at times used confusingly. I will try to avoid as many of these as possible. Luckily, many sentences containing such words only refer to a qualitative dark-to-light scale of no particular definition. The only term whose meaning might initially appear self evident is "amount of light", even though it usually means "amount of light per unit area (or subtended solid angle) per unit time".

The brightness scale that interests us is the scale with a constant difference limen. This scale may then be multiplied by a constant so that the difference limen is equal to 1.

The brightness difference limen is constant over a very wide range when expressed as the just noticeable percentage difference in the amount of light (yes, per unit solid angle subtended at the eyes and per unit time) coming from two adjacent patches (Pratt 1978). This difference is about 2%. The required brightness scale is therefore a logarithm or some similarly curved function of the amount of light. Note in contrast the value of the intensity difference limen in sound - around 1 dB or 26% (Moore 1989).

It is reasonable to assume that this nonlinearity is taken into account in the mapping of incident light to pixel value inherent in a system for producing digital images, and in the mapping from pixel value to radiant light inherent in a system for displaying digital images. This is the case for the present system, as shown in the upper third of Figure 6.1. These sixteen different greys are the nearest the system can get to a continuous grey scale. The point to note is that the subjective difference between adjacent greys does not particularly increase from right to left or from left to right. Some brightness steps do appear more pronounced than others; this is a feature of the printing software and is not the case on the screen.

For this reason, the grey scale, or pixel value in the case of black and white scenes, is taken here as the required brightness scale needing only a multiplying constant to become a difference-limen (DL) scale.

In order to convert the pixel value to units of one difference limen, it is necessary to know how many different values can be distinguished. Real scenes can be satisfactorily displayed with 64 grey levels. Scenes with areas of smoothly varying brightness, such as a face, show unsatisfactory contouring when 32 grey levels are used, while scenes with much detail can be satisfactorily displayed with only 16 grey levels (Pratt

1978, Gonzalez & Wintz 1987). The multiplying factor from grey scale to DL scale is therefore equal to the required number of grey levels divided by the original number of grey levels.

Let us refer to this candidate psychophysical representation as CPR A.

6.2.2 Effect of spatial separation

Unfortunately, the above procedure is based only on adjacent greys and does not model the increase of difference limen with separation. It is very difficult to compare the brightness of patches that are not adjacent.

Consider the central third of Figure 6.1. Here the grey scale is divided into only five values. Number the panels one to five and consider the central panel, n° 3. Compare two patches, patch A in the left half of panel 3, near panel 2, and patch B in the right half of panel 3, near panel 4. It is a well known effect, known as the Mach-band effect, that patch A looks brighter than patch B, even though the amount of light coming from each is physically the same. One can convince oneself of this by covering up panels 2 and 4, whereupon the difference disappears.

Similarly, panels 2 and 5 in the bottom third of Figure 6.1 are the same grey, but panel 2 looks brighter because of its dark surround. The strength of the effect varies with subtended angle, as can be explored by placing Figure 6.1 at the far side of the room and looking at it from different distances.

One way of explaining the Mach-band effect is to say that low spatial frequencies are relatively less important to human vision than higher frequencies. This is an easily acceptable statement if one considers the lowest and next lowest spatial frequencies as compared to some much higher frequency. Take any natural scene. Corrupt the lowest spatial frequency (sometimes called the DC component). All that results is an overall shift in brightness, which in a complete scene, as opposed to a picture in a frame with a surround that does not change, is simply not noticeable. Similarly, corrupting the next lowest frequency results in a vague lightening of one half of the scene and darkening of the other half, again hardly noticeable. On the other hand, corrupting a much higher frequency by the same amount results in quite objectionable stripes across the scene.

The relative importance of different spatial frequencies has been studied systematically. Mannos & Sakrison (1974) produced a frequency weighting function which peaks arbitrarily at 1 at a frequency of 8 cycles per

degree, falling to 0.05 at frequency 0. Remembering what was said above about the difference between scenes and pictures, we might make the weighting function fall to zero at frequency zero. Both functions are shown in Figure 6.3, but the curves are indistinguishable in Figure 6.2.

Note in particular the different nature of each side of the graph. In Figure 6.3, based on wavelength or feature size, the left side of the graph merely represents the best the eye can do under the various physical and physiological constraints present in the eye and in the nature of light, while the right side reflects the relative importance of different scales in the scene. This is the side of current interest.

A standard demonstration of the variation of sensitivity of the eye to spatial frequencies is Figure 6.4. The poor quality is due to the figure being produced by dot-matrix printer, and some indulgence is requested. The figure is formed by sinusoidal swings between black and white along the top of the figure, diminishing progressively in amplitude to constant grey along the bottom. The wavelength is about 8 mm at the centre of the figure. To place this wavelength at the peak sensitivity reported by Mannos & Sakrison, the figure should be viewed at a distance of around 4 m. At that distance, all of the far left of the figure appears a

constant grey, the grey area being narrower towards the top of the figure. Note that this demonstration concerns the "best the eye can do" side of the curve that does not at the moment interest us.

Striking evidence of the lack of importance attached by the eye to longer spatial wavelenths is illustrated in Figure 6.5 (Schroeder 1983), in which all the high frequencies of the original photograph have been removed and replaced by the lines that form the figure, but in a clever way which leaves the lower frequencies intact. However, it is not possible to gain access to these lower frequencies. The information they contain can be seen if the figure is placed so far away that they cease to be low frequencies. Another way is to place the figure some centimetres behind frosted glass, thus removing the spurious high frequencies contained in the lines. Only by doing one of these things is it possible to tell that the man is wearing glasses.

Now an optophone, as does any other piece of equipment designed to capture scenes, will have its own physically limited resolution. In terms of cycles per degree, this may even be variable, either electronically or by means of a zoom lens. Thus both the position of the peak in Figures 6.2 and 6.3 and the entire "best the eye can do" side of the curve are for the present purpose of little interest, since they refer specifically to human vision.

We are therefore free to use similar curves peaking at whatever spatial frequency is most appropriate.

6.2.3 Scenes as filtered pixel brightnesses

Two new candidate psychophysical representations (vectors with DL scales) now suggest themselves. One is to express the scene as a vector of spatial-frequency coefficients by means of a Fourier or cosine transform, and weight the coefficients according to the Mannos & Sakrison curve. Call this CPR B. The other is to add a further step, namely reconvert the weighted coefficients into a filtered version of the scene and use the resulting pixel values. Call this CPR C. The question is whether either of these is a true PR.

Consider first CPR B, consisting of Fourier transform coefficients. Take a scene transformed from 256 x 256 pixels to 256 x 128 Fourier magnitudes and 256 x 128 Fourier phases. If the magnitudes are numbered according to spatial frequency, the numbering goes from -128 to 128 in one direction and 0 to 128 in the other, thus covering all orientations. The frequencies are from 0 to 128 cycles/scene.

Unfortunately, there are in vision the visual equivalent of critical bands in hearing (Julesz 1971, p67), and

these visual critical bands are over an octave wide. Suppose that the scene, or part of it, has a strong component at 70 cycles/scene in some orientation. The existence of critical bands of more than an octave means that weaker components from 50 to 100 cycles/scene in the same orientation are masked, the nearer the component to 70 cycles/scene the greater the masking effect, in the same way as auditory masking was described when discussing the TPT. This means that the difference limen of one number in CPR B depends on the size of other numbers, and CPR B is not a true PR.

CPR C is more similar to CPR A than CPR B is. CPR A didn't work because whether a change in the value of one of the numbers was noticeable depended on what other numbers were changed at the same time. In particular, if a sinusoidal change were applied, then the height of the just noticeable sinusoid increased with the sinusoid wavelength. Apply a similar just noticeable sinusoidal change to CPR C. Does its magnitude still depend on its wavelength?

The filtering involved in creating CPR C multiplies the magnitudes of the sinusoids constituting a scene, and of sinusoidal changes to it, by the weighting of the Mannos & Sakrison curve. The result is intended to be a set of sinusoids of equal visual importance. If CPR C is a filtered scene consisting of sinusoids of equal visual

importance, then changes of the same magnitude will be equally noticeable regardless of scale, and the smallest noticeable change (the difference limen) will be the same regardless of scale. We will therefore take CPR C to be a true PR.

6.2.4 Statistics of scene PR

Suppose that as a first stage an optophone captures a 512 x 512 pixel unfiltered scene subtending an angle of 120°. As discussed above, assume the peak spatial-frequency sensitivity to be 1 cycle/degree instead of 8 cycles/degree, since 8 cycles per degree is even less than the pixel separation.

For simplicity, and because distant pixels are hardly correlated, take a 16 x 16 pixel picture in the scene. This is helped by taking an adjacent-pixel correlation ρ of 0.9. Let the vector representation of this picture be a 256 element column vector p with the first row of the scene as the first 16 elements, the second row as the second 16 elements, and so on. The covariance matrix of the vector is a 256 x 256 element matrix C, with c_{ij} as the covariance of pixels p_i and p_j. Thus C_{ii} is the variance of pixel p_i. Since all pixels have the same variance, C is equal apart from a scale factor to the correlation matrix R:

$$C = \sigma^2 R \tag{1}$$

where σ^2 is the pixel variance.

Figure 6.6 has six panels, numbered 1 to 3 along the top half and 4 to 6 along the bottom half. The top left panel (panel 1) is a representation of R (or C), with the black diagonal representing the highest correlation (= 1) and the other elements given by

$$\rho_{ij} = \rho^{\sqrt{(row_i - row_j)^2 + (col_i - col_j)^2}} \tag{2}$$

where ρ is the adjacent-pixel correlation as discussed earlier.

Panel 2 of the figure is a rearrangement of R so that each 16 x 16 submatrix is a map of the scene itself and shows the correlation of each pixel with the pixel shown darkest.

Panel 3 is an extension at the same scale of any of the squares of panel 2 beyond the picture boundaries.

Now consider the correlation properties of the PR consisting of the filtered scene. Let F be a filter similar to the Mannos & Sakrison filter described above, so that the filtered scene p' is given by

$$p' = F p \tag{3}$$

Then the covariance matrix C' of p' is (Pratt 1978)

$$C' = F C F^{*T} \qquad (4)$$

where superscript * denotes complex conjugation and T transposition. Since F is real, the * need not concern us.

The bottom three panels of Figure 6.6 show the covariance matrix C' in the same three ways as C in the top three panels. Note that correlation between pixels now stretches much less far, as expected.

C' was in fact not calculated by two 256 x 256 matrix multiplications as implied by equation (4), since many of the numbers in C' are the same. Instead, panel 6 was obtained directly from panel 3 by the process illustrated in Figure 6.7. Panel 1 is the same as panel 3 of Figure 6.6, namely the correlation of a near-central pixel of the 512 x 512 scene to all the others. It turns out that the same effect as equation (4) can be obtained by filtering this correlation image by the two-dimensional circularly symmetric version of the square of the spatial frequency sensitivity curve. This spatial filter is shown in panel 4 (the small panel), with the frequencies numbered 0 to 128 cycles/scene starting in the top left corner. The calculation was done using two-dimensional forward and inverse fast Fourier transforms. For information, panel 3 shows the same filter on a

wavelength instead of a frequency base, to the same scale as the 512 x 512 scene in panel 1. Note the peak at only a few pixels distance from the origin.

6.2.5 Decorrelation of scene PR

Karhunen-Loève transforms P and P' for both p and p' have been derived from their covariance matrixes (the English plural is deliberate) C and C'. Each row of P is an eigenvector of the covariance matrix C (Gonzalez & Wintz 1987). The inverse P^{-1} of P, used to reconstruct a scene vector p from an uncorrelated vector v, is equal to the transpose P^T of P, so from equation (3) of Chapter 2

$$p = P^T v + m \tag{5}$$

The vector v can then be understood as containing a list of weights multiplying the columns of P^T, each of which is a basis function for constructing a scene. The bottom two panels in Figure 6.8 show P^T and P'^T. In the same way as panel 2 of Figure 6.6 is a rearranged version of the covariance matrix in panel 1, so the columns of P^T and P'^T, the basis functions, have been rearranged into the 16 x 16 pixel squares of the top two panels of Figure 6.8.

These square basis functions are shown enlarged in Figures 6.9 and 6.10, one from P^T and one from P'^T. They

can be readily understood as building blocks for making 16 x 16 pixel scenes. The basis functions derived from the filtered and unfiltered scene statistics are remarkably similar, so much so that having forgotten to make a note I can't even tell which is which.

The big difference is between the variances of the elements of v and those of v'. This is shown in Figure 6.11. The main conclusion is that p' is already largely decorrelated and there is much more to be lost by discarding the higher-numbered coefficients of v' than is the case with v.

6.3 Statistics of full-spectrum sounds

6.3.1 Notional auditory time-frequency filter

A psychophysical representation (PR) for steady sounds has been derived in section 3.2 dealing with the theoretical performance test. This PR turned out to be excitation levels on an erb abscissa, excitation levels being taken instead of sound intensity levels so as to deal with masking.

Sounds in general are not steady, and masking takes place in both the frequency direction and the time direction.

A general PR is required based on a time-frequency plane in much the same way as a spectrogram is. An attempt was made to derive a two-dimensional masking pattern reaching forwards and backwards in time as well as into adjacent frequencies. The object was to obtain something like a two-dimensional impulse response which could be applied at any location in the time-frequency representation of a sound to determine the additive contribution to the overall time-varying excitation pattern by the sound power sampled at that location.

Researchers have studied masking of short probe tones by many types of sound pattern, including pure tones, chirps and noise. (References are given under **PSYCHOPHYSICS - HEARING - MASKING - Temporal.**) Probes have been placed above and below the masker in frequency, before and after the masker in time, and in both frequency and time gaps in the masker.

All the maskers have been extended in frequency, time, or both, since because of the uncertainty principle (Gabor 1946) it is not possible to have a point-like sound in the time-frequency plane. This in itself would not seem to preclude the derivation of the time-frequency masking pattern of a notional point-like sound, provided that, when used additively, the pattern could reproduce the effects measured with extended maskers.

In the event, it was not possible to reproduce some of the measured effects or to reconcile others. For example, Penner (1979) showed that forward masking diminishes more rapidly the louder the masker, making forward masking not additive. Houtgast (1977) found that forward masking by two pure tones can be less than by one alone. This can be construed as being additive if the masking pattern around the tip of a stopped pure tone is negative in a region after the tip in time and lower in frequency. Kohlrausch (1988), on the other hand, found masking between down-chirps to be considerably less than masking between otherwise similar up-chirps. If masking were subdivisible and additive, then on the basis of Houtgast one would expect the opposite effect in Kohlrausch, since between up-chirps the probe is after and below much more masker than between down-chirps.

Faced with this evidence, a sensible compromise approach is to take a point-sound masking pattern such that when summed along a pure tone it reproduces the simultaneous masking patterns of Moore & Glasberg (1983) and when summed along a gap in a noise it reproduces the temporal masking patterns of Moore et al (1988). It is not necessary to vary the shape of the point-sound masking pattern to deal with transient sounds such as clicks with a wide spread of frequencies if the pattern is only used on valid time-frequency representations of sounds which have the proper spread built in.

The point-sound masking pattern can be considered an impulse response or point-spread function. Reversed in time and frequency this becomes a filter. The function is given by the equation

$$W = W_t W_f \tag{6}$$

where

$$W_t = W_{t1} + W_{t2} \tag{7}$$

$$W_{t1} = (1 - v)\ (1 + p_1|t_{out}-t_{in}|)\ e^{-p_1|t_{out}-t_{in}|} \tag{8}$$

$$W_{t2} = v\ (1 + p_2|t_{out}-t_{in}|)\ e^{-p_2|t_{out}-t_{in}|} \tag{9}$$

$$v = 0.0001 \tag{10}$$

$$p_1 = \begin{cases} 2/t_1^- & t_{out} > t_{in} \\ 2/t_1^+ & t_{out} < t_{in} \end{cases} \tag{11}$$

$$p_2 = \begin{cases} 2/t_2^- & t_{out} > t_{in} \\ 2/t_2^+ & t_{out} < t_{in} \end{cases} \tag{12}$$

$$t_1^- = 0.006\ s \tag{13}$$

$$t_1^+ = 0.003\ s \tag{14}$$

$$t_2^- = 0.030\ s \tag{15}$$

$$t_2^+ = 0.015\ s \tag{16}$$

and

$$W_f = (1 - w)\ (1 + q|f_{out}-f_{in}|)\ e^{-q|f_{out}-f_{in}|} + w \tag{17}$$

$$w = 0.0001 \tag{18}$$

$$q = 4/ERB \tag{19}$$

$$ERB = af_{out}^2 + bf_{out} + c \tag{20}$$

$$a = 0.00000623 \tag{21}$$

$$b = 0.09339 \tag{22}$$

$$c = 28.52 \tag{23}$$

In the above equations, subscripts in and out refer to sound and excitation, subscripts 1 and 2 refer to tip and skirt, and superscripts - and + refer to backward and forward masking respectively.

The equations for W are dimensionless and refer to ratios

of sound power. Note that W_f is none other than the power attenuation PA producing the steady excitation pattern of Figure 3.1 (Moore & Glasberg 1983), while W_t is a similarly derived time window for broad-band sounds (Moore et al 1988).

It is easy to show that

$$\int_{-\infty}^{\infty} W \, dt = k_t \, W_f \tag{24}$$

and

$$\int_{0}^{F} W \, df = k_f \, W_t \tag{25}$$

where F is some inaudibly high frequency and k_t and k_f are constants. Thus, to within multiplicative constants required to standardise the peaks at 1, this notional auditory time-frequency filter will simulate both the steady auditory filter of Moore & Glasberg (1983) and the auditory time window of Moore et al (1988).

Figures 6.12 and 6.13 show the notional point filter resulting from the above equations, the tip in terms of power ratio and the skirts in terms of decibels, scaled to peak at 1. In these figures, f_{in} is the ordinate and

f_{out} is 1000 Hz. Similarly, t_{in} is the abscissa and t_{out} is 0.

Figures 6.14 and 6.15 show the corresponding notional impulse response. In these figures, f_{in} is 1000 Hz and f_{out} is the ordinate. Similarly, t_{in} is 0 and t_{out} is the abscissa. Note that while the filter is taken to be symmetrical in terms of linear frequency (which is more or less true), the impulse response isn't. This is because, in terms of linear frequency, the filter becomes wider when centred at a higher frequency.

Note that the filter and impulse response are by no means circular, being instead pinched in off the major axes. This goes some way towards reproducing the suppression effect (see under PSYCHOPHYSICS - HEARING - MASKING - Suppression), which would require hollows off the major axes even going negative in some quadrants.

Programs: \lwork\kernel.wk3 and \lwork\psf.wk3.

6.3.2 PR based on time-frequency representation

Several short passages of speech and music have been analysed with a view to deriving the statistics of real sounds. Figure 6.16 shows the PR of an extract from a financial bulletin read by an American lady saying "[The]

pound benefitted from the dollar's weakness and rose to the giddy height of one dollar ninety point four two, that's the highest for f[our years, before dropping back]". The PR, as expected, looks very much like a spectrogram, with high intensity shown dark.

The PR has the following features. First, the ordinate is frequency on an erb scale, giving constant resolution in that direction. Because of this the fundamental and first two harmonics along the bottom of the PR have much the same separation as the top two formants along the top of the PR. As an aside, these top two formants are present in most vowels and voiced consonants, but the top formant disappears during /n/ and /m/ (see "from", "and" and "ninety"), the next formant down disappears during the American /r/ (see "dollar" and "four"), and both disappear during /w/ (see "weakness" and "one").

Second, the time and frequency resolution are deliberately kept down to what is audible by smudging the primitive time-frequency representation derived from the sound itself by the notional time-frequency auditory impulse response of Figure 6.14. The impulse response being asymmetrical in time and frequency, and variable with frequency, this is easier said than done.

Following Loughlin et al (1993) we take our required time-frequency energy distribution of the one-dimensional

sound signal as given by the following equation:

$$E'(t, f) = \iint W_f(t-t', f-f')\, E(t', f')\, dt'\, df' \qquad (26)$$

where W_f is the W of equation (6) centred on frequency f, and E is the raw Wigner distribution of the energy of the sound signal s. The Wigner distribution itself is impossibly precise and therefore goes negative, while any real-world filter W does the necessary blurring and makes the result E' positive. For more detail see the references under **MATHEMATICS - SIGNAL PROCESSING - TIME-FREQUENCY ANALYSIS.**

Still following Loughlin et al (1993), a more useful formulation, since we do not have the Wigner distribution ready, is

$$E'(t, f) = \iint V_f(\phi, f-f')\, S(f'+\phi/2)\, S^*(f'-\phi/2)\, e^{j2\pi\phi t}\, df'\, d\phi \qquad (27)$$

where ϕ is frequency lag, V_f is the Fourier transform of W_f in the first of its arguments, and S is the Fourier transform of the sound signal s.

Now a Fourier transform of a time signal gives values at preset equally spaced values of frequency. We on the

other hand require a PR at frequencies equally spaced on an erb scale. This is solved by rewriting Equation (27) as

$$E'(t, f) = \int_{-\infty}^{\infty} F(\phi, f)\, e^{j2\pi\phi t}\, d\phi \qquad (28)$$

where

$$F(\phi, f) = \int_{-\infty}^{\infty} V_t(\phi, f-f')\, S(f'+\phi/2)\, S^*(f'-\phi/2)\, df' \qquad (29)$$

may be calculated for arbitrary f at which S need not be known.

For more details, pick the code of \cwork\progs\hearstat.c.

The PR for sounds derivedhere is designed to have the resolution of human hearing. As an interesting crossmodal exercise in resolution, measure the distance at which the PR of Figure 6.16 can be comfortably seen. For me this is about 4 m. At this distance the vertical angle subtended by the whole frequency range in the speech is around 0.7°. The straightforward piano transform attempts to map the complete vertical field of vision of say 120° on to this frequency range.

6.3.3 Statistics of sound PR

Figure 6.17 shows the PR of the full text (apart from the initial "The") together with a bottom row of 10 panels showing the correlation between the excitation at one frequency and the excitation at all other frequencies and a range of time lags. High correlation is shown dark. Panel 1 and panels 9 and 10 are to be disregarded, the relevant base frequency being outside the range of frequencies in the signal.

Figures 6.18 and 6.19 show similar information for two musical extracts, one a rapid dance, the conga, and one the opening bars of Brahms's 4th symphony.

Bearing in mind the enormous computation required to decorrelate a 16 x 16 pixel picture, it was decided to investigate simultaneous correlation of the excitation pattern by itself, with no time lag. This is given by a section up the left-hand edge of each panel in the bottom row in Figures 6.17 to 6.19. These sections are plotted in Figure 6.20 for the speech and Figure 6.21 for the Brahms.

The fall-off of correlation with frequency separation appears to be divided into a steeper central portion and a shallower skirt, and to be largely independent of centre frequency. In order to try to extract this trend,

all correlation curves are superimposed and averaged in Figure 6.22 for the speech and in Figure 6.23 for the Brahms, and the two averages again superimposed in Figure 6.24. Given the scatter, the difference between speech and music in this respect is considered insignificant, and a common line fitted to both sets of points:

$$\rho = \max\left(1 - 0.278\,\Delta g,\ 0.78\,e^{-0.066\,\Delta g}\right) \qquad (30)$$

6.3.4 Decorrelation of sound PR

A covariance matrix based on equation (30) was generated and KL basis functions extracted in the same way as explained for scenes. Figure 6.25 shows every fifth basis function obtained. The corresponding figure for scenes is 6.9 or 6.10.

Figure 6.26 shows the KL transform coefficient variance. As expected, the variance falls off rapidly with coefficient number in a similar way to the variances for unsharpened pictures in Figure 6.11.

Figure 6.27 shows a randomly generated excitation pattern using random numbers, the variances of Figure 6.26, the basis functions of Figure 6.25, and equation (2) of Chapter 2. Excitation patterns generated in this way

only rely on statistical similarity to real ones, and have no other safeguard against being impossible (slope too steep in dB/erb).

Programs: \cwork\progs\speckl.c and \lwork\speckl.wk3.

6.4 Statistics of sparse-spectrum sounds

6.4.1 Preamble

If most possible sinusoids are not present in a sound (often effectively the case) a smaller number of numbers results from specifying only those that are present, which means that their frequencies must count as variables as well as their loudnesses. The statistics of sounds defined in such a way were examined by generating sets of sinusoids at random frequencies and loudnesses and sifting out the valid ones, defined as those in which none of the sinusoids is masked by the others. This proved easier said than done.

For sounds like these, with N sinusoids, the PR vector fed to the KL decorrelating procedure consists of a list of the N frequencies followed by the N loudnesses, both in units of their resptective difference limens, and is therefore 2N long.

6.4.2 Generating procedure

First, the number N of sinusoids (or spectral peaks) is chosen. Next, each is assigned at random a frequency lying comfortably in the audible range (3 to 33 erbs). Third, the sinusoids are numbered in increasing order of frequency.

Next comes the assignment of loudnesses. If any sinusoid masks another, the masked sinusoid effectively ceases to exist, and the chosen number of numbers is false. Random assignment of loudnesses, except when the sound consists of very few sinusoids, is impractical, since so many of the sounds produced are invalid. Some forethought is required.

Suppose the first loudness is chosen at random (within a specified range) and assigned at random to one of the sinusoids, say sinusoid A. A second sinusoid B is chosen at random from those remaining. The presence of the first sinusoid changes both upper and lower limits on the loudness of the second, raising the lower limit (so that A doesn't mask B) and lowering the upper limit (so that B doesn't mask A). In the general case, instead of sinusoid A, there will already be several sinusoids whose loudness has been fixed, but the principle is the same.

Unfortunately, this doesn't work either, except when there are very few sinusoids, because there soon appears a pair of limits with the upper limit lower than the lower.

Instead, it is necessary to start with a pattern known to be valid, with the loudnesses all set half way between the overall limits. These overall limits are set at threshold (Figure 3.5) for the lower limit and at some chosen maximum level for the upper limit, boosted at high and low frequencies by the inverse of the dBA weightings of Figure 7.11.

Each sinusoid, chosen at random as before, is then assigned a new loudness. The problem is to calculate the limits to this new loudness so that the sinusoid isn't masked by any of the others nor masks any of them. Call the sinusoid being adjusted sinusoid B, power p_B in W/m², loudness l_B in dB SPL, frequency f_B in Hz and g_B in erbs.

First the lower limit p_B^-. This is set at one difference limen above the excitation level at frequency B due to the loudnesses of all the other sinusoids, whether adjusted or not:

$$p_B^- = p_{0B} + PRL \sum_{i \neq B} p_i A_{iB} \qquad (31)$$

where $A_{f_{in}\, f_{out}}$ is the power attenuation matrix element given

by the equation for PA in Figure 3.1 or by equation (17), the inclusion of the threshold power p_{08} is justified in Chapter 3 (Figures 3.3 and 3.4), and PRL is the power ratio limen corresponding to the decibel difference limen dBDL:

$$PRL = 10^{dBDL/10} \tag{32}$$

The upper limit to the loudness of sinusoid B is set by the most vulnerable of the other sinusoids. Call this sinusoid i. Sinusoid i must be left at least one difference limen above the excitation pattern produced by all other sinusoids including B:

$$p_i > p_{0i} + PRL \sum_{j \neq i} p_j A_{ji} \tag{33}$$

Replacing item B with item i in the summation and rearranging, and taking the minimum over all i, the upper limit to the loudness of B is given by

$$p_B^* = \min_{i \neq B} \frac{\frac{(1+PRL)\ p_i - p_{0i}}{PRL} - \sum_{j \neq B} p_i A_{ji}}{A_{Bi}} \tag{34}$$

For some reason it is still possible with this procedure to produce sets of sinusoids in which one or more are masked, as shown by curve D in Figure 8.9.

6.4.3 Decorrelation of sparse-spectrum PR

The statistics of sounds generated in this way were examined as follows. First, as many sounds are generated as are necessary to obtain 1000 valid sounds (all sinusoids distinct). Each is considered to be a vector of length 2N, with as the first N elements the sinusoid frequencies in GDLs and as the second N elements the loudnesses in dBDLs. For this purpose, GDL is taken equal to a constant 0.1 erbs, and dBDL equal to 3 dB.

A covariance matrix was derived from the valid sounds generated and KL basis functions extracted in the same way as explained for scenes. These are shown in Figure 6.28. The corresponding figure for scenes is Figure 6.9 or 6.10 and for full-spectrum sounds Figure 6.25.

Figure 6.29 shows the KL transform coefficient variance. The corresponding figure for scenes is Figure 6.11 and for full-spectrum sounds Figure 6.26.

Figures 6.28 and 6.29 together lead to the conclusion that, as is known intuitively, the frequencies of formants or sinusoids are of far greater significance than their loudnesses.

The figure corresponding to Figure 6.27 showing how a random excitation pattern generated in the KL domain is Figure 3.1 (with error bounds added). It is in fact impossible to reproduce Figure 3.1 (or 6.27) except by photocopying, since every recalculation uses different random numbers and results in a completely different excitation pattern.

Programs: \cwork\progs\ranspekl.c, \lwork2\ranspekl.wk3, \lwork\excite.wk3.

6.5 Matching of scene and sound basis functions

6.5.1 Ambiguous sounds necessary

As mentioned above, full-spectrum excitation patterns generated in the KL domain (Figure 6.27) can be impossible in the sense of being too steep in dB/erb. The corresponding event in the case of sparse-spectrum sounds is not an impossibility but the disappearance by masking of one or more of the sinusoids.

One way to reduce the frequency of this happening, applied in preparing Figure 3.1 but not Figure 6.27, is

to reduce by some reduction factor (called squeeze in the programs) the range of the KL coefficients generated. This results on the one hand, as desired, in less frequently ambiguous sounds (frequency of some component inaudible), but on the other hand reduces the gamut of sounds available for representing scenes.

The inescapable reason for this is the central limit theorem (Papoulis 1984), which states that the sum (the result of an inverse KL transform) of uncorrelated random variables (the KL coefficients generated) tends to have a more normal distribution the more numbers there are in the sum, whatever the distribution of the individual numbers summed. Thus it is not possible, by for instance giving the KL transform coefficients rectangular distributions, to ensure nice sharp cutoffs to the parameters of the sounds produced.

Worse still, KL coefficients not generated at random but derived from a scene would themselves be largely normally distributed already.

6.5.2 Ambiguous matching necessary

Two points to note from the decomposition of scenes into basis functions. First, it can be seen from Figure 6.11 that, although ranked in order of variance, some of the

basis functions have the same variance and are equally weighted. Figures 6.9 and 6.10 show that this occurs when two basis functions are the same apart from a 90° rotation. Basis functions 1 and 2 are an example (counting starts at 0). Such pairs are ordered at random.

Second, looking at Figure 6.9, if basis function 2 is a rotation of basis function 1, why has basis function 3, with two white quadrants and two black, no such partner? The answer is that, for basis function 3, rotation by 90° is the same as multiplication by -1. This is counted as the same basis function because its use would only involve changing the sign of the weighting attached to it in the vector v. The point to note is that the sign of basis function 3, and of every other basis function in Figures 6.9 and 6.10, has been chosen at random.

Now let's return to the purpose of the exercise, which is to do the same for sounds and then match the basis functions one for one. The question is: does it matter which ordering and which sign are chosen? Suppose a scene is decorrelated by this method (say for storage purposes) and then reconstructed with the same weights but with random reordering of equally weighted basis functions and with random selection of a sign for each weight. The result is total confusion. And yet such random choices are precisely what the method requires.

The method therefore results in a very large number of different mappings.

How are we to choose the best? We can choose two at random and compare them, but what will that tell us? Such a plodding approach is ruled out by the training required for each mapping, discussed in Chapter 3. Not knowing what to do next with the KL method, we go back to the drawing board and find ourselves having what turns out to be another stab at the piano transform.

Figure 6.1

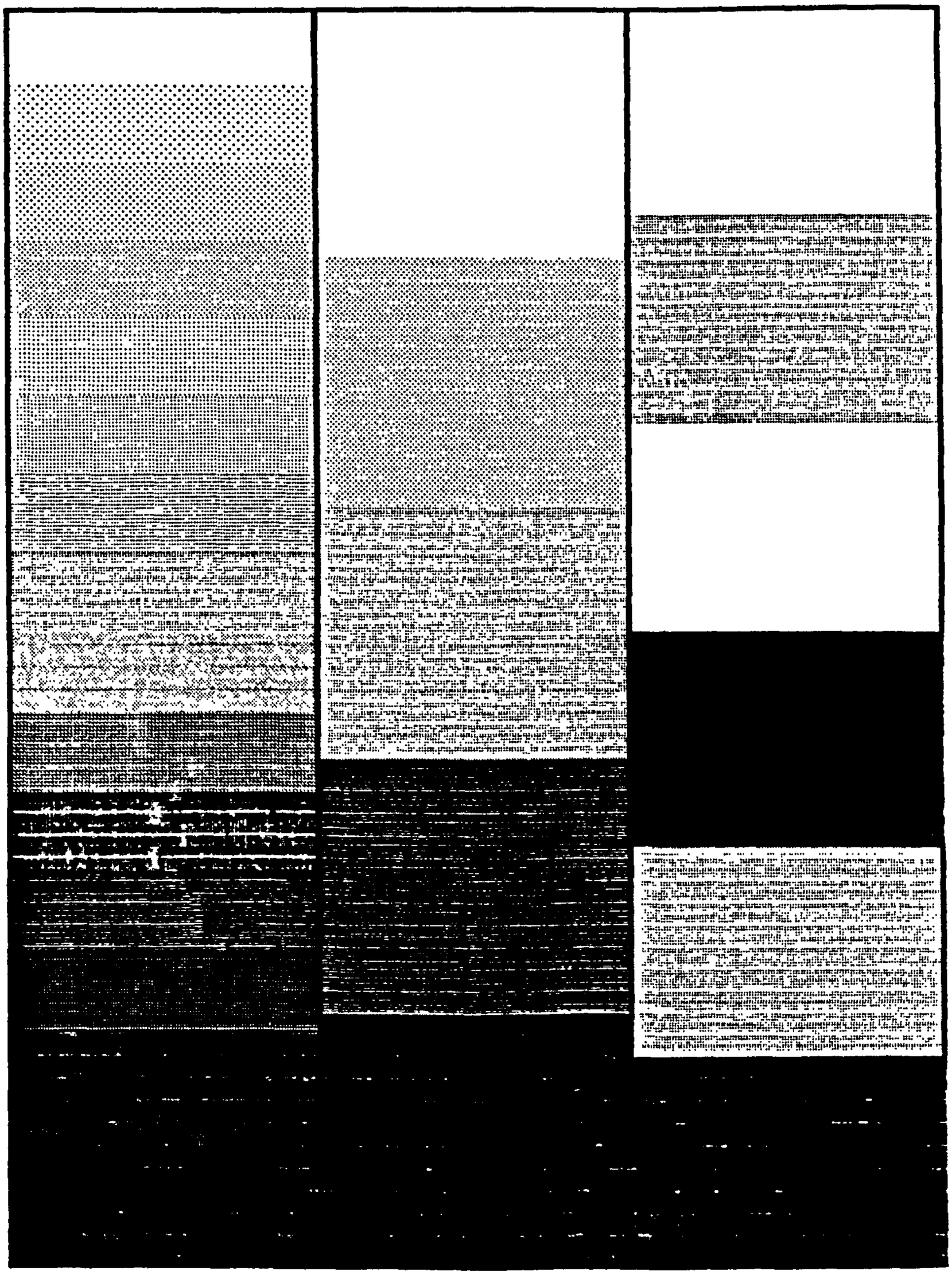


Figure 6.2

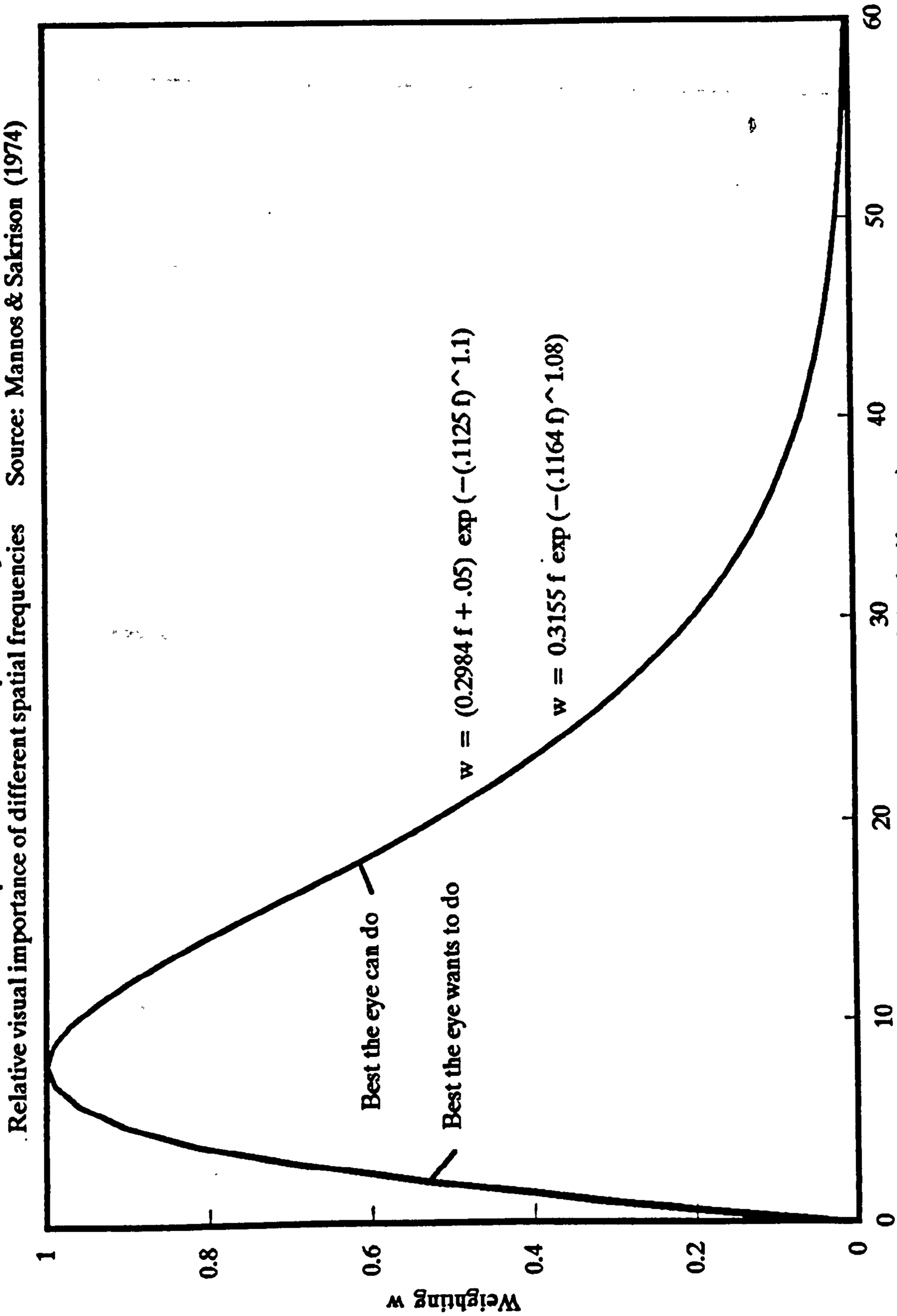

Spatial frequency sensitivity
Relative visual importance of different spatial frequencies Source: Mannos & Sakrison (1974)
Relative visual importance w
Best the eye can do
Best the eye wants to do
w = (0.2984 f + .05) exp (−(.1125 f)^1.1)
w = 0.3155 f exp (−(.1164 f)^1.08)
Spatial frequency f (cycles/degree)
Weighting w
1
0.8
0.6
0.4
0.2
0
10
20
30
40
50
60

Figure 6.3

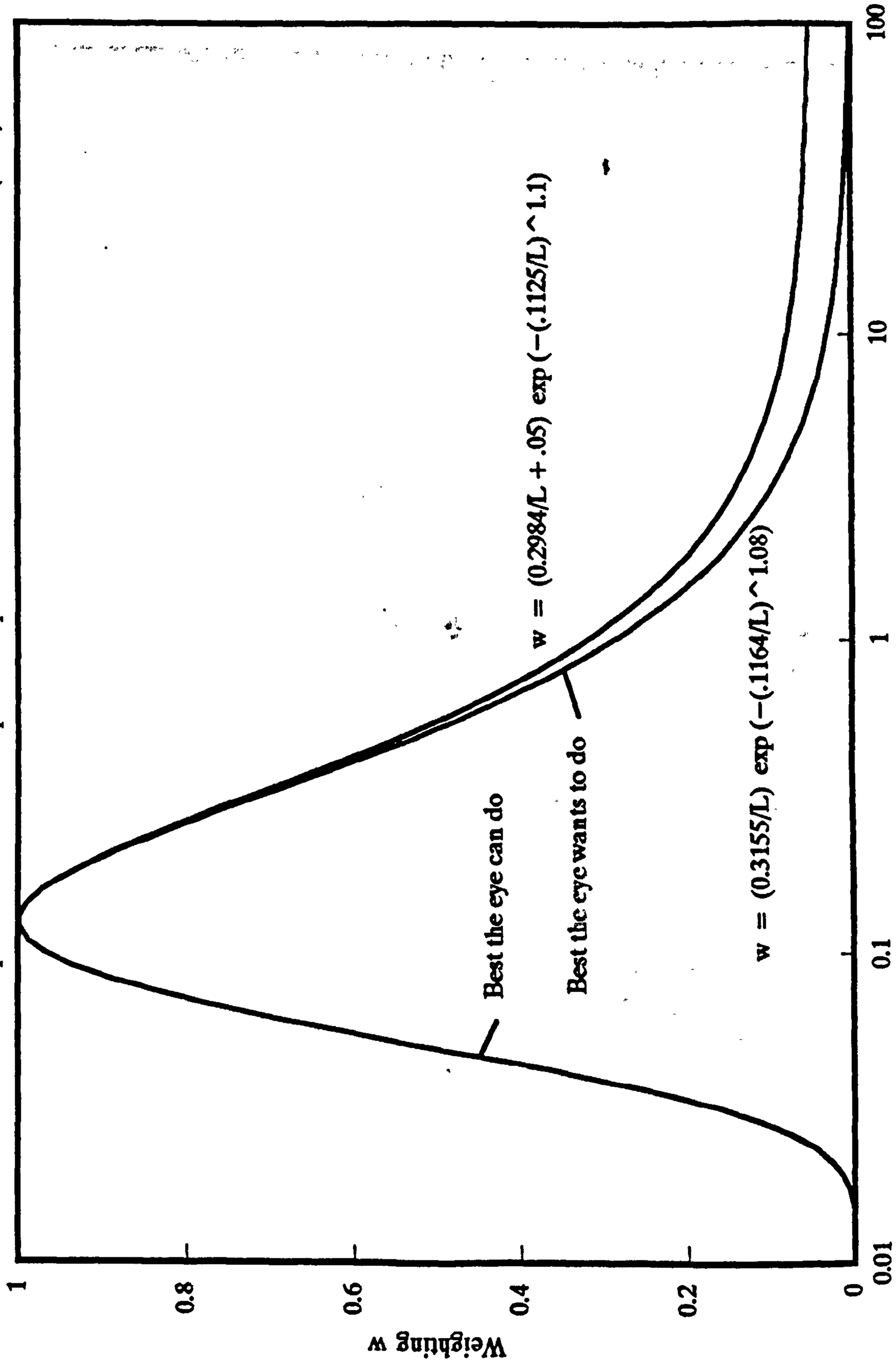

Spatial frequency sensitivity
Relative visual importance of different spatial frequencies
Source: Mannos & Sakrison (1974)
Best the eye can do
Best the eye wants to do
w = (0.2984/L + .05) exp (−(.1125/L)^1.1)
w = (0.3155/L) exp (−(.1164/L)^1.08)
Spatial wavelength L (degrees)
Weighting w
100
10
1
0.1
0.01
1
0.8
0.6
0.4
0.2
0

Figure 6.4

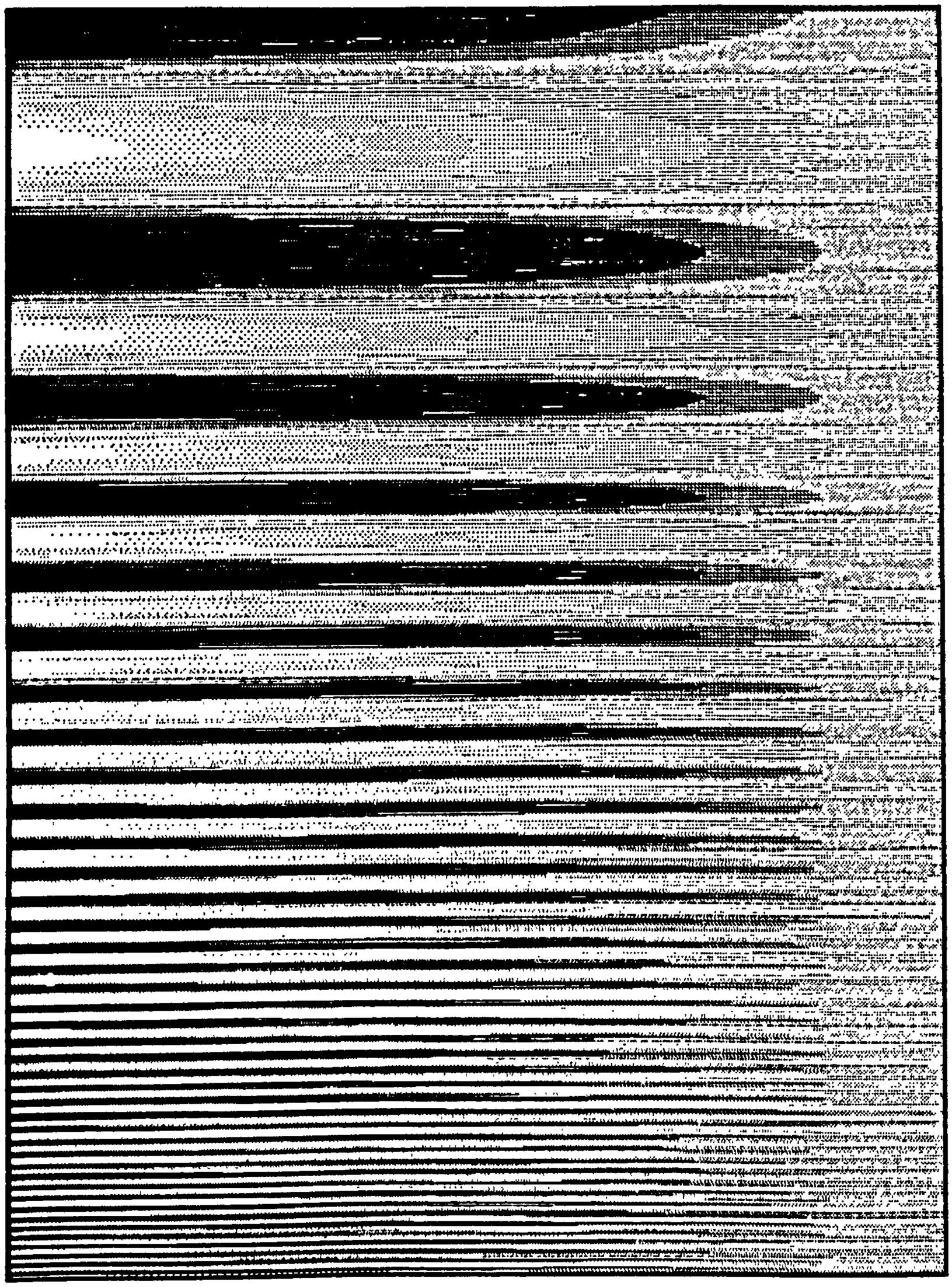


Figure 6.5

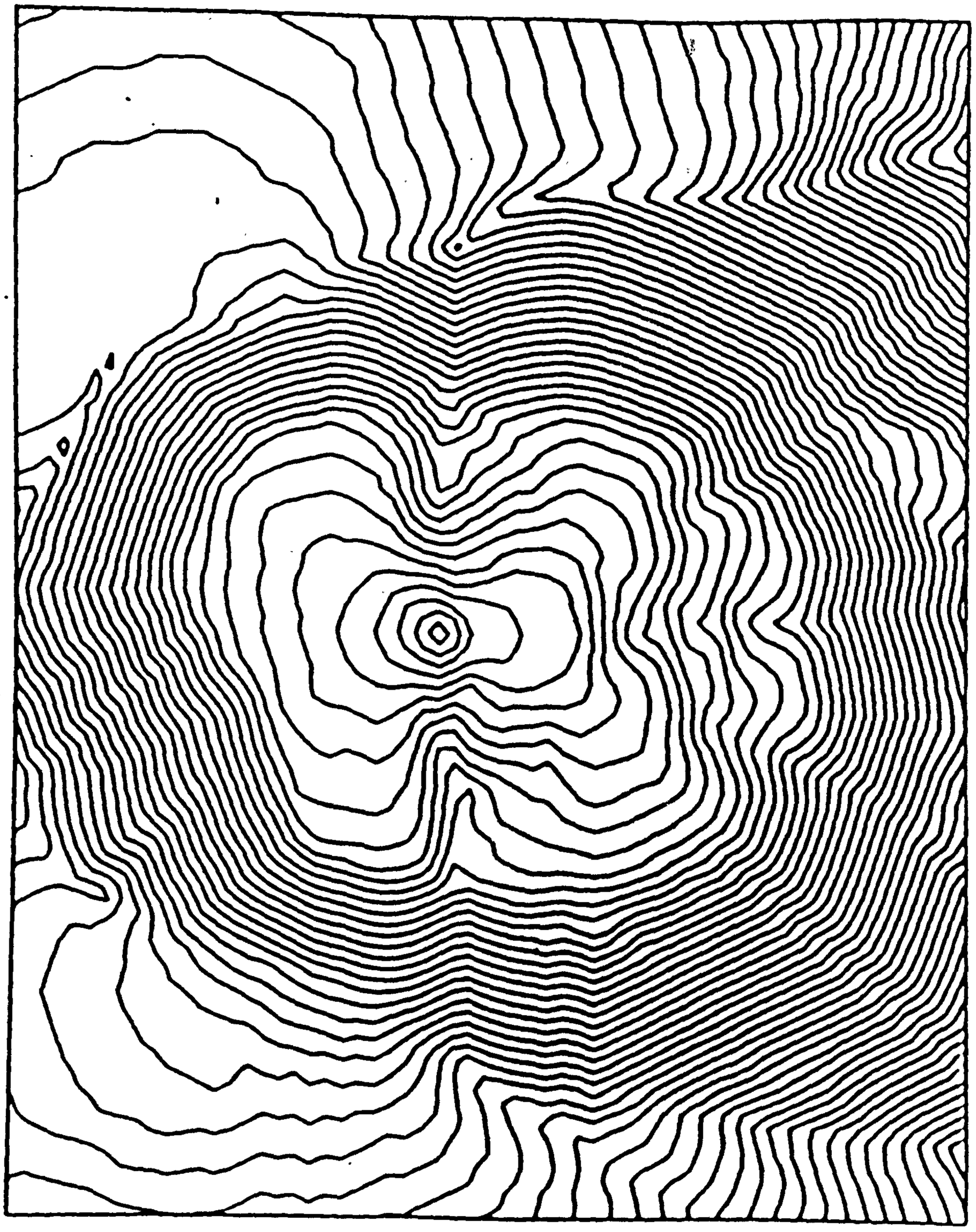


Fig. 1. Eikonal portrait of and by *Wolfgang Möller* [from *Schroeder*, 1983].

Figure 6.6

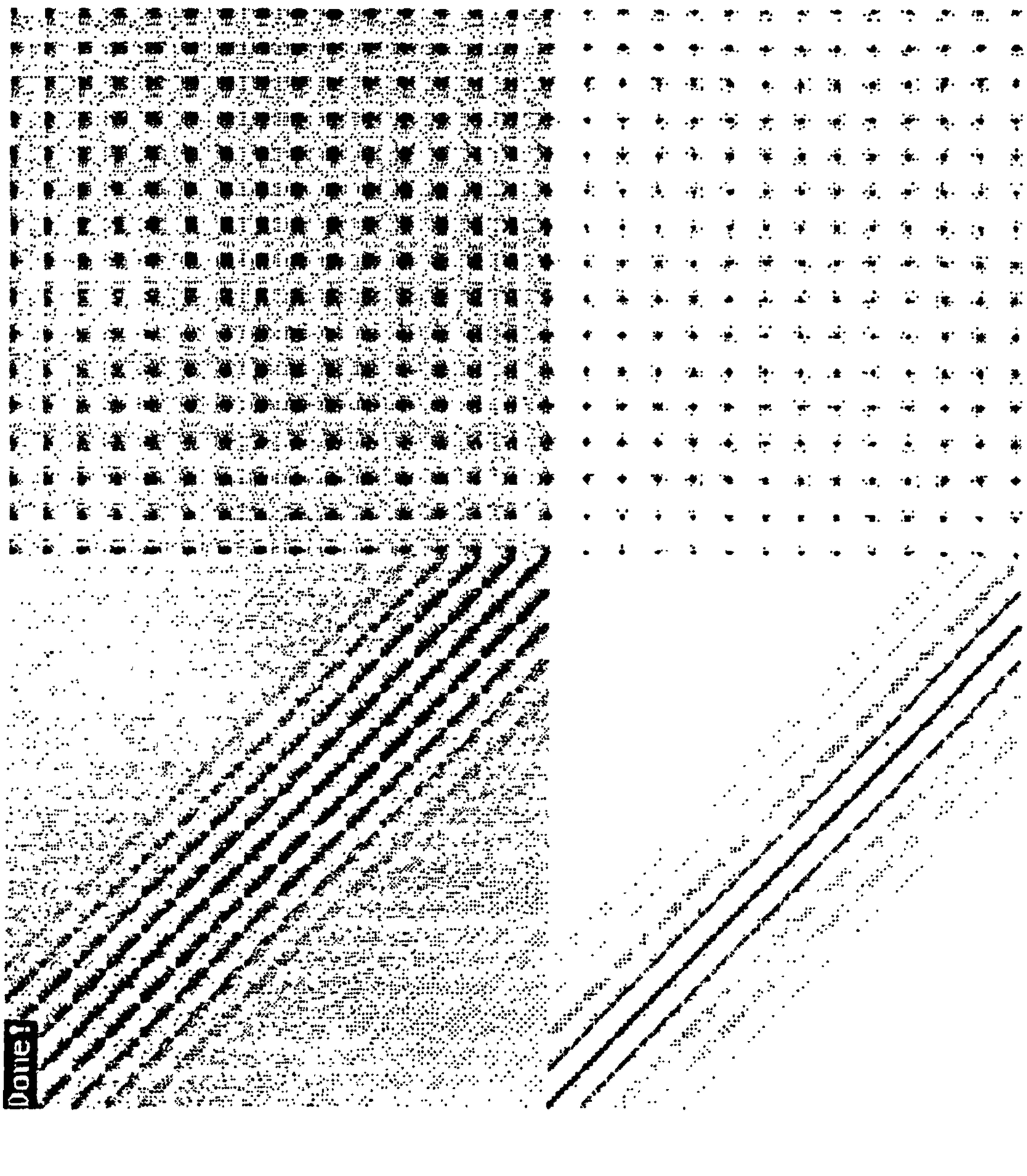


Figure 6.7

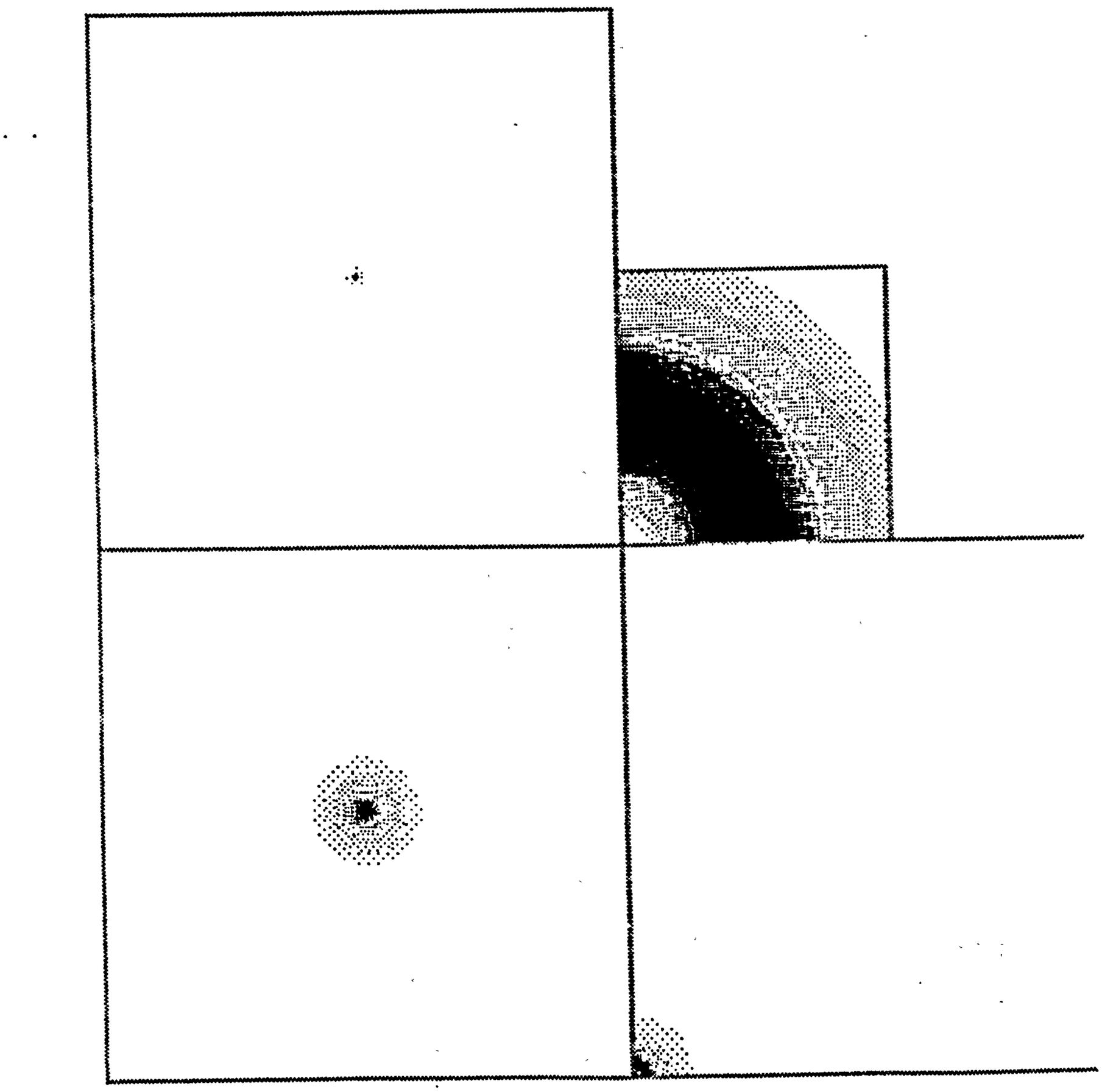


Figure 6.8

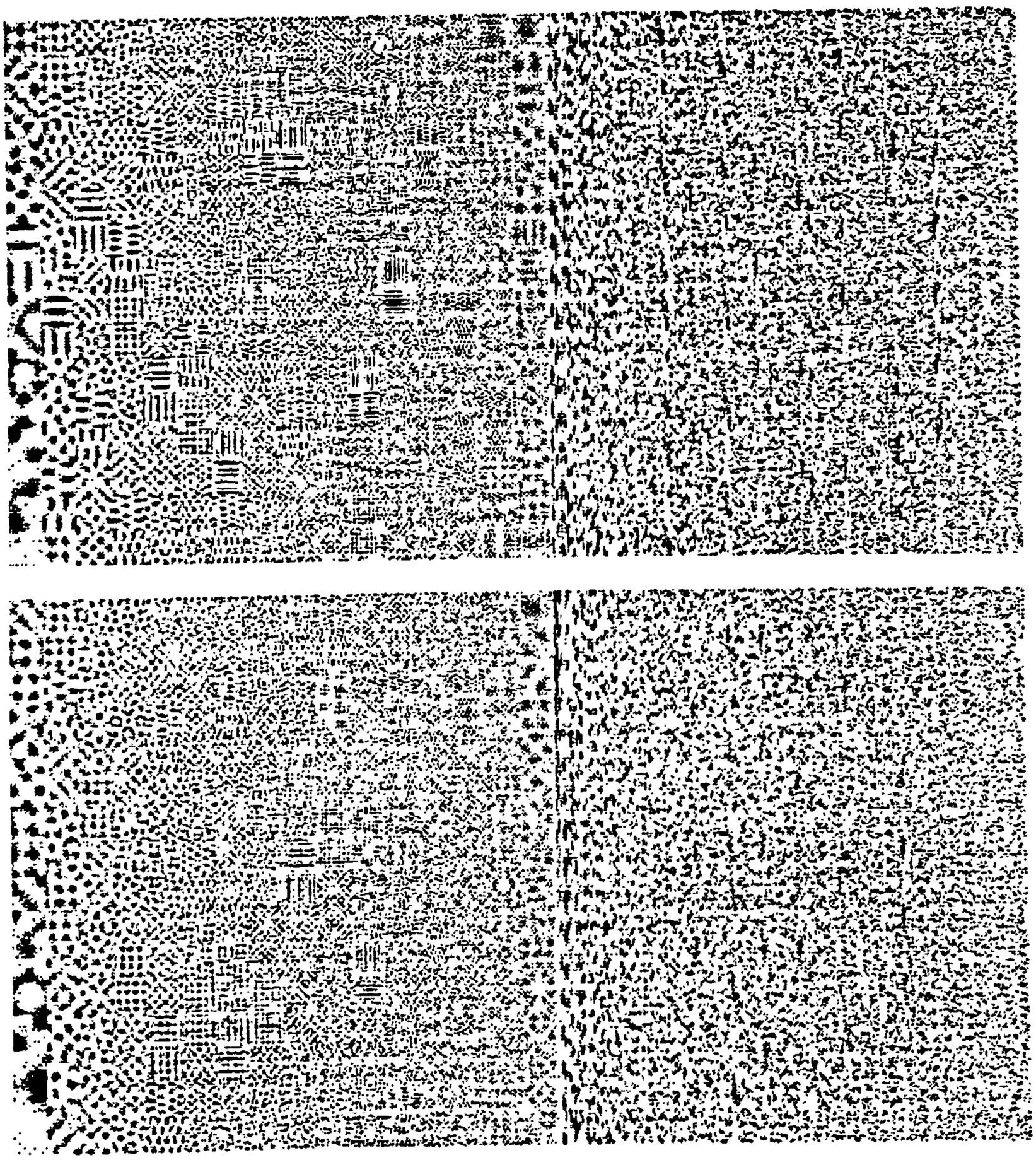


Figure 6.9

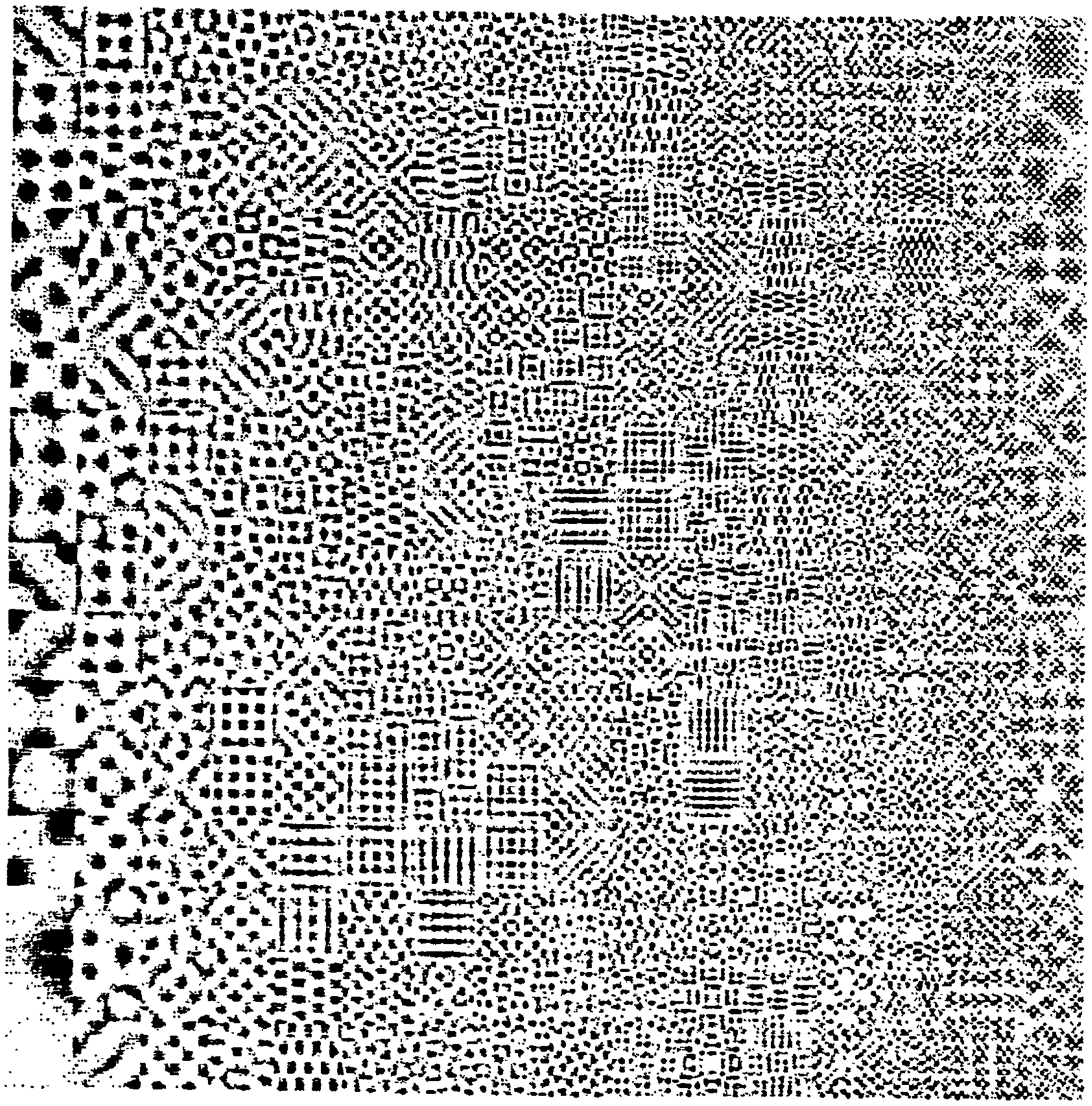
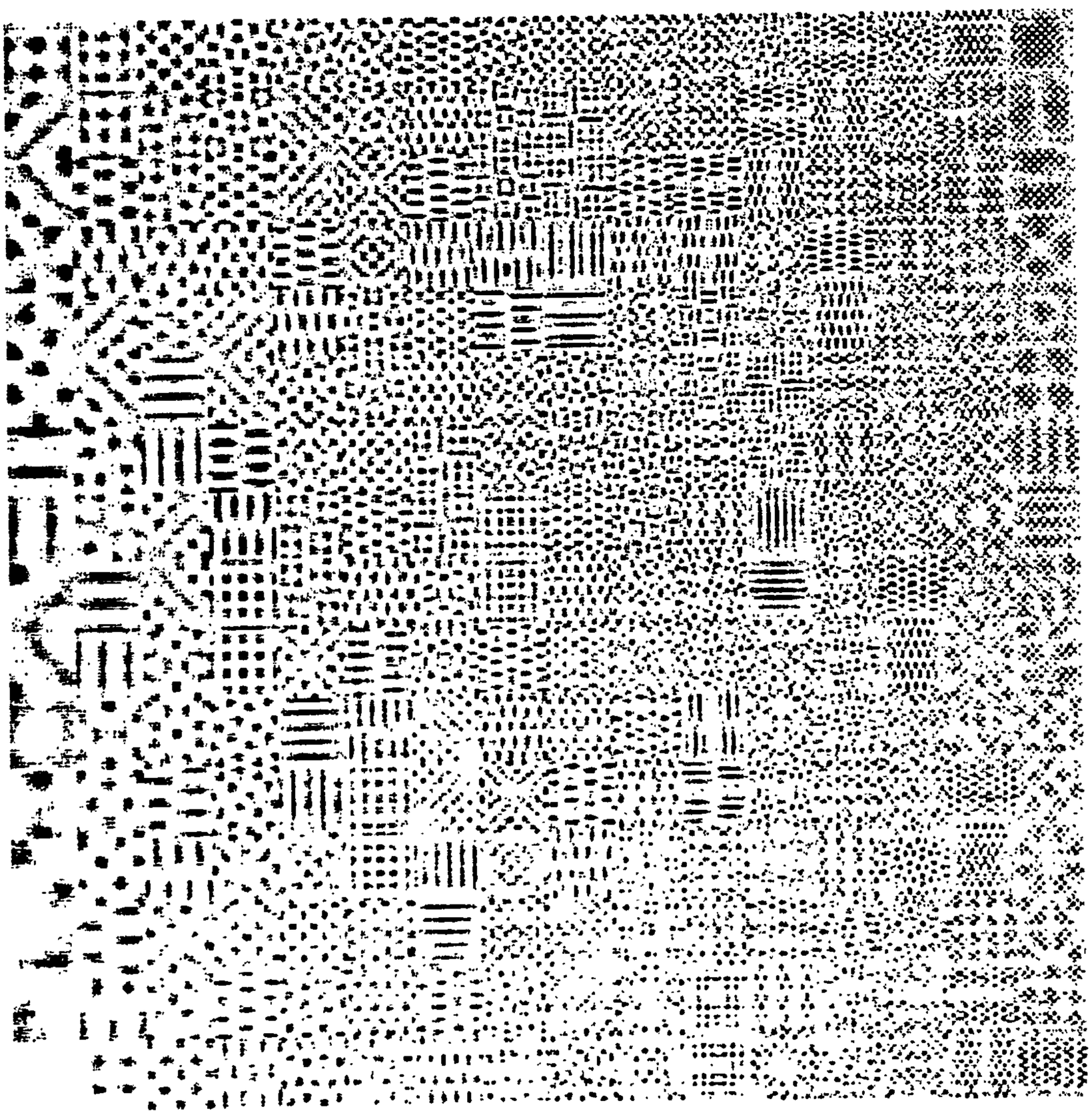


Figure 6.10



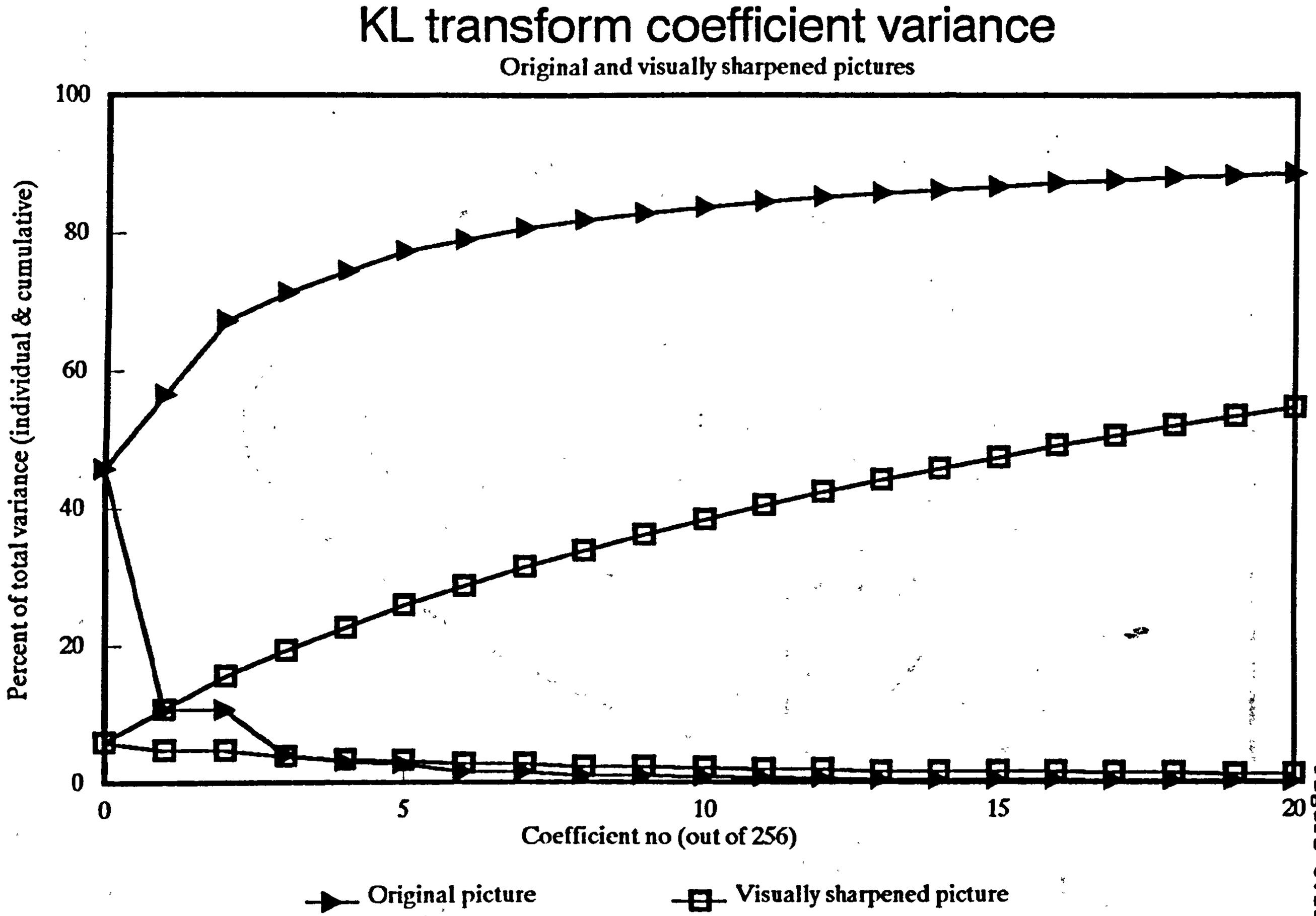


Figure 6.11

Based on 16x16 patch from 512x512 picture subtending 120 degrees. Adjacent pixel correlation 0.9.
Visual sharpening based on Mannos and Sakrison.

Figure 6.12

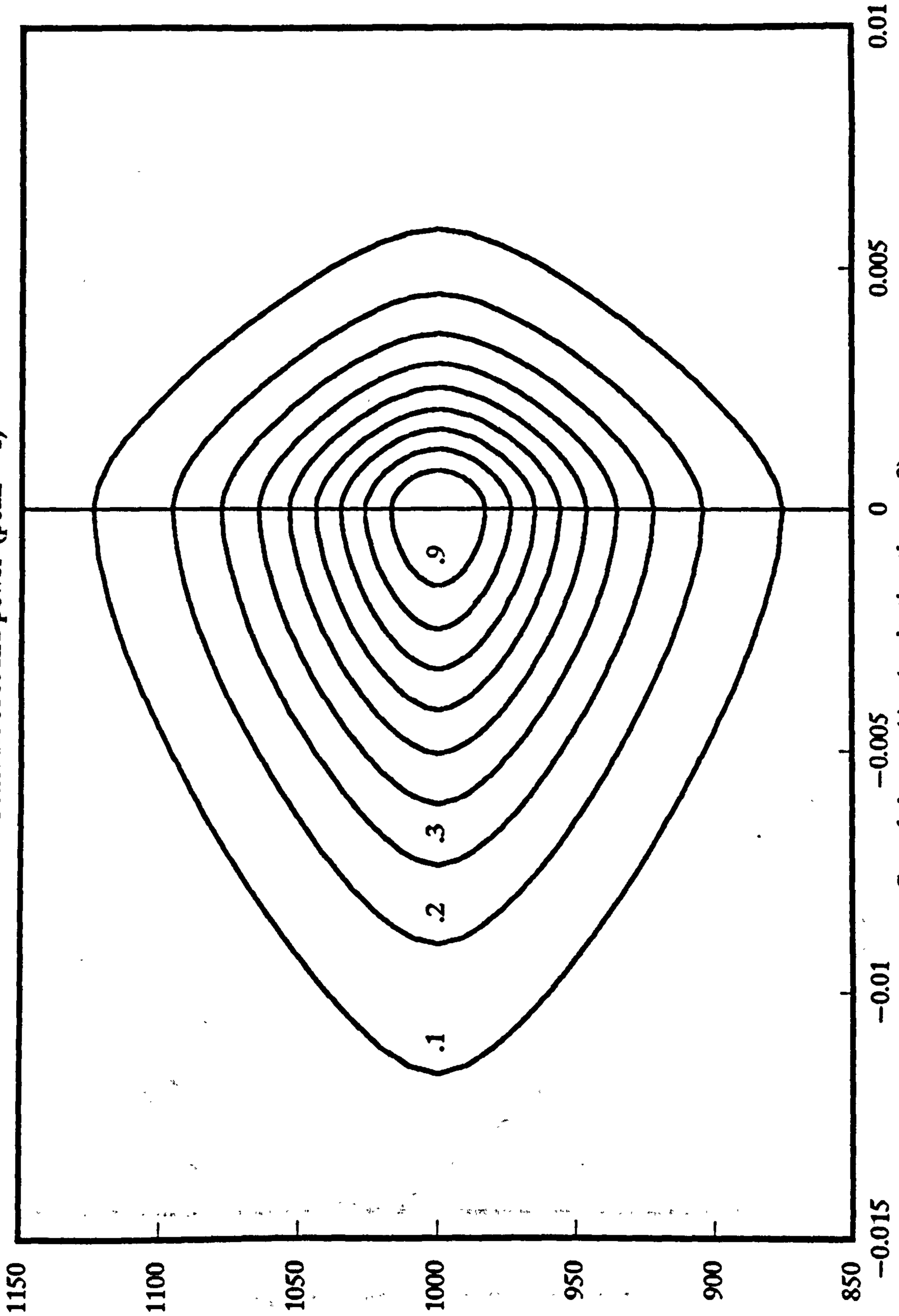

Notional auditory time-frequency filter
Contours of sound power (peak = 1)
.9
.3
.2
.1
0.01
0.005
0
−0.005
−0.01
−0.015
Sound time t (s) (excitation time = 0)
1150
1100
1050
1000
950
900
850
Sound frequency f (Hz) (excitation frequency = 1000 Hz)

Figure 6.13

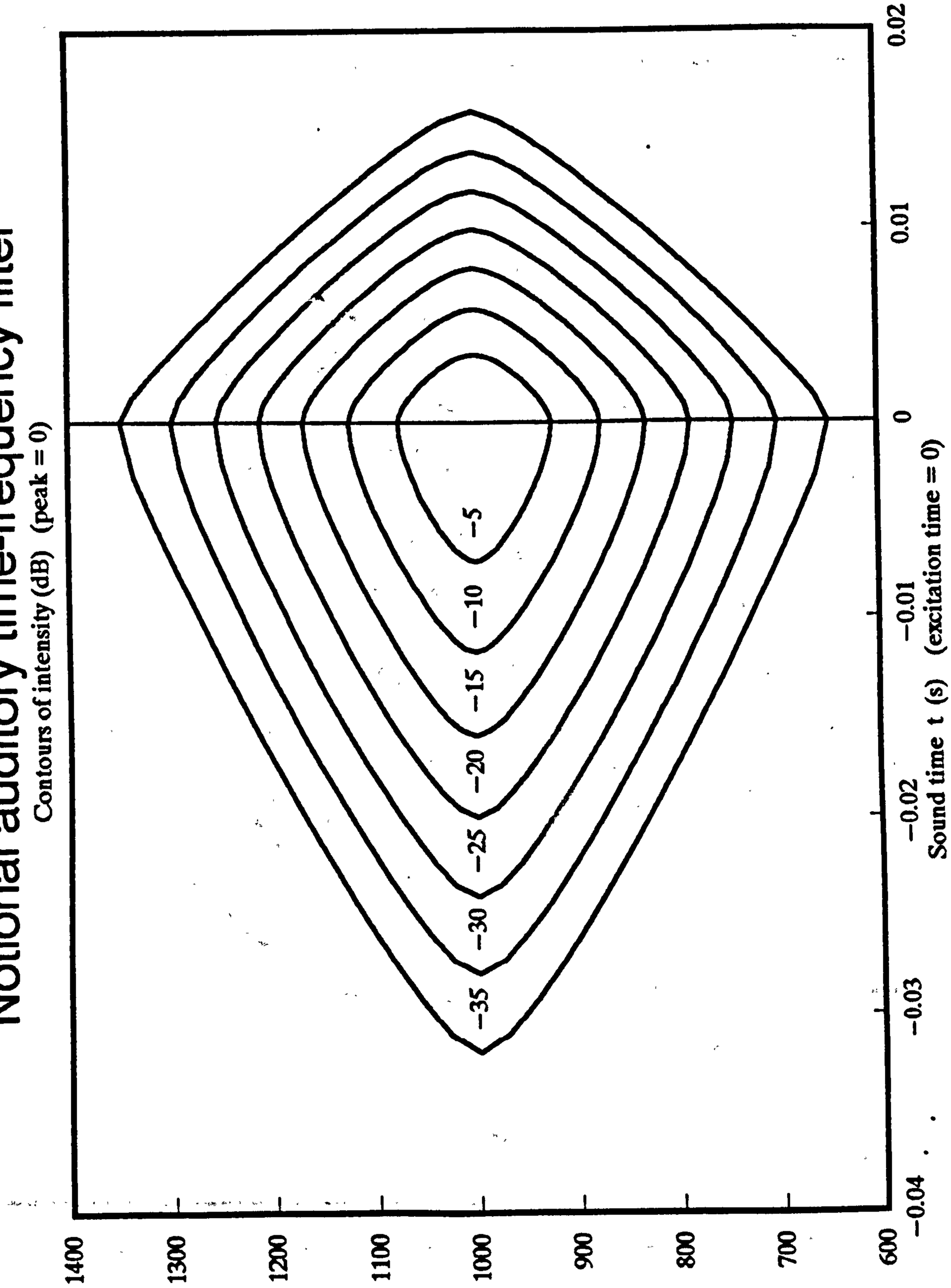

Notional auditory time-frequency filter
Contours of intensity (dB) (peak = 0)
0.02
0.01
0
−0.01
−0.02
−0.03
−0.04
(excitation time = 0)
Sound time t (s)
−5
−10
−15
−20
−25
−30
−35
1400
1300
1200
1100
1000
900
800
700
600
Sound frequency f (Hz) (excitation frequency = 1000 Hz)

Figure 6.14

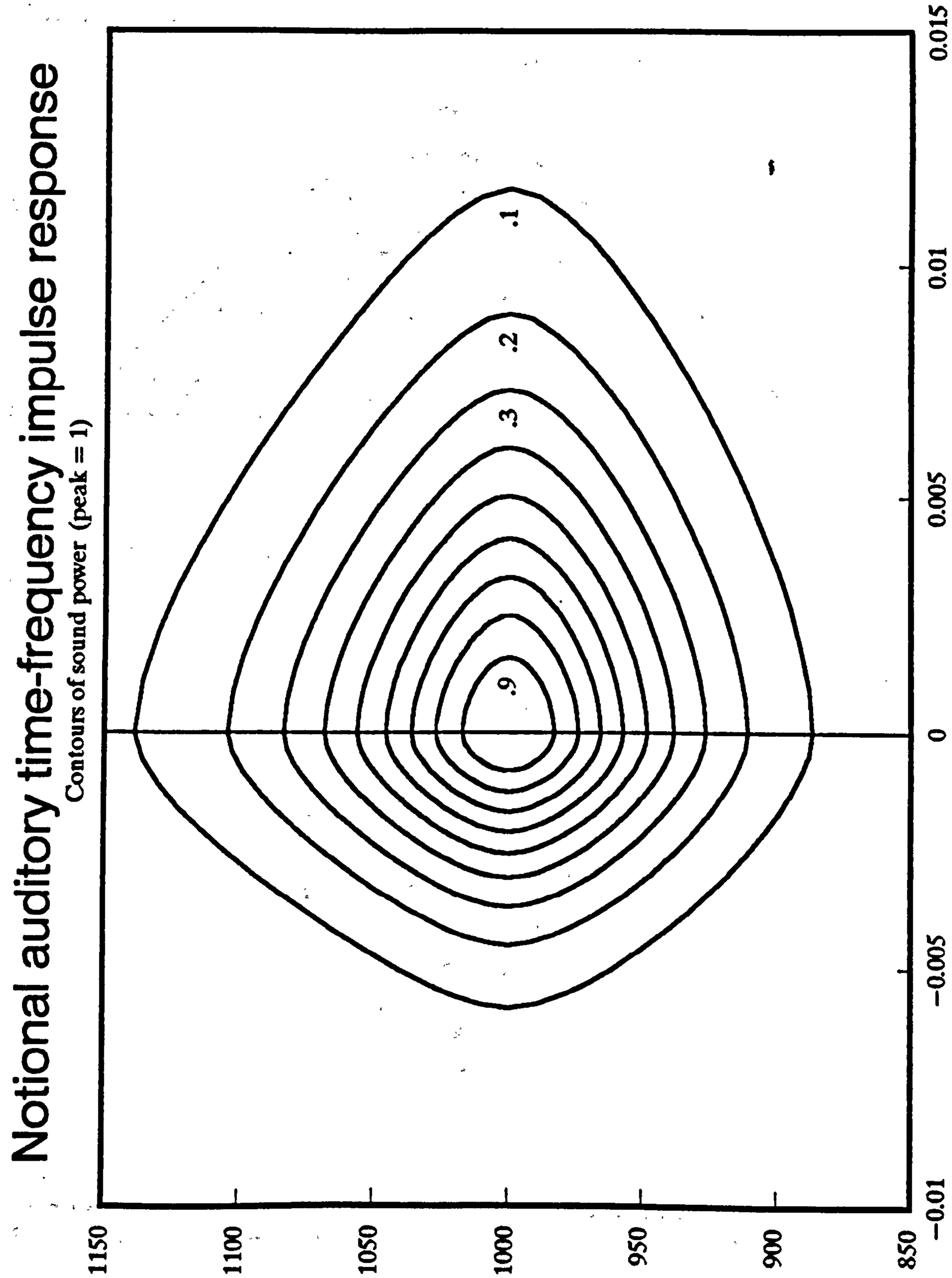

Notional auditory time-frequency impulse response
Contours of sound power (peak = 1)
.1
.2
.3
.9
0.015
0.01
0.005
0
−0.005
−0.01
Excitation time t (s) (sound time = 0)
1150
1100
1050
1000
950
900
850
Excitation frequency f (Hz) (sound frequency = 1000 Hz)

Figure 6.15

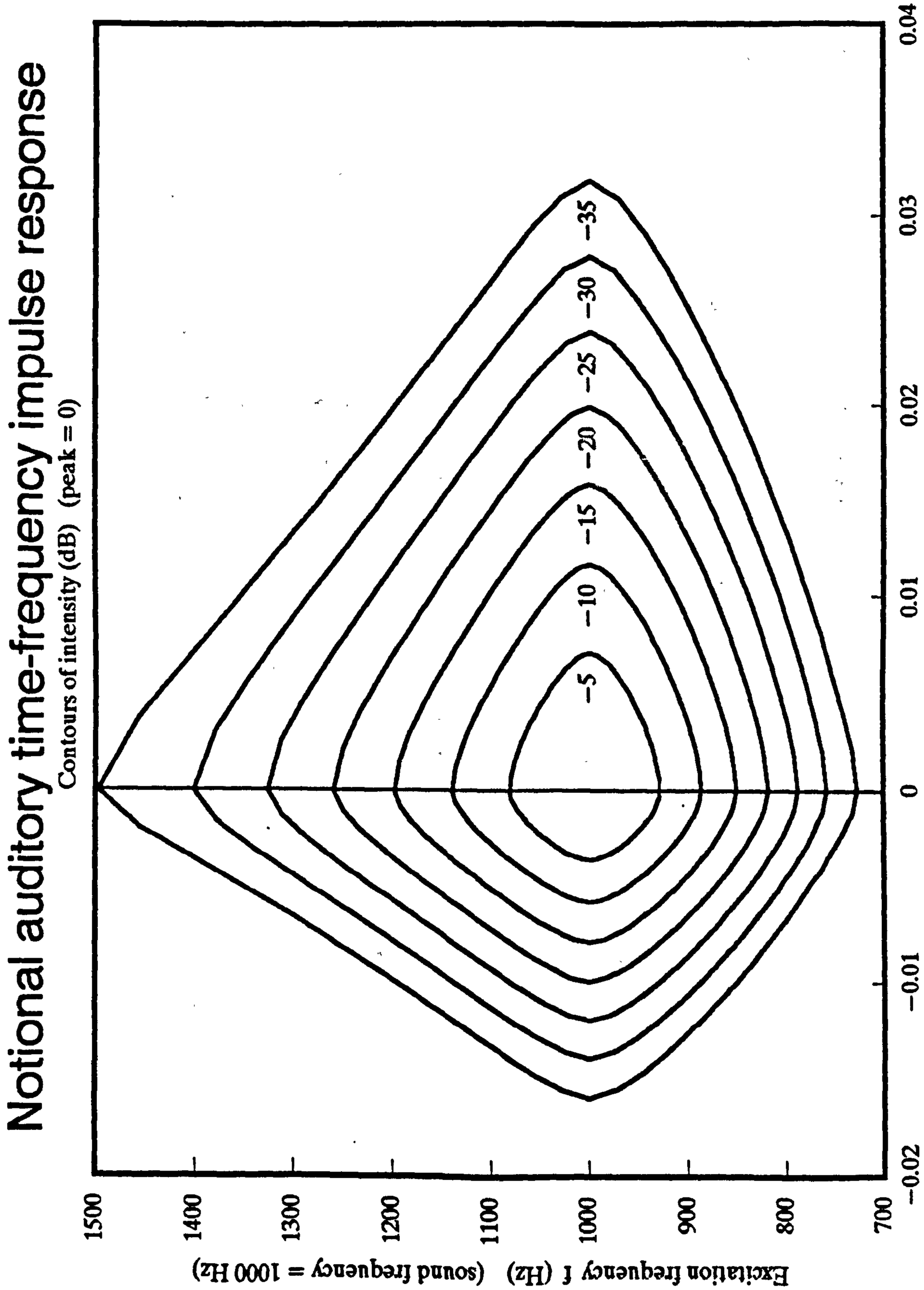
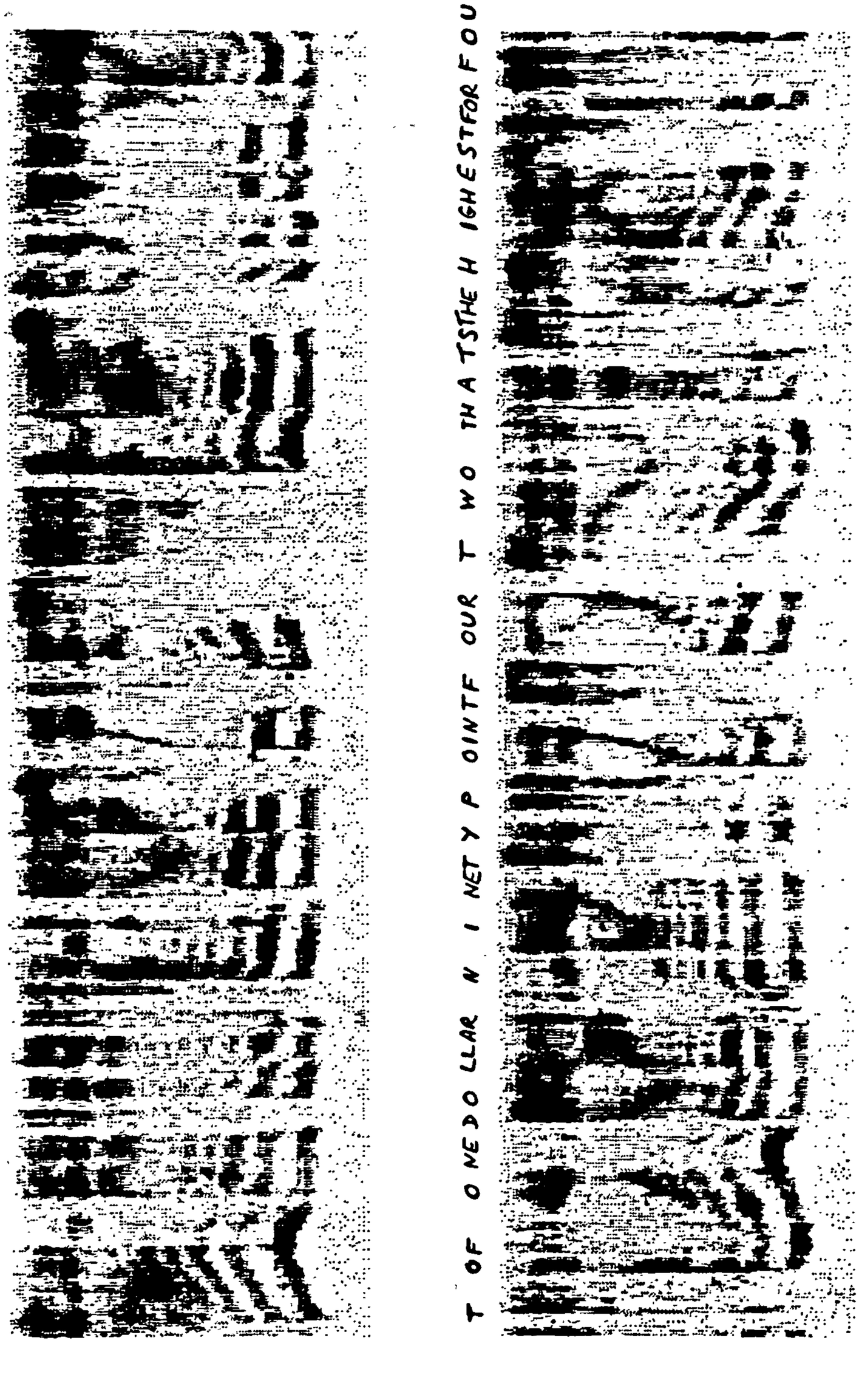

Notional auditory time-frequency impulse response
Contours of intensity (dB) (peak = 0)
−5
−10
−15
−20
−25
−30
−35
Excitation frequency f (Hz) (sound frequency = 1000 Hz)
1500
1400
1300
1200
1100
1000
900
800
700
Excitation time t (s) (sound time = 0)
0.04
0.03
0.02
0.01
0
−0.01
−0.02

Figure 6.16


THE GIDDY H EIGH
RS VEAKNE SS (BREATH) ANDR O SETOTHE
P OU N DBENE F I TTE DFR O MTHE D O LLARS
T OF ONE DO LLAR N I NET Y P O
OINTF OUR T WO TWA TSTHE H IGHESTFOR FOUR...

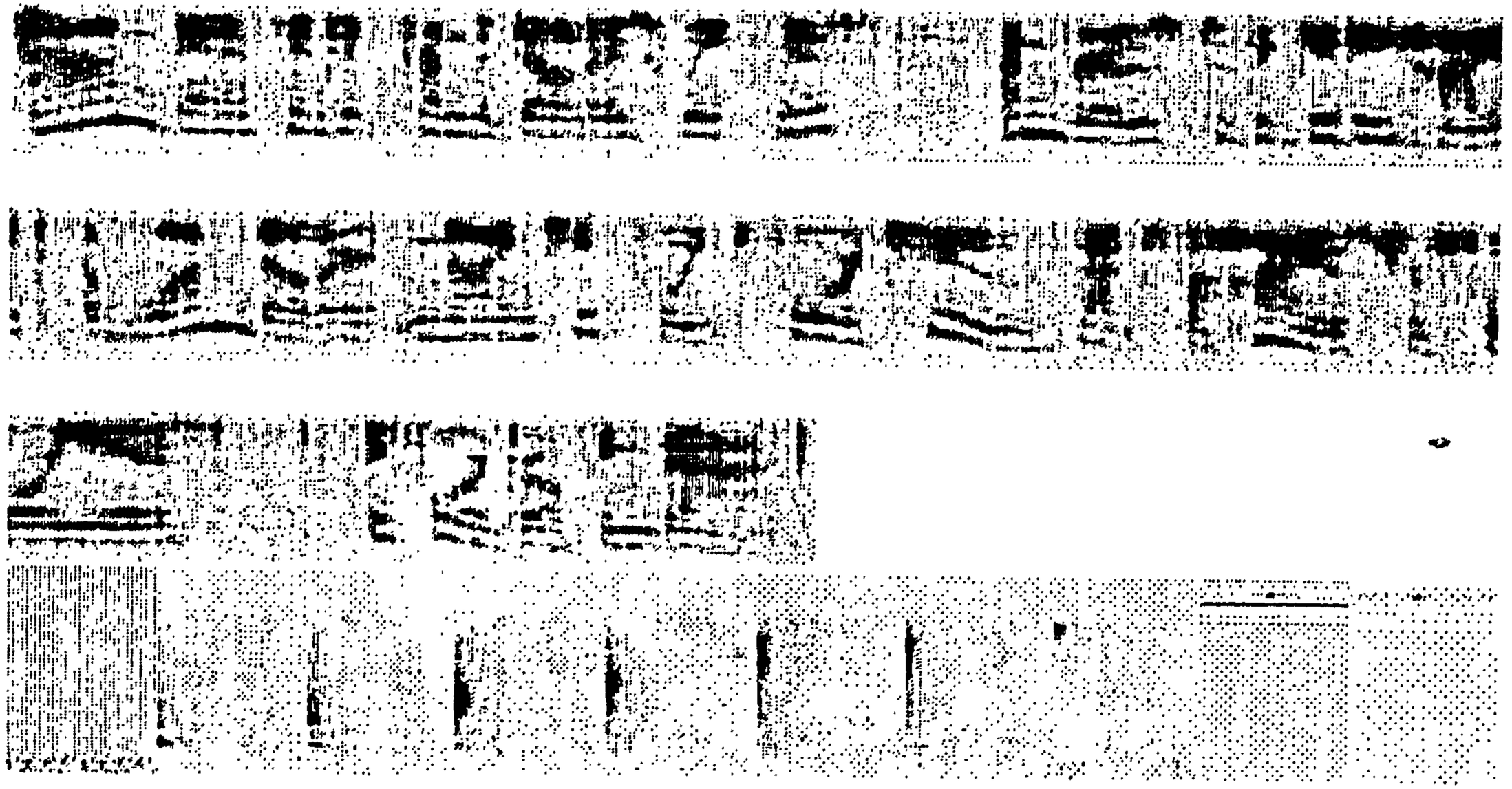
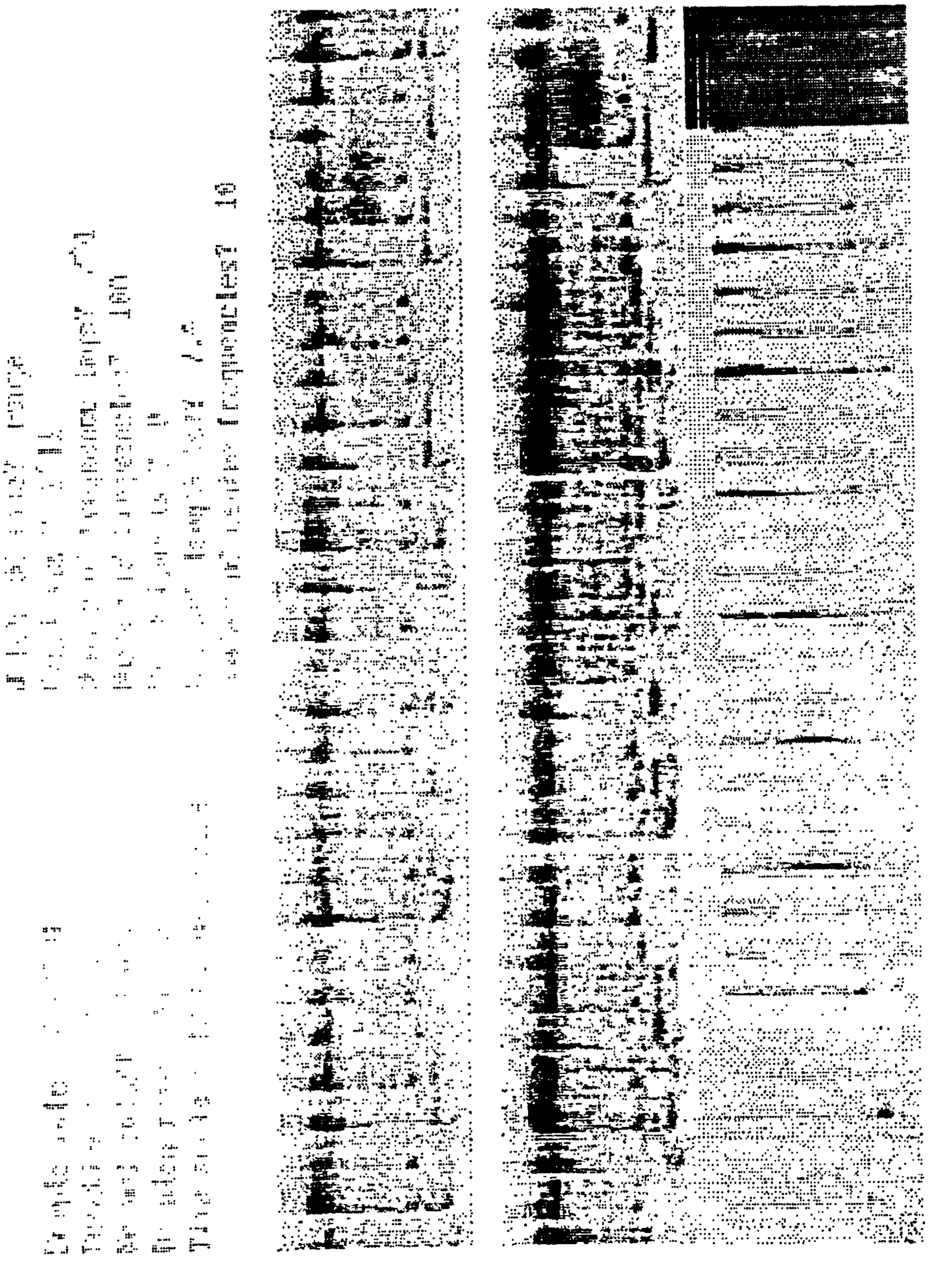


Figure 6.17

Figure 6.18



```
Sample rate = 43178 Hz
Duration:    24.23 s
Kernel cutoff = 3.000
Display rate = 170 Hz
Time scale = 3.77 s per frame

Which VOK file? brahms4
Chunk size? 2^12
Number of frequency lags? 2^4
Number of frequencies? 80
Start time (s)? 0
Extract length (s)? 11.2
Number of centre frequencies? 10
Number of time lags? 64
```

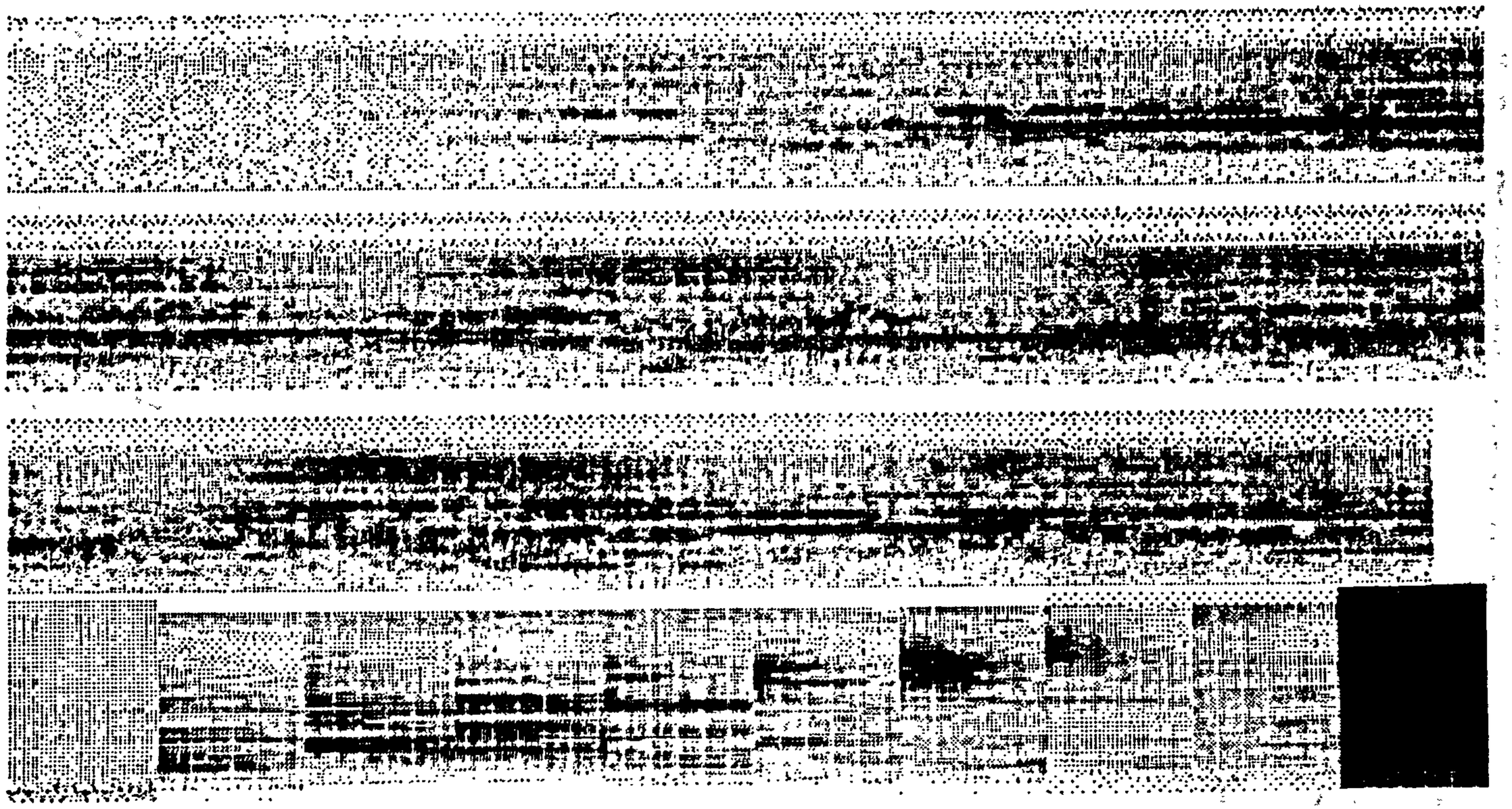


Figure 6.19

Figure 6.20

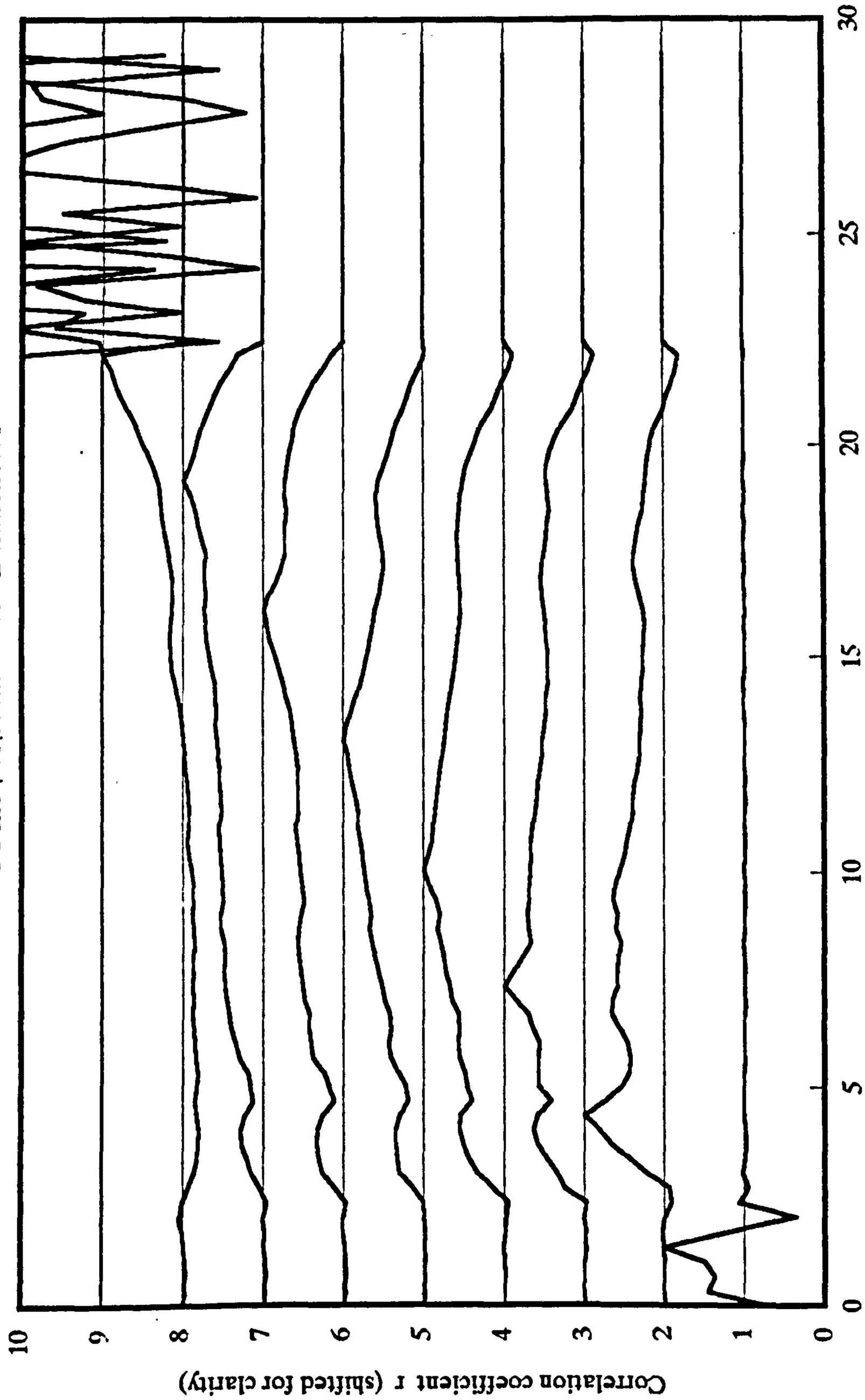
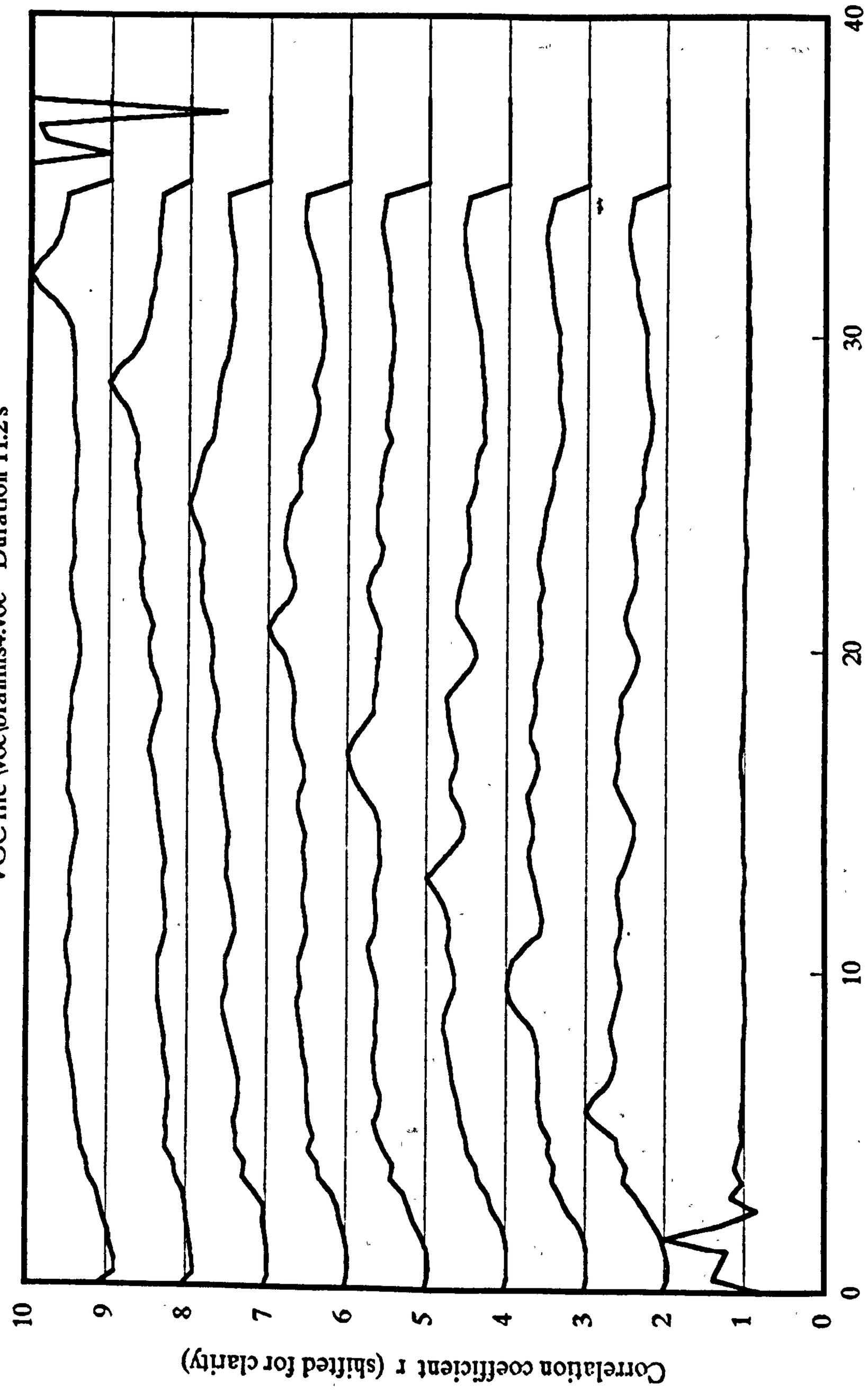

Simultaneous excitation-level correlation
VOC file \voc\bbcnews.voc Duration 9.4 s
Correlation coefficient r (shifted for clarity)
10
9
8
7
6
5
4
3
2
1
0
0
5
10
15
20
25
30
Frequency g (erbs)
r = 1 at each centre frequency.
Bottom and top two centre frequencies were outside signal bandwidth.

Figure 6.21



r = 1 at each centre frequency.
Bottom and top centre frequencies were outside signal bandwidth.

Figure 6.22

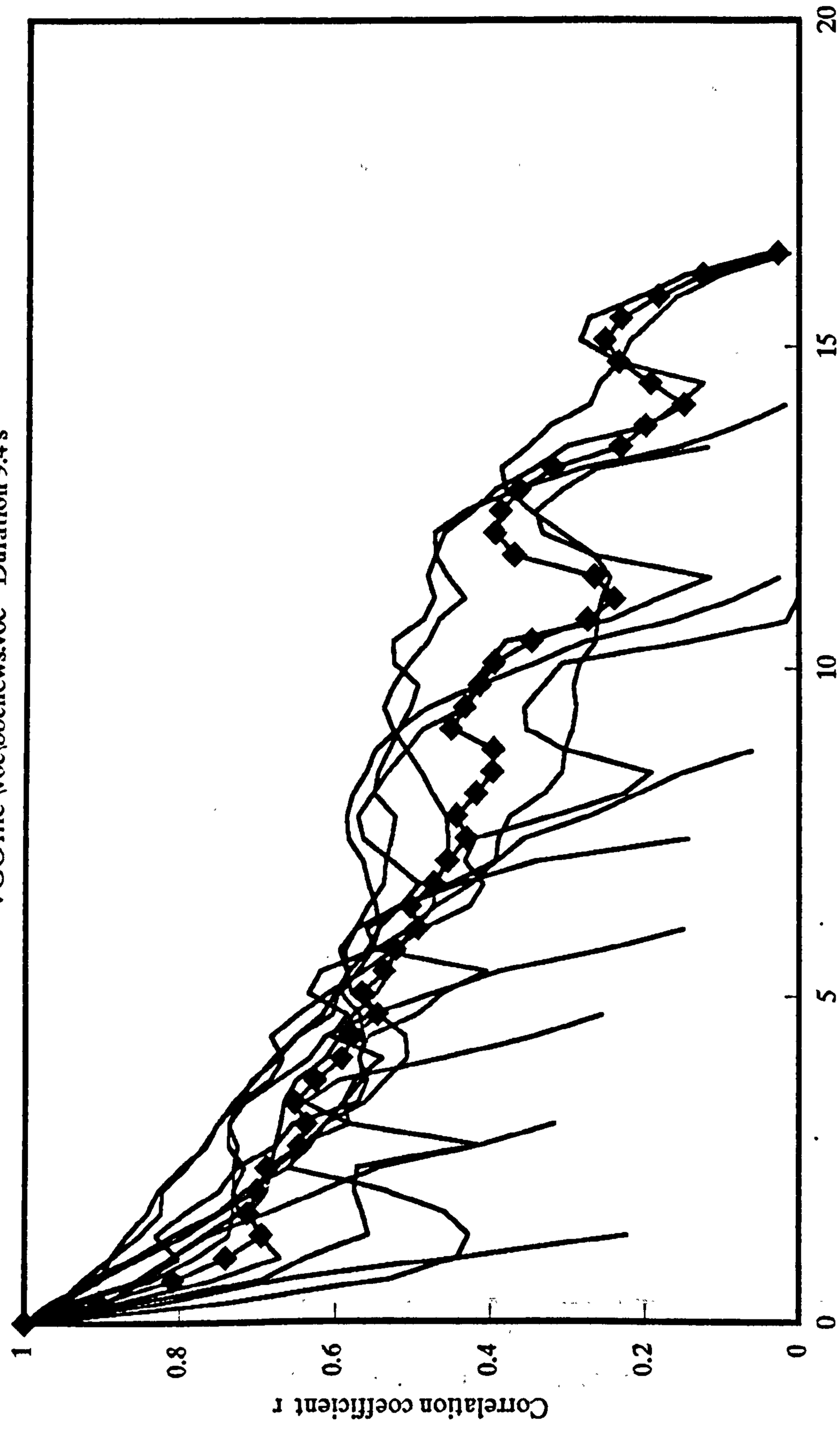

Simultaneous excitation-level correlation
VOC file \voc\bbcnews.voc Duration 9.4 s
Correlation coefficient r
1
0.8
0.6
0.4
0.2
0
Frequency separation dg (erbs)
0
5
10
15
20
Selected centre frequency
Average
Average uses both positive and negative frequency separations.

Figure 6.23

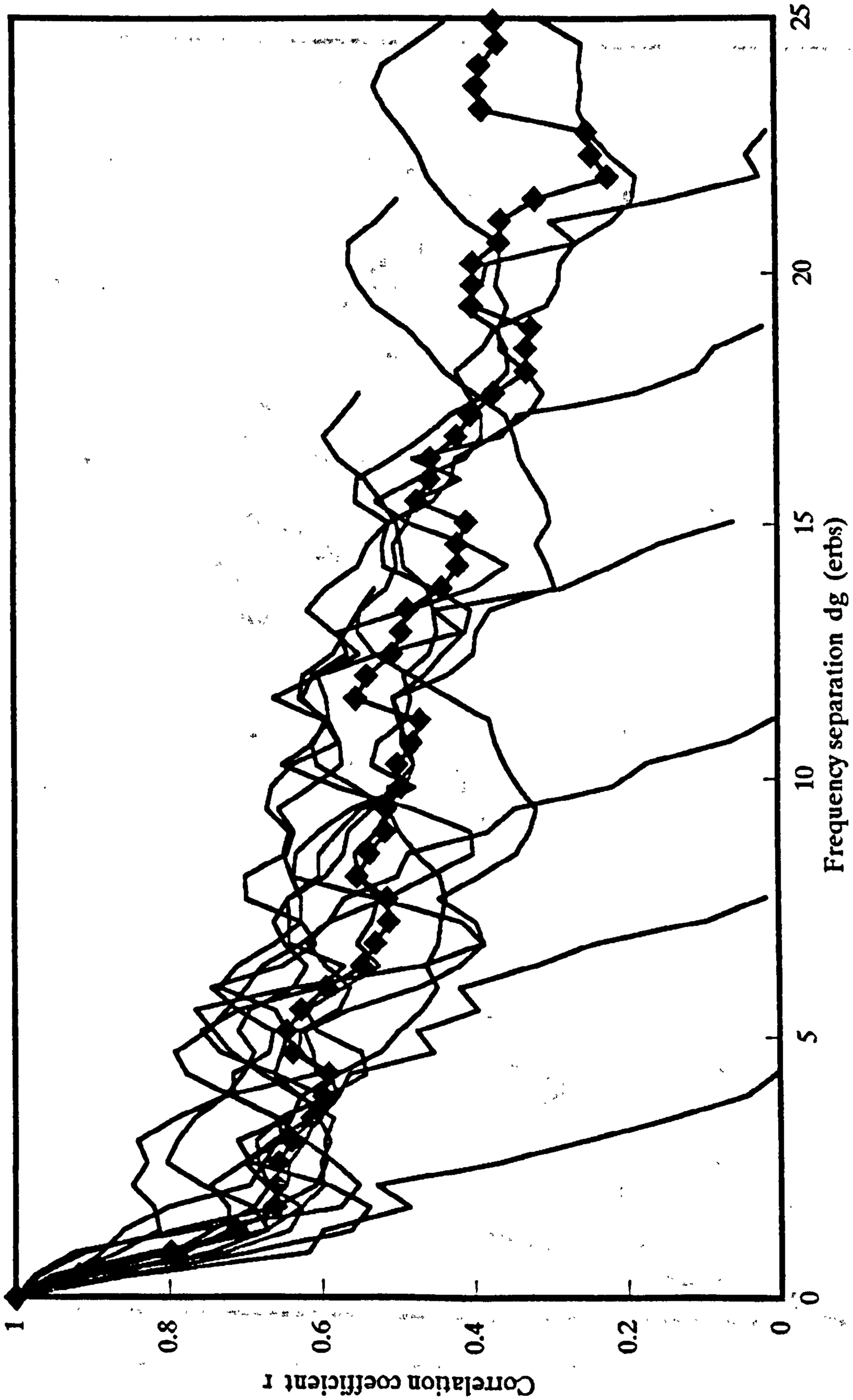

Simultaneous excitation-level correlation
VOC file \voc\brahms4.voc Duration 11.2s
Correlation coefficient r
1
0.8
0.6
0.4
0.2
0
Frequency separation dg (erbs)
0
5
10
15
20
25
Average uses both positive and negative frequency separations.
Selected centre frequency
Average

Figure 6.24

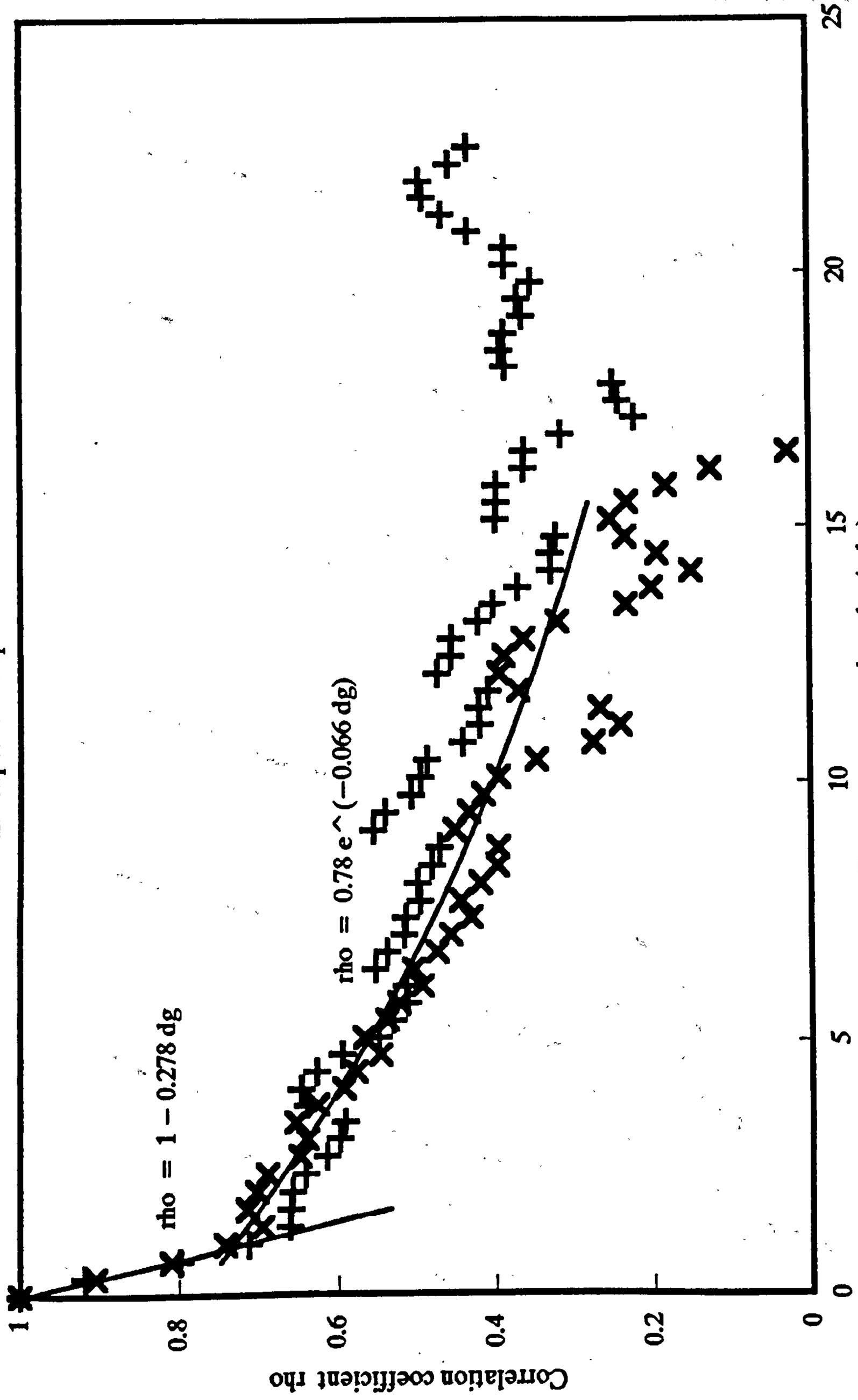

Simultaneous excitation-level correlation
Examples from speech and music
rho = 1 − 0.278 dg
rho = 0.78 e^(−0.066 dg)
bbcnews.voc
brahms4.voc
Fitted lines
Frequency separation dg (erbs)
Correlation coefficient rho
0
5
10
15
20
0.2
0.4
0.6
0.8
1

Figure 6.25

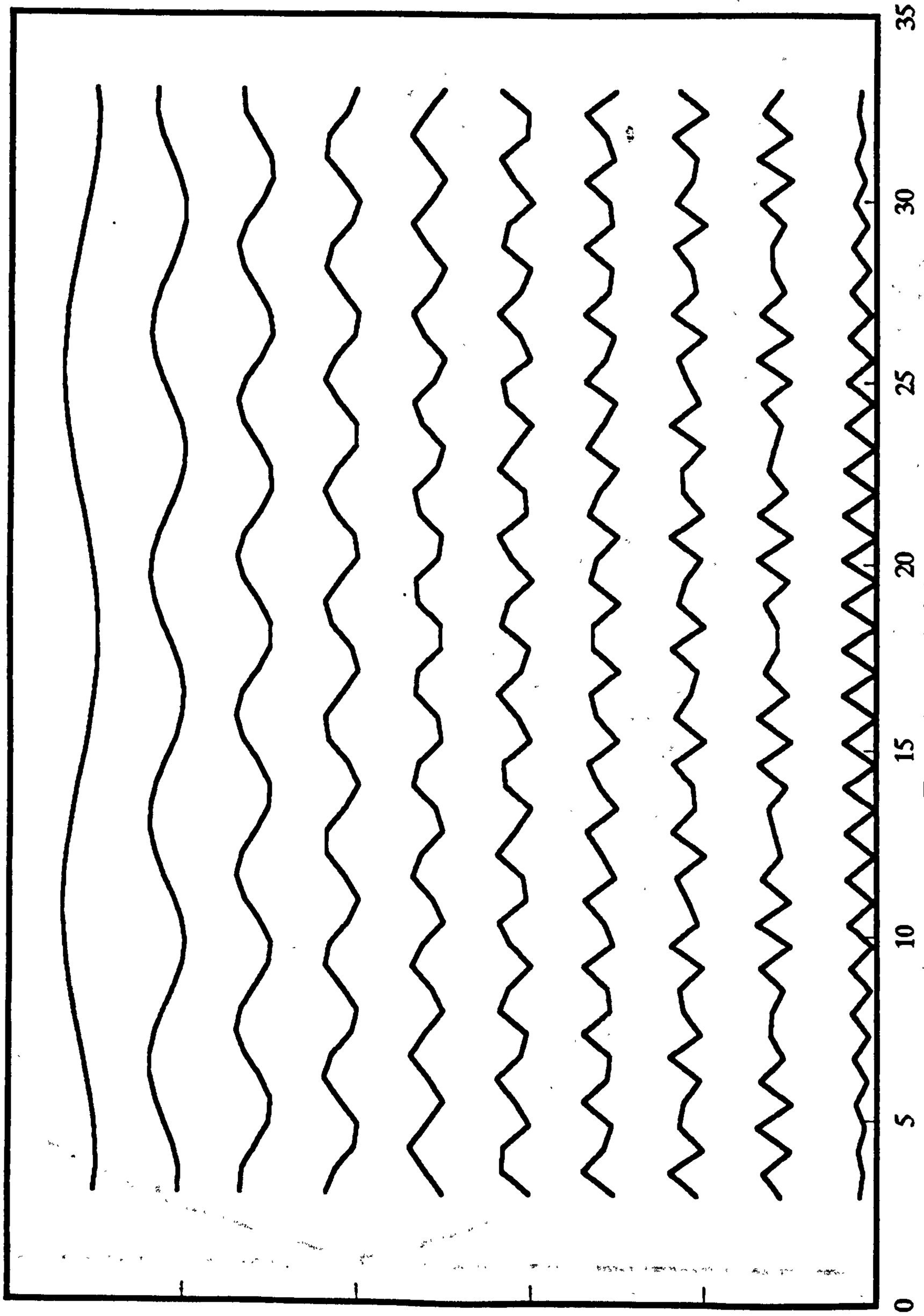

KL basis functions for steady sounds
Excitation pattern sampled at 50 equal intervals from 3 to 33 erbs
5
10
15
20
25
30
Frequency g (erbs)
50
40
30
20
10
0
Function n° (out of 50 – every 5th function shown)

Figure 6.26

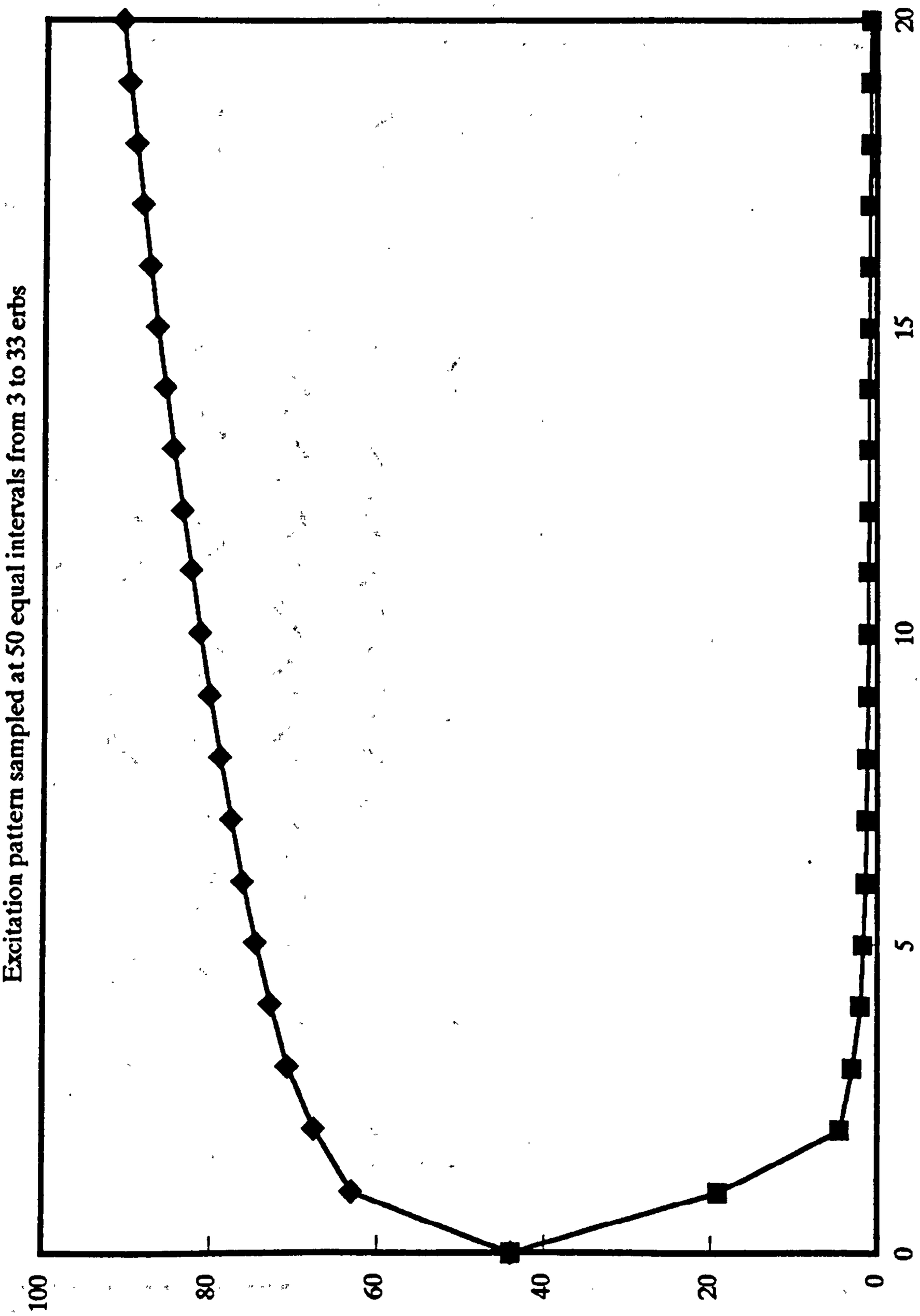

KL transform coefficient variance
Excitation pattern sampled at 50 equal intervals from 3 to 33 erbs
Coefficient n° (out of 50)
Percent of total variance (individual & cumulative)
0
5
10
15
20
0
20
40
60
80
100

Figure 6.27

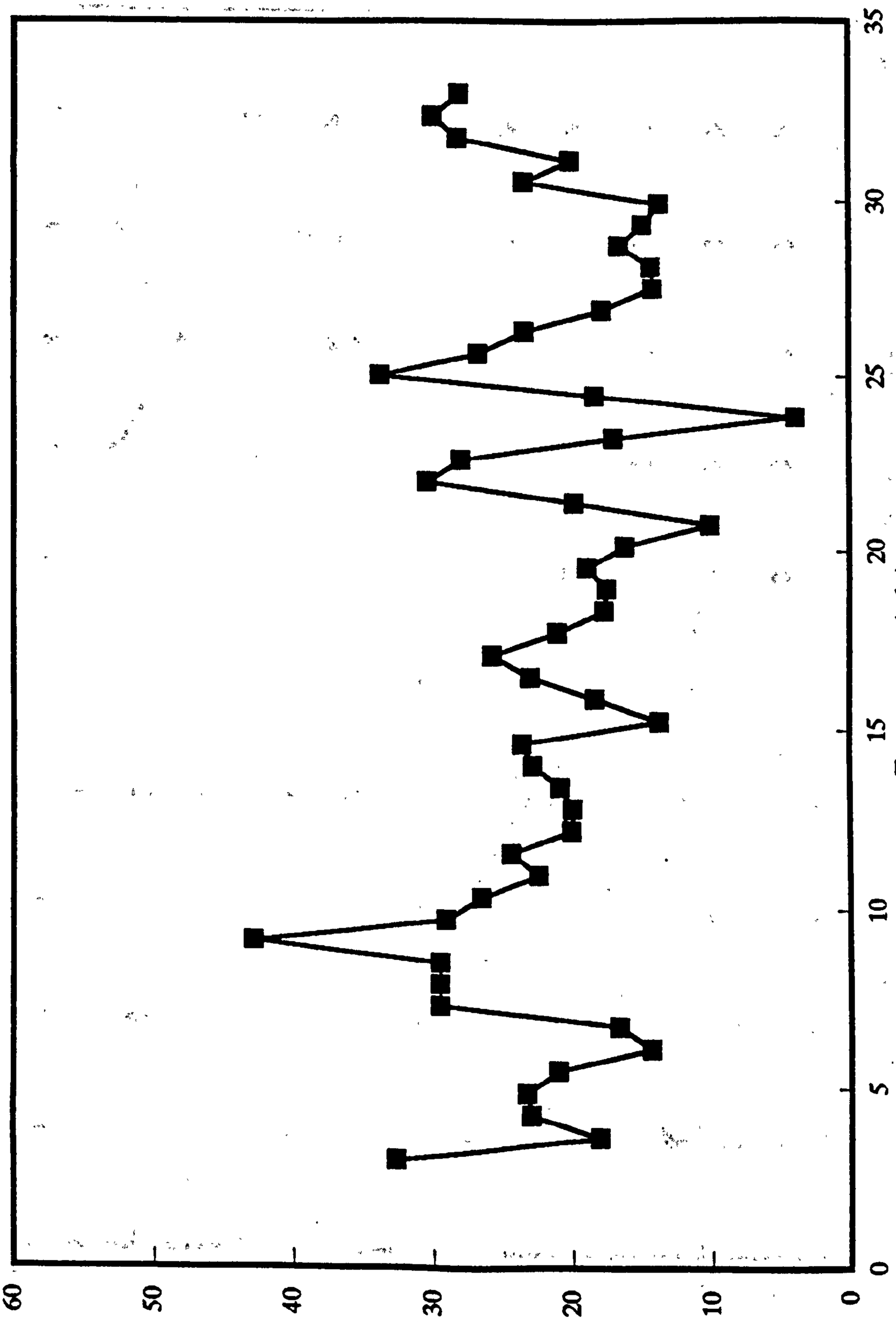
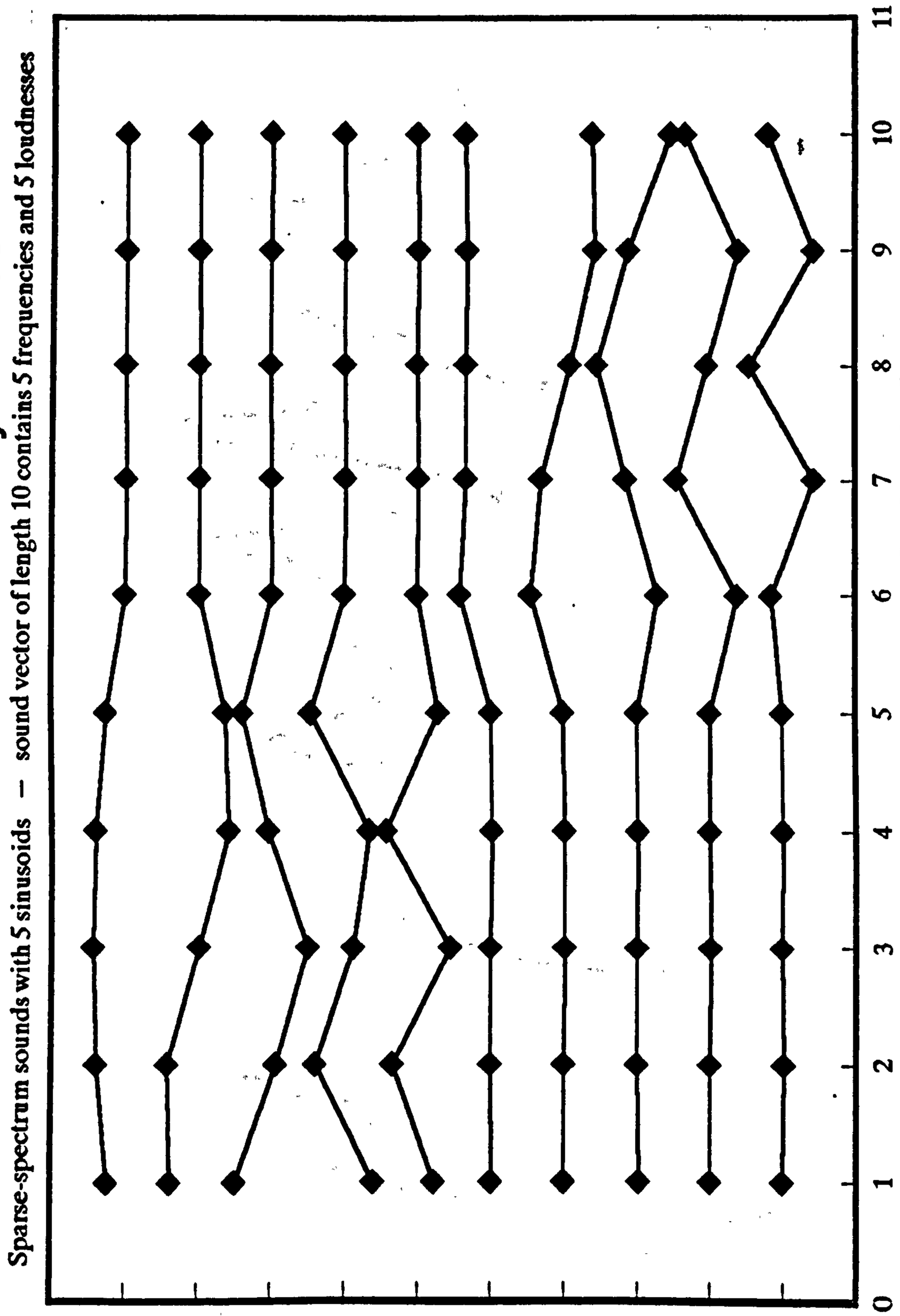

Random excitation pattern
Generated from 50 random numbers and inverse KL transform
Excitation level e (dB)
Frequency g (erbs)
0
10
20
30
40
50
60
0
5
10
15
20
25
30
35

Figure 6.28


KL basis functions for steady sounds
Sparse-spectrum sounds with 5 sinusoids – sound vector of length 10 contains 5 frequencies and 5 loudnesses
Function n°
Frequency n°
Loudness n° + 5
Frequency difference limen GDL = 0.1 erbs
Loudness difference limen dBDL = 3 dB

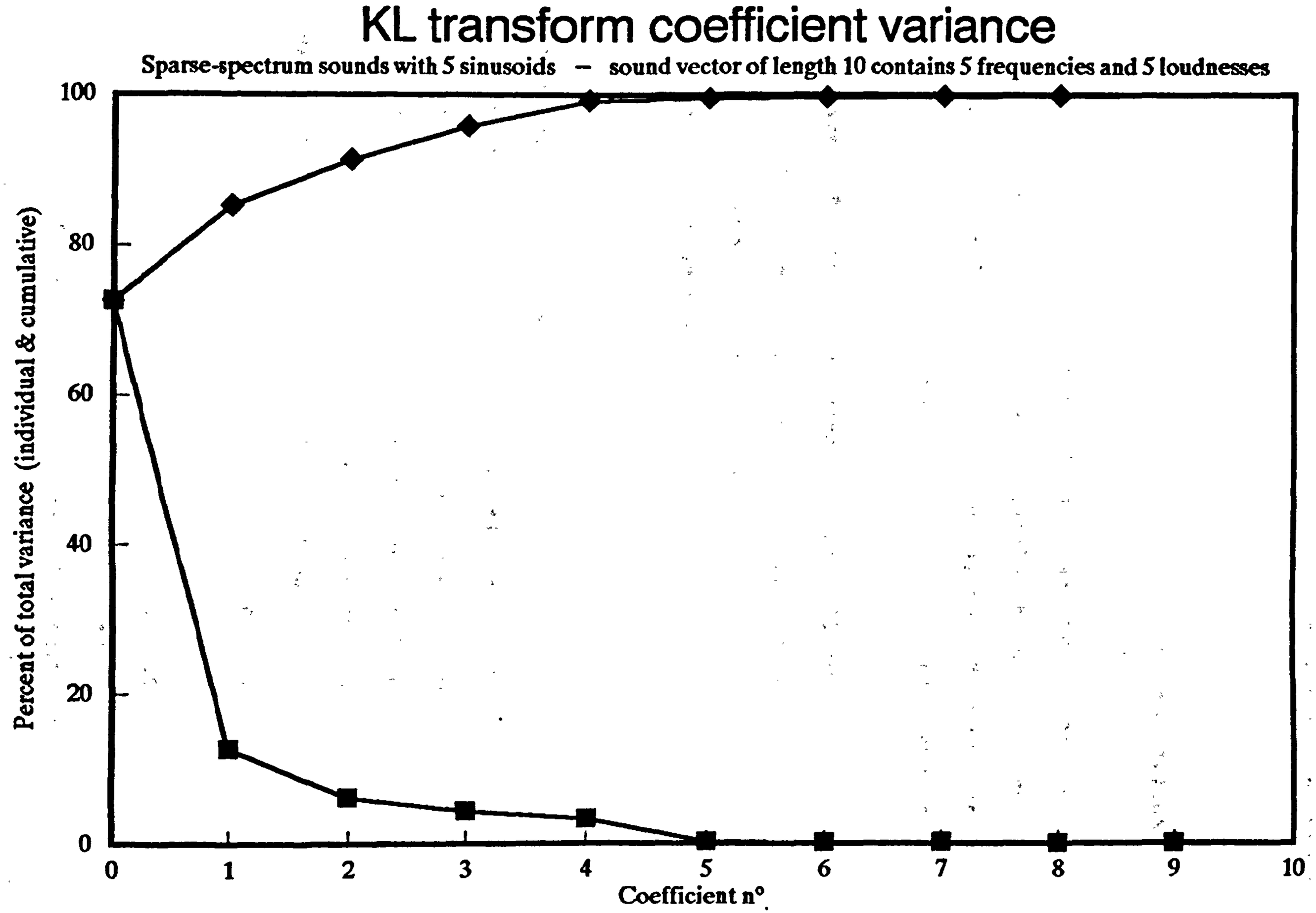

KL transform coefficient variance
Sparse-spectrum sounds with 5 sinusoids – sound vector of length 10 contains 5 frequencies and 5 loudnesses
Percent of total variance (individual & cumulative)
100
80
60
40
20
0
0
1
2
3
4
5
6
7
8
9
10
Coefficient n°
Frequency difference limen GDL = 0.1 erbs
Loudness difference limen dBDL = 3 dB

Figure 6.29

CHAPTER 7 SCHEME 3 - POLAR PIANO TRANSFORM

7.1 Motivation

The first go at the piano transform, described in Chapter 4, was abandoned rather hastily, without wondering whether its deficiencies could be put right. To recap, these deficiencies were

1 Insufficient stressing of edges. The eye is very sensitive to edges, while the ear is sensitive to spectral peaks. Some differentiation or high-passing of the scene seems called for.

2 Very poor resolution. Comparing the auditory ERB of one thirtieth of the auditory bandwidth with the visual peak spatial-frequency response of 8 cycles/degree in say a 120° field of view, the unfoveated piano transform has a solid-angle resolution 1000 times coarser than human vision.

3 Dubious invariance. Take the piano keyboard to be vertical. Invariance to horizontal translation of the scene is excellent - merely a time shift in the sound. Vertical

translation of the scene corresponds to a frequency shift in the sound. Mild shifts produce the famous Donald Duck effect. Shifts of over an octave make sounds very hard to recognise. Divers on helium require special processing of their speech to be understood. Worst still is the effect of size change, causing changes in high frequencies opposite to those in low frequencies.

4 No colour.

For historical accuracy, it should be pointed out that the present work on the piano transform did not develop as an improvement on the cartesian piano transform of Chapter 4 as might be inferred from the above. It was instead the result of "going back to the drawing board" and "starting again from scratch" after the inconclusive results of Chapter 6.

7.2 Remedies

7.2.1 Invariance and resolution

The compromise chosen for the polar piano transform is to map scene size to sound delay, thus obtaining excellent

size invariance. Scene rotation is mapped to sound frequency shift, thus matching those two limited invariances.

Unfortunately, the second sound invariance - invariance to speed of presentation with frequencies unchanged - is left unused by this method. Thus either horizontal or vertical translation of the scene results in a distortion of the sound, though perhaps not sufficient to make an object unrecognisable if it remains roughly centred.

The mapping is achieved by first representing the scene by a standard (r, θ) polar grid of pixels. The scene is therefore circular. For two reasons, it is decided to sound the scene in two halves, first the left and then the right. The first reason, as will be shown below, is that this gives a good balance between resolution and display duration. The second reason is to do with left-right symmetry, which in vision is readily recognisable. A symmetric or near-symmetric scene, such as a face or some letters of the alphabet, may by this method be given either a symmetric sound, where the second half of the sound is a time reversal of the first, or a repeated sound, where the second half is a repeat of the first, such sounds being readily recognisable as such by the listener.

For descriptive purposes, place the origin at the centre

of the scene and start the angle at zero at the x axis, increasing anticlockwise in the usual way. Map scene angle θ of 90° to the highest sound frequency and proceed through increasing angles and decreasing frequencies to the lowest sound frequency at scene angle $\theta = 270°$. Thus only the left half of the scene is mapped to the whole sound freqency range.

The slot scans the scene as shown in Figure 7.1. Start the slot radius r at its maximum value R at the scene circumference. As sound time increases from t = 0, so scene radius r decreases from R until r = 0 when t = T/2, by which time the left half of the scene has been sounded.

Let the unit of measurement be the pixel spacing in the original cartesian scene available for processing. Let the polar resolution initially be isotropic, with

$$\Delta r = r \Delta\theta \qquad (1)$$

If each ring of polar pixels takes the same time Δt to sound, equation (1) implies a logarithmic relationship between slot radius r and sound time t.

With a usual number of pixels in a digital scene and with the sound spectrum specified at a sensible number of frequencies, we have at the circumference $\Delta r >> 1$. As r decreases from R towards 0, it reaches a radius at which

r = 1. From there on inwards, there is insufficient resolution in the original cartesian picture to justify further reduction in slot speed, so inside this circle dr/dt remains constant. In any case, some sort of deviation from an ever diminishing slot speed is required, or the centre would never be reached.

Integrating equation (1) gives the number of radial pixels, which is equal to $T/2\Delta t$ if T is the time taken to sound both halves of the scene. Thus

$$\frac{T}{2\Delta t} = \frac{S}{2\pi}\left(\ln\frac{2\pi R}{S} + 1\right) \qquad (2)$$

where S is the number of circumferential pixels and S/2 is the number of discrete frequencies specifying the sound.

Similarly, the equation giving the progress of the slot in time is

$$r = \begin{cases} \frac{S}{2\pi}\, e^{2\pi\frac{T/2 - t}{S\Delta t} - 1} & t < \frac{T}{2} - \frac{S\Delta t}{2\pi} \\ \frac{T/2 - t}{\Delta t} & t > \frac{T}{2} - \frac{S\Delta t}{2\pi} \end{cases} \qquad (3)$$

where $S/2\pi$ is the changeover radius, at which the polar resolution becomes bigger or smaller than the cartesian resolution.

The derivative of equation (3) gives the inward scanning speed of the semicircular slot.

As promised, we can now examine the trade-off between resolution and display duration. In slot-based schemes, the minimum time necessary to play the slot without loss of information (this is after a long period of training) can't be expected to be less than an erd. This is another invented word intended to be the time equivalent of the erb, with d for duration instead of b for bandwidth, and is the duration of the rectangular time window having the same area as the auditory time window of Moore et al (1988). Integrating equation (8) of Chapter 6 (equation (9) has little effect), the ERD of the time window comes to

$$ERD = t_i^- + t_i^+ = 0.009\ s \quad (4)$$

One erd is therefore defined as a time unit of 0.009 s.

With the pixel numbers of Figure 7.1, let the sound spectrum (for reasons discussed under colour, below) be specified at 73 discrete frequencies, in 6 octaves of 12 semitones, each semitone considered to correspond to one circumferential pixel. So S = 146. With a Δt of 1 erd, the time taken to sound the whole scene, from equation (2), is T = 1.4 s. If on the other hand the scheme had mapped the same frequencies to one whole

circle of pixels, the both radial and circumferential resolution would have been halved, and the scene duration would only have been a quarter as much: 0.35 s, or almost three scenes a second. Similarly, if each of the four quadrants had been mapped on to the whole frequency range, with a quartercircular slot swept inwards four times to sound the whole scene, resolution would have been doubled in each direction and the whole scene would have taken 5.6 s.

Figure 7.2 shows the raw cartesian scene used, a face called Shanti. Inset panels show various stages of computation. Panel 1, the top left panel, shows the geometric side of affairs, namely the mapping described so far. The ordinate of panel 1, as of all the other inset panels, is frequency and the abscissa is time, but the content of panel 1 is still recognisably the scene.

The reason for presenting the panel as an inset to the raw cartesian scene is to show it at the same scale in terms of resolution, with one polar pixel in the panel the same size as one cartesian pixel in the main picture. Thus while the cartesian scene is 480 pixels high, the panels are only 73 notes high.

For discussion, panel 1 is reproduced enlarged as Figure 7.3. A vertical line down the centre of the figure corresponds to a single point at the polar origin.

The horizontal line from top left corner to top dead centre is identical to the line from top right corner to top dead centre, and corresponds to the radius at $\theta = 90°$. Similarly, the horizontal line from bottom left corner to bottom dead centre is identical to the line from bottom right corner to bottom dead centre, and corresponds to the radius at $\theta = 270°$.

As it is, the second half of the polar scene is scanned from the centre outwards. Since the scene is nearly symmetrical, the second half of the sound is nearly a time reversal of the first. (By scanning the second half of the scene inwards, it can be easily arranged, instead, for the second half of the sound of a symmetrical scene to be a repeat of the first.)

The sound is invariant to scene size in that a smaller face would merely delay the first half of the sound and bring forward the second half. (In the case of both halves of the scene being scanned inward, both halves of the sound would be delayed.)

For future reference, panel 1 consists of numbers in the range 0 to 1.

7.2.2 Stressing of edges

Going downwards from top left, the second panel of Figure 7.2, reproduced as Figure 7.4, is a differentiation of the first panel, a procedure intended to correspond to the linear rise from zero of the start of the spatial frequency sensitivity curve of Figure 6.2. Edges are shown black, brightness gradients grey.

Panel 2 is then blurred to produce panel 3, reproduced as Figure 7.5, intended to be a measure between 0 and 1 of fineness of scale, with a high response corresponding to areas of detail. The use of this will become clear further on.

7.2.3 Colour

The colour mapping tried out here is a mapping from hue to musical key. Now hue is a one-dimensional circular thing (Figure 2.5), while musical keys lie on a torus (Figure 2.6). There is one circular way to list musical keys so that they are all used up. Start from C in the top left corner of Figure 2.6 and proceed through e and G down to E on the bottom line. This E reappears next to the original C on the top line. Carry on in the same direction down to G♯ on the bottom line, and so on. It only takes three circuits of the torus to come back to

the original key of C.

Figure 7.6 shows how this circular list of keys may be mapped in a continuous way to hue. Panel 4 (Figure 7.7) shows the result of converting hue to note loudness according to the principles of Figure 7.6. The calculation in fact differs from Figure 7.6 in two ways. First, the hue used is Oleari's hue, so the colours on the bottom row of Figure 7.6 are not necessarily in the right place, either as concerns relative spacing or the anchoring of green to C major. Second, the result of the mapping of Figure 7.6, which is a weighting from 0 to 1 for each note, is further multiplied by the saturation (also a value from 0 to 1) to produce the weightings of Figure 7.7.

For any hue, three notes per octave are sounded, so a coloured area in the scene has to occupy an octave or more of the slot in order to impart its hue. Note that an octave is three or four times coarser than an erb (Figure 3.2), and chromatic resolution is three or four times coarser than achromatic resolution (section 2.6.2). For areas smaller than this, calculation of hue is not helpful. This is the reason for the measure of fineness of scale constituting Figure 7.5, which works as follows.

Panel 5 (Figure 7.8) is simply the maximum of panels 3 and 4 (Figures 7.5 and 7.7). The result is a loudness

weighting w from 0 to 1 emphasising contrast in areas of detail and colour in areas of less detail.

Panels 6 and 7 (Figures 7.9 and 7.10) are almost final spectrograms for the left ear and right ear respectively, calculated as follows. Let b be a number in panel 1 (for "brightness"). Let l and r refer to panels 6 and 7. Then $l = w(1 - b)$ and $r = wb$. Alternatively, if z is a zero-mean version of panel 1, then $l = w(0.5 - z)$ and $r = w(0.5 + z)$.

Finally, panels 6 and 7 are scaled for overall loudness and frequency-dependent "pre-emphasis" according to the dBA scale (Figure 7.11) before being sounded.

7.2.4 Implementation

See program \cwork\progs\soundpic.c.

Polar piano transform

Radial contraction scan of semicircular slot

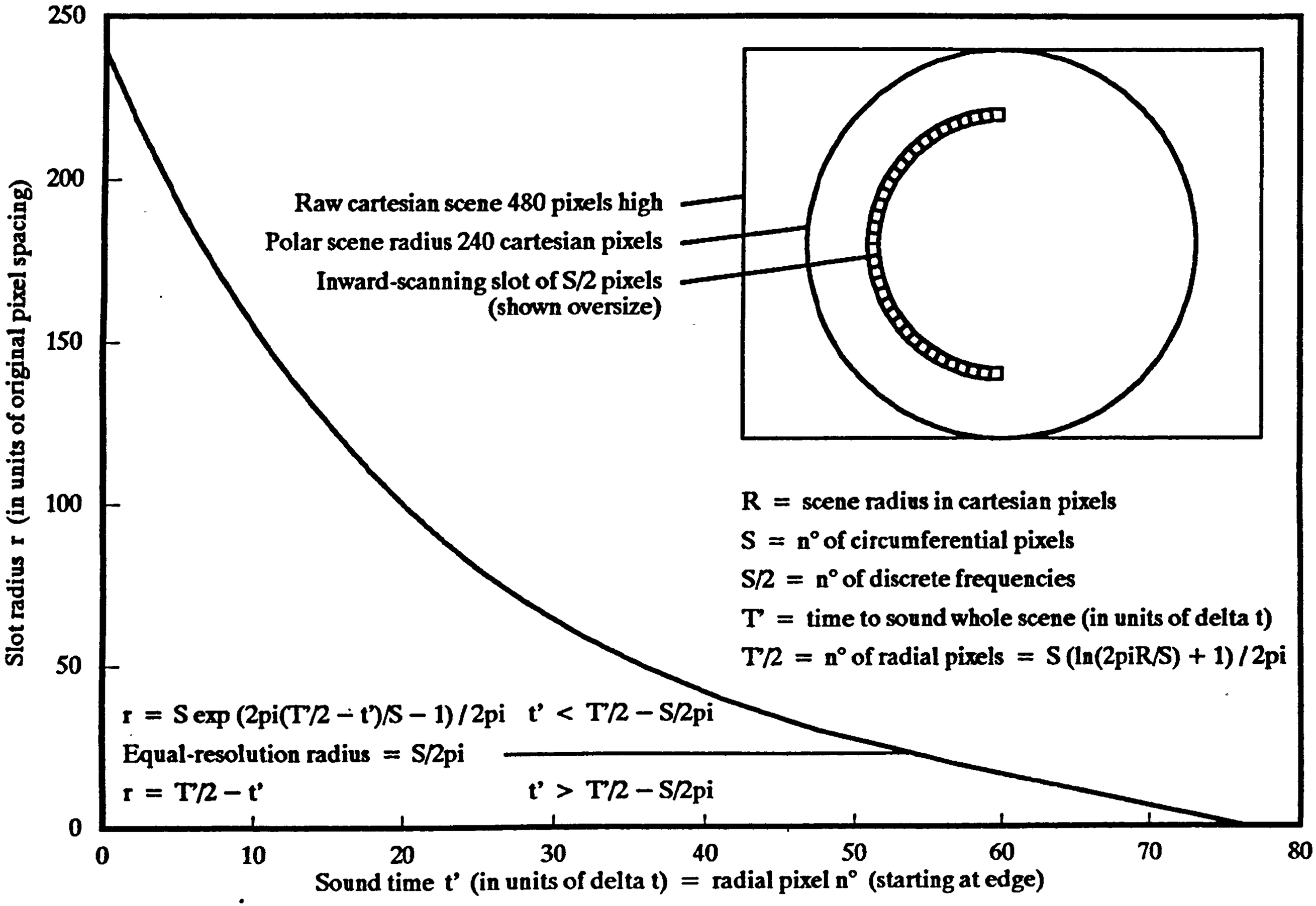


Figure 7.1

Figure 7.2

Figure 7.3

Figure 7.4

Figure 7.5

Figure 7.6

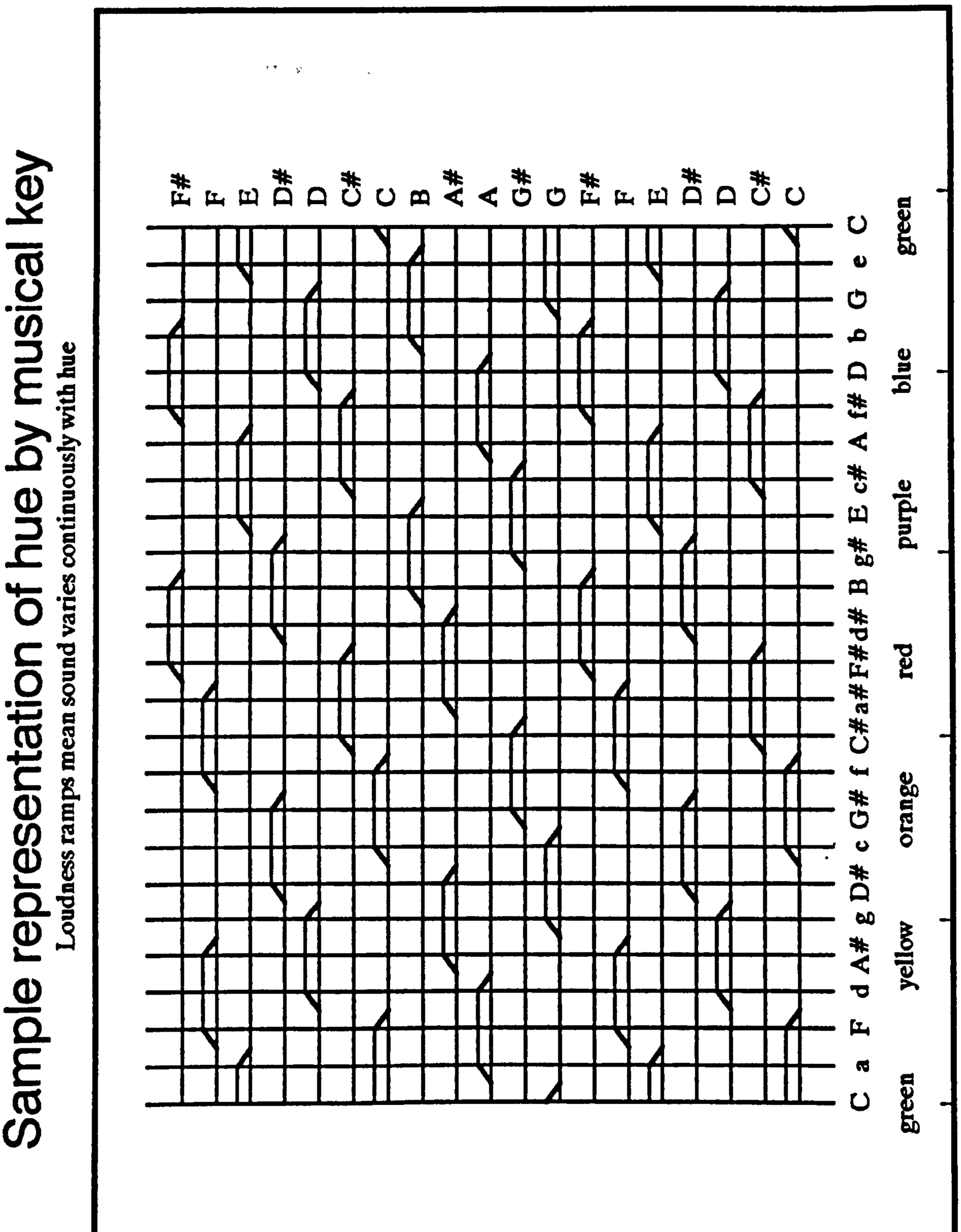

Sample representation of hue by musical key
Loudness ramps mean sound varies continuously with hue
F# F E D# D C# C B A# A G# G F# F E D# D C# C
C a F d A# g D# c G# f C# a# F# d# B g# E c# A f# D b G e C
green yellow orange red purple blue green
Hue and musical key (small letters refer to minor keys)
Musical note

Figure 7.7

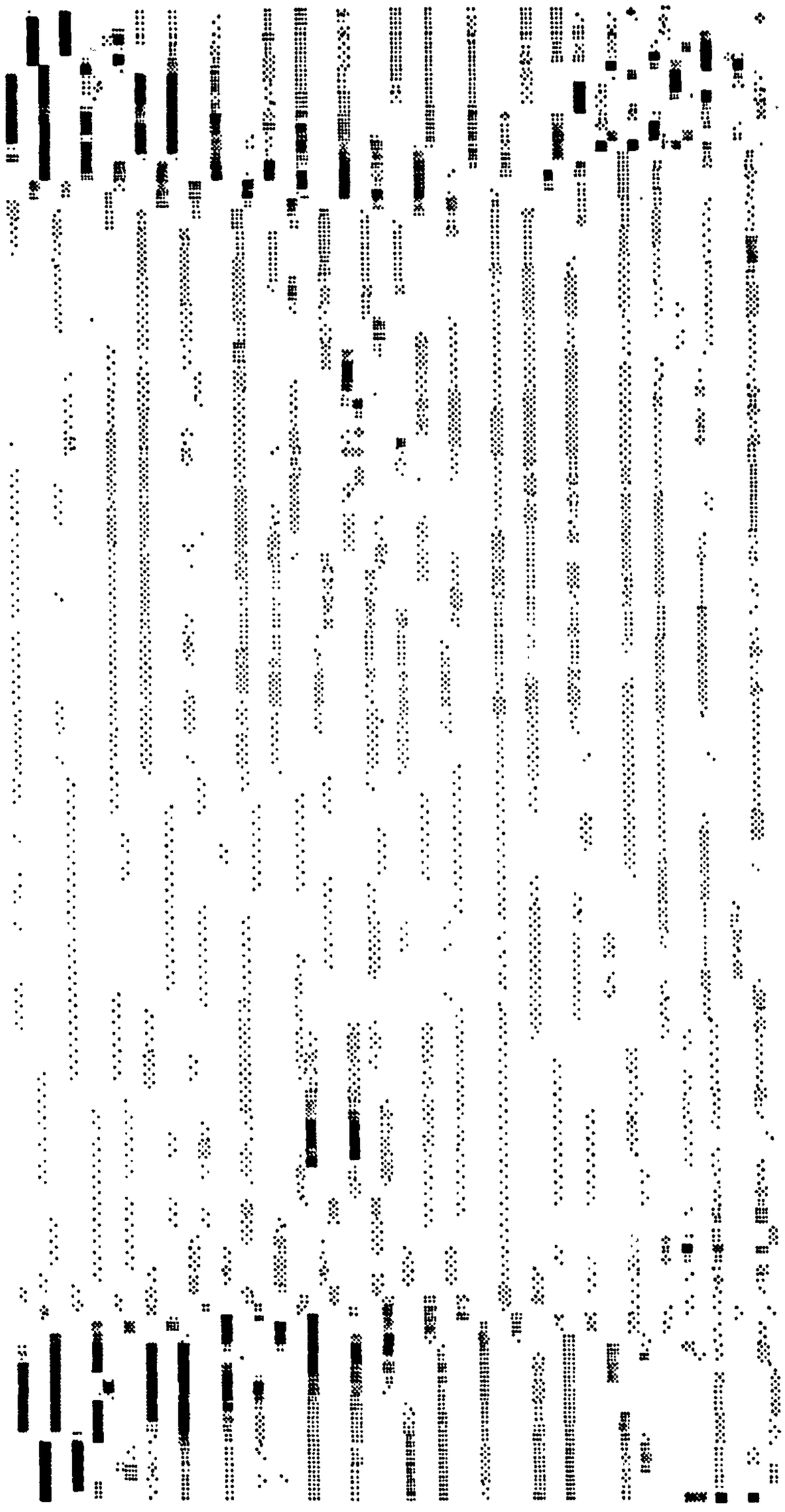


Figure 7.8

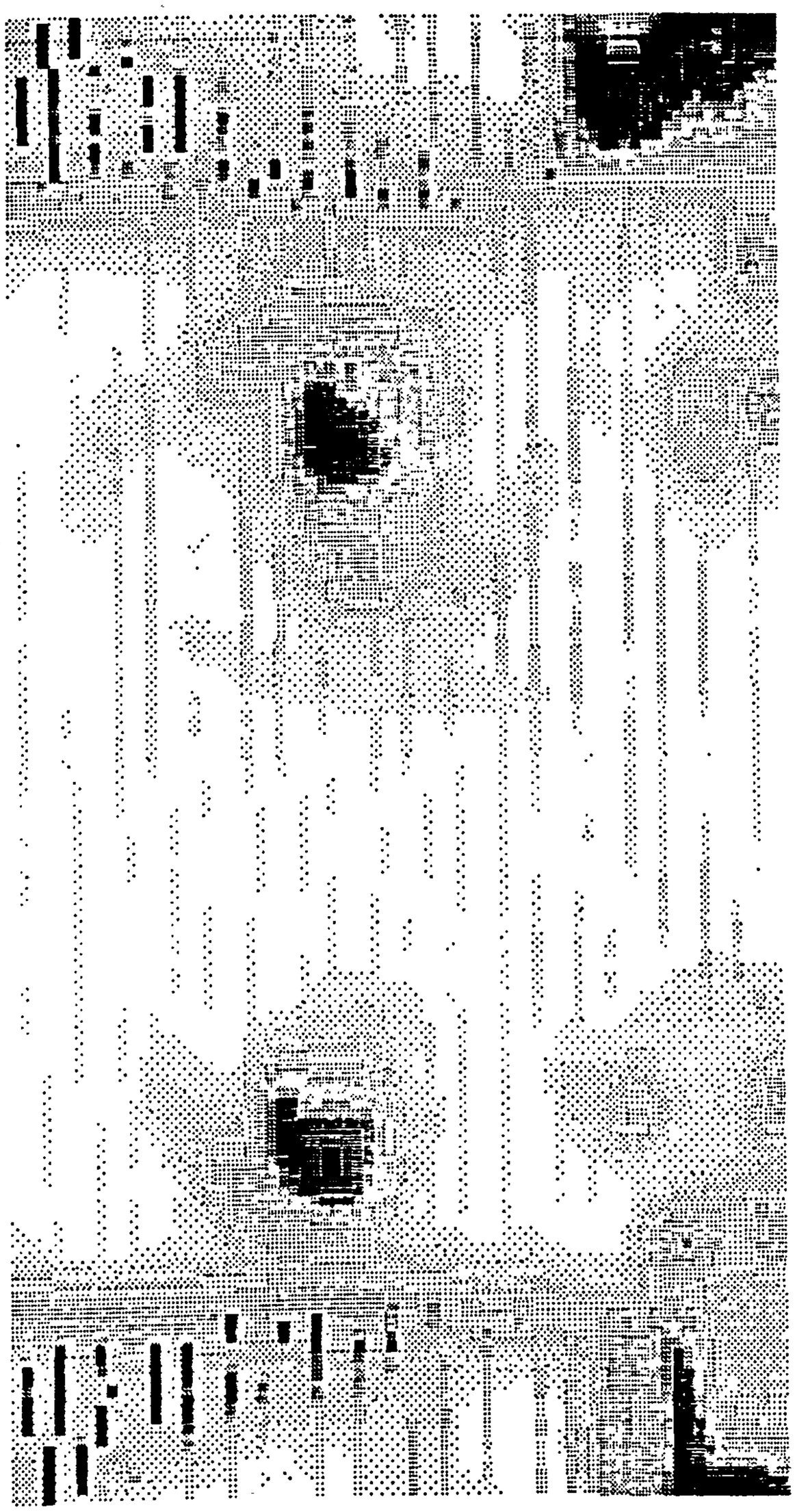
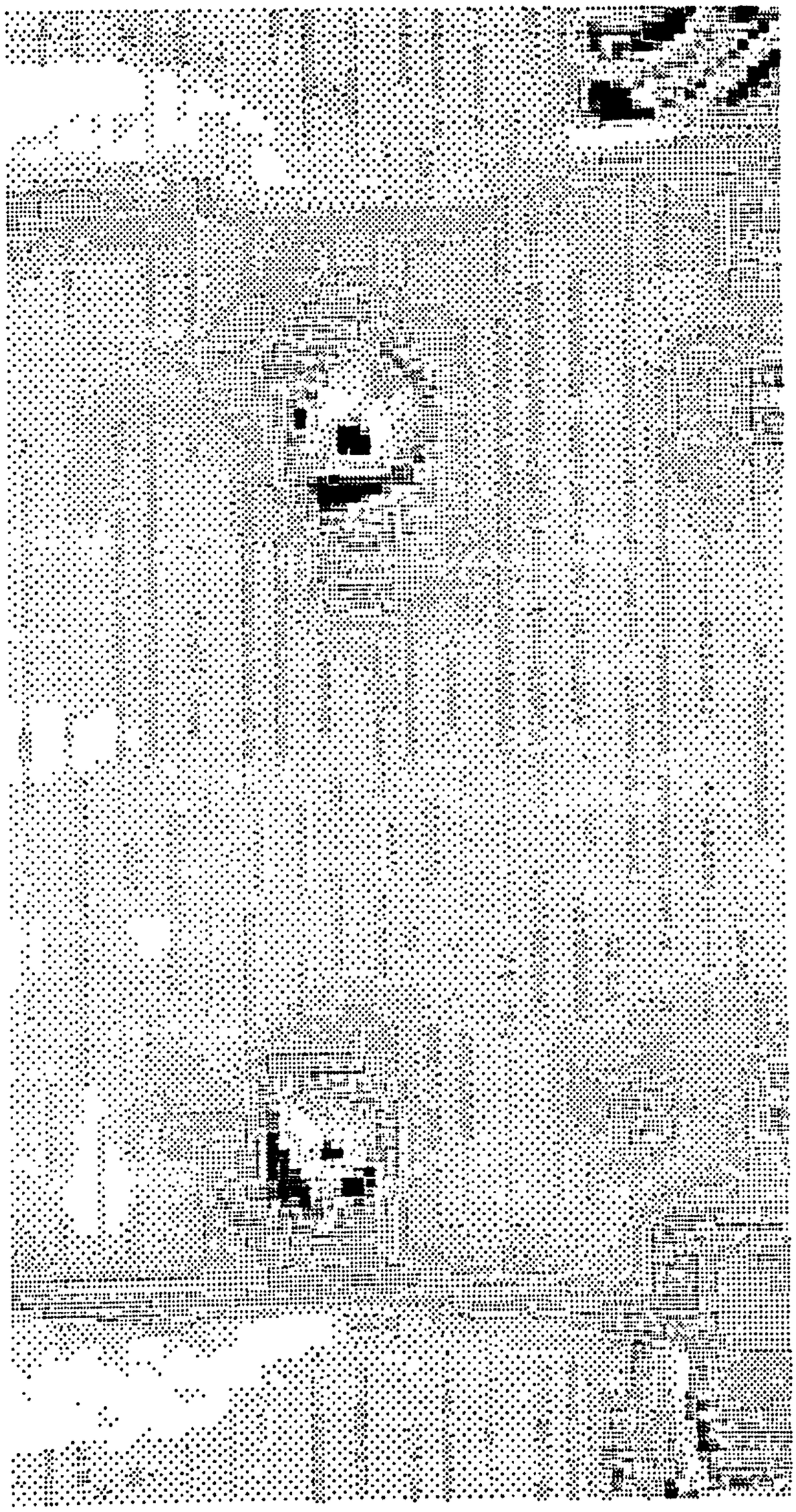


Figure 7.9

Figure 7.10



dBA weighting and equation

Source: Haughton (1980)

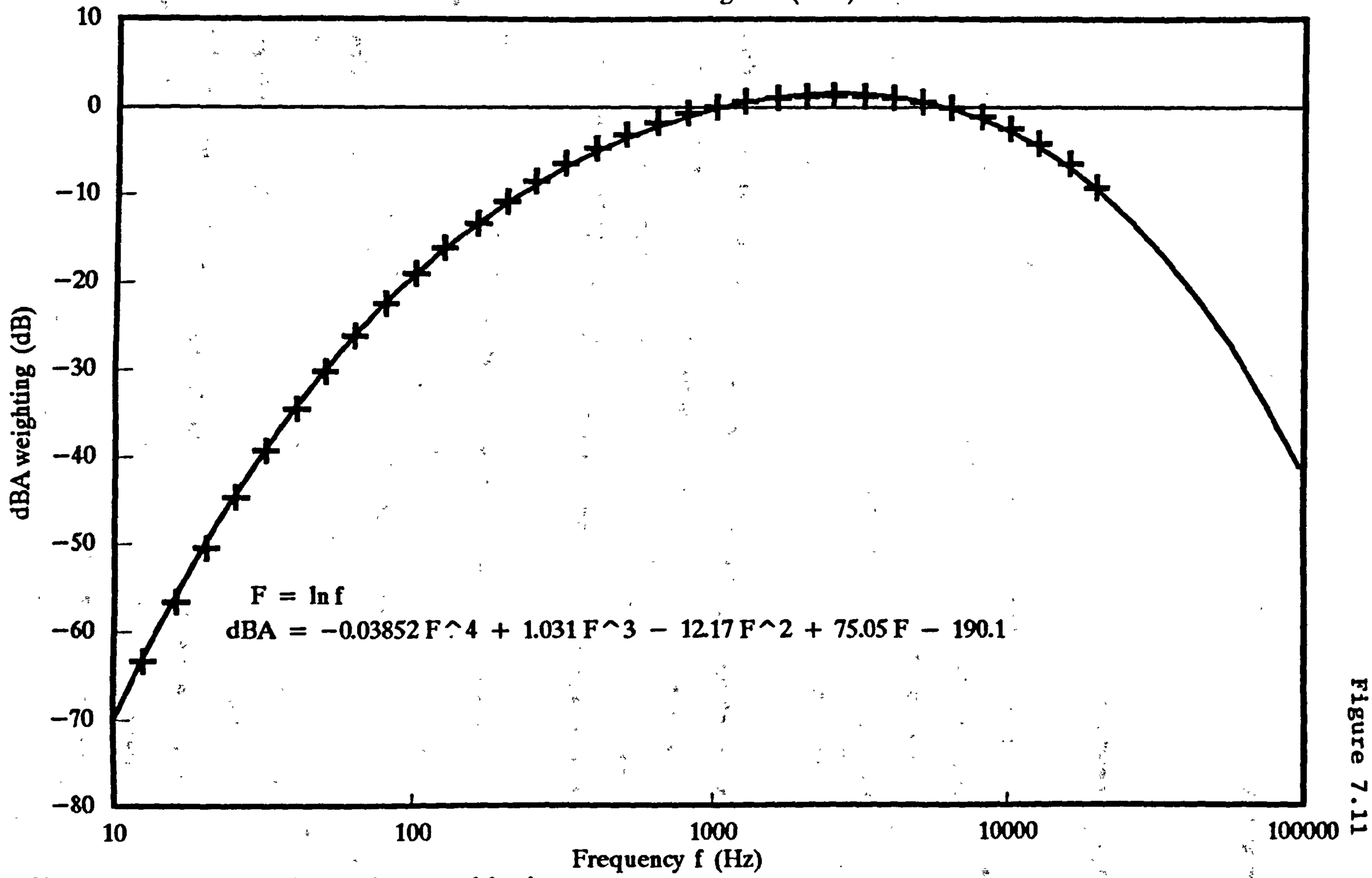


Figure 7.11

Use of fewer significant figures gives surprising inaccuracy.

CHAPTER 8 NUMBER OF NUMBERS REQUIRED TO SPECIFY A SPECTRUM

8.1 Preamble

For the purpose of our next scheme, described in Chapter 9, we need for economy to know the smallest number of numbers required to specify a sound spectrum for human consumption, that is to say without noticeable degradation.

A spectrum can be specified either as a set of loudnesses at a prearranged set of frequencies, or as a number of frequency-loudness pairs. There are other ways, but these are ways associated with known difference limens.

The first of these two ways was used in Chapters 6 and 7, where the spectrum was specified by 100 numbers for the purposes of analysis (Figures 6.16 to 6.19), by 50 numbers for the purposes of demonstration (Figures 6.25 to 6.27), and by 73 numbers for the purposes of synthesis (inset panels of Figure 7.2). Our ultimate purpose is synthesis. We do not want to degrade the scene more than the user's ears will. On the other hand, overspecification of the spectrum will merely raise the cost of the optophone to no good effect.

8.2 Argument based on correspondance of erbs and erds

Several different lines of argument come to mind. First, compare the question to the similar one in the time dimension. If it is thought unnecessary to specify the spectrum more often than once per erd (section 7.2.1), then, having regard to the similarities between the time and frequency dimensions (Figures 6.12 to 6.15), it should be unnecessary to specify the spectrum more often than once per erb, which if the spectrum is 30 erbs wide requires some 30 numbers.

8.3 Argument based on frequency difference limens

Second, it might be thought appropriate to relate the frequency spacing to the frequency difference limen. There appear at first to be three different frequency difference limens, but luckily they can be reduced to a single frequency difference limen which is a function of the duration and bandwidth of the sound whose frequency is being varied. The three are as follows.

1 The frequency difference limen of a single pure tone. This is remarkably small, as low as 0.2% in the range 300 to 3000 Hz (Figure 8.1), provided it is sounded for 0.1 s or more. This method would require some 1150 numbers in this range alone. The

single pure tone frequency difference limen jumps significantly at around 5 kHz, above which frequency is not tracked by phase locking of aural pulse trains (Moore 1989).

Program: \lwork\fdl.wk3.

2 The frequency difference limen of a single narrow noise band. A narrow noise band sounds like an unsteady pure tone. Moore (1973) found that at centre frequencies of 2 kHZ and 4 kHz, the frequency difference limen of a single narrow noise band was higher than that of a pure tone, but by a factor of less than 2 (compare curves with points on left axis of Figure 8.2). At 6 kHz, there was negligible difference, again explained by the absence of phase locking. Moore's stimuluses (the English plural is deliberate) lasted 100 ms. The differences between Figure 8.1 and the points on the left axis of Figure 8.2 are not explained, and probably only indicate the use of different subjects.

The formants (spectral resonance peaks) of Gagné & Zurek (1988) produced with a white noise source also fall in the category of a narrow noise band. Plotted as the solid square points in Figure 8.2, the centre-frequency difference limens show no trend associated with centre frequency from 300

to 2000 Hz, provided the DLs are expressed in erbs.

Program: \lwork\formant.wk3.

3 The frequency difference limen of a voiced formant (formant with periodic source). The difference here is the presence of other sinusoids associated with the source. The pitch of the sound is that of the source, whereas with a white noise source the pitch of the sound is that of the formant frequency. The greater scatter of Gagné & Zurek's results with a periodic source is ascribed to the coincidence or otherwise of one of the sinusoids with the formant centre frequency.

The important thing to note however is that there is no overall difference between the centre-frequency difference limens of narrow noise bands and of voiced formants, meaning that the presence of components other than at the frequency being varied does not necessarily impair detection of change.

The formant frequency difference limen (FFDL) was found by Gagné & Zurek (1988) to depend on the width of the formant according to the formula

$$FFDL = 0.079\frac{F}{\sqrt{Q}} \qquad (1)$$

where F is formant resonant frequency in Hz and Q is the parameter in the resonant filter with response |H(f)| given by

$$|H(f)| = \frac{1}{\sqrt{(1 - (f/F)^2)^2 + (f/QF)^2}} \quad (2)$$

For our more general purposes we want FFDL in terms of the resonant freqency and the width of the formant. The formant half-power width W in Hz is given approximately by

$$W = F/Q \quad (3)$$

Substituting,

$$FFDL = 0.079\sqrt{WF} \quad (4)$$

This equation is not satisfactory as it has FFDL going to zero as the formant bandwidth W goes to zero, and so can only locally be true.

On a more natural erb scale, if FEDL is the formant erb difference limen and B is the half-power width of the formant in erbs, then

$$FEDL = \frac{FFDL}{ERB} \quad (5)$$

and

$$B = \frac{W}{ERB} \qquad (6)$$

with ERB given by equation (20) of Chapter 6.

As noted above, expressing the difference limen in erbs eliminates dependence on frequency, at least in the range of Gagné & Zurek's tests. The resulting equation (Figure 8.2) is

$$FEDL = \Delta G = 0.209\, B^{0.467} \qquad (7)$$

Suppose the spectrum to be specified consists of spectral peaks 1 erb wide. Then the relevant difference limen (Figure 8.2) is about 0.2 erbs, and if the whole spectrum is 30 erbs wide then a total of 150 numbers is required to specify it. This doesn't appear a very sensible arrangement, since if a number is available every 0.2 erbs, one ought on the Nyquist argument be able to specify spectral peaks every 0.4 erbs. If the peaks are 0.4 instead of 1 erb wide, the relevant difference limen is no longer 0.2 erbs. In fact, from equation (7), the difference limen which gives a bandwidth of two DLs is equal to 0.1 erbs, and the whole spectrum would then require 300 numbers.

Program: \lwork\formant.wk3.

8.4 Argument based on spectral modulation depth

Third, there might be some clue in comparing the depth of modulation in the excitation pattern caused by sinusoids of different spacings to the intensity difference limen of 1 or 2 dB. This is done in Figures 8.3 and 8.4. Sinusoid spacings for these two difference limens fall closely on either side of 1 erb, which could well mean something. This method brings us back down to 30 numbers.

Program: \1work\specmod.wk3.

8.5 Argument based on information theory

8.5.1 General

Fourth, because of the blurring effect of the masking pattern, there ought to be a spacing below which nothing more is to be gained in the way of information.

Figure 8.5 shows the relation between the range of a variable expressed in difference limens and the information contained in knowledge of its value. Both Gaussian and rectangular distributions are considered, but the equation for the Gaussian case breaks down near

the origin. For calculation purposes a rectangular distribution will be assumed.

In the case of a number of such variables, as contained in the PR of a sound, the simple way of calculating the total information, by summing the information in each number (or multiplying by the number of numbers if they are equally informative), is invalid because the numbers are correlated. The correlation of the elements of a sound vector was examined in Chapter 6 and found to follow the equation in Figure 6.24. This equation was used in Chapter 6 to decorrelate a 50-element sound vector.

The solid squares in Figure 8.6 show further work of this nature, using sound vectors from 10 to 150 numbers long. The hope was that the information contained in these vectors would level off beyond a certain vector length, corresponding hopefully, in view of Figures 8.3 and 8.4, to a spacing somewhere in the region of 0.5 to 1 erb. Unfortunately, nothing of the sort happens (solid symbols, Figure 8.6), and this is put down to inaccurate modelling of the correlation coefficient at close spacings (equation (30) of Chapter 6).

8.5.2 Fine correlation of auditory excitation pattern

An improved description of the correlation of excitation patterns at close spacings can be derived as follows. Consider a sound spectrum specified by a vector **s** in decibels, **n** numbers long. The PR of this is a blurred and scaled version of **s**, with the blurring caused by the width of the auditory filter. Let **F** be an **n** x **n** matrix with an auditory filter in each row, centred on the frequency corresponding to the row number (that is, centred on the diagonal element). The elements of **F** are given by the equation for PA in Figure 3.1, multiplied by the local element spacing in Hz and normalised to have a total (not peak) weighting of 1. Now PA stands for power attenuation, and **F** can only operate on a sound vector s_p in units of sound power (W/m^2). The result of applying the filter, also in units of sound power, is

$$e_p = F s_p \qquad (8)$$

which can then be expressed in decibels if required.

If the correlation matrix of s_p is R_{sp}, then the correlation matrix R_{ep} of e_p is (Pratt 1978)

$$R_{ep} = F R_{sp} F^{*T} \qquad (9)$$

where superscript * denotes complex conjugation and T transposition. Since **F** is real, the * need not concern

us.

We are interested in the correlation at close quarters of e_p, where auditory blurring is dominant. A good approximation and lower bound on the correlation at close quarters of e_p is therefore obtained by taking as R_{ep} the unit matrix I, giving

$$R'_{ep} = F F^T \tag{10}$$

with each element of R'_{ep} given by

$$R'_{ep}(i, j) = \sum_{k=1}^{n} F(i, k)\, F(j, k) \tag{11}$$

where i and j refer to the two values of f_{out} on which two filters are centred, and k refers to all the frequencies f_{in} at which the sound has energy.

From equation (11) it is clear that the correlation coefficient between the excitation-pattern power at any two close frequencies f_a and f_b is given by

$$\rho_p = \int_{-\infty}^{\infty} W_a\, W_b\, df \tag{12}$$

where

$$W_a = \left(1 + \frac{4|f_a - f|}{ERB_a}\right) e^{-\frac{4|f_a - f|}{ERB_a}} \tag{13}$$

with ERB_a given by f_a in equation (20) of Chapter 6, and similarly for W_b. Note that the skirt (w in equation (17) of Chapter 6 or t in Figure 3.1) is irrelevant and set to zero.

Because of the discontinuity in (13), it is necessary to partition equation (12). Since (12) is commutative in f_a and f_b, assume without loss of generality that $f_a < f_b$. Equation (12) then becomes

$$\rho_p = \int_{-\infty}^{f_a} W_a^- W_b^- \, df + \int_{f_a}^{f_b} W_a^+ W_b^- \, df + \int_{f_b}^{\infty} W_a^+ W_b^+ \, df \qquad (14)$$

where

$$W_a^- = \left(1 + \frac{4(f_a - f)}{ERB_a}\right) e^{-\frac{4(f_a - f)}{ERB_a}} \qquad (15)$$

and

$$W_a^+ = \left(1 + \frac{4(f - f_a)}{ERB_a}\right) e^{-\frac{4(f - f_a)}{ERB_a}} \qquad (16)$$

and similarly for W_b^- and W_b^+.

In general,

$$\int (1 + r(f-p))\, e^{-r(f-p)} \; (1 + s(f-q))\, e^{-s(f-q)} \, df = (Pf^2 + Qf + R)\, e^{sf+T} \qquad (17)$$

where

$$p = \frac{rs}{S} \tag{18}$$

$$Q = -(1 + P(p + q + 2/S)) \tag{19}$$

$$R = -\frac{Q + T + pqrs}{S} \tag{20}$$

$$S = -(r + s) \tag{21}$$

and

$$T = pr + qs \tag{22}$$

Using subscripts 1, 2 and 3 for the three integrals in equation (14), we have

$$p = f_a$$
$$q = f_b$$

$$r_1 = -\frac{4}{ERB_a} \qquad r_2 = \frac{4}{ERB_a} \qquad r_3 = \frac{4}{ERB_a} \tag{23}$$

$$s_1 = -\frac{4}{ERB_b} \qquad s_2 = -\frac{4}{ERB_b} \qquad s_3 = \frac{4}{ERB_b}$$

The resulting excitation-power correlation coefficient ρ_p is shown as a function of frequency separation as the

lowest curve in Figure 8.7.

Program: \cwork\progs\speckl.c, \lwork2\tiprho.wk3.

8.5.3 From power correlation to decibel correlation

Unfortunately, the psychophysical representation of a sound, whose information content can be calculated, is in decibels, and we need to know the correlation at close quarters of the PR. The relationship between the correlation of two numbers and the correlation of their decibels is not obvious and was examined by generating correlated pairs of numbers and measuring the correlation of their conversions.

For two reasons, the generation was done in decibels, and the sound-power correlation measured, not the other way round. First, it guarantees positive sound powers and thus allows negative numbers to be generated with impunity. Second, and more important, is the feeling that the normal or rectangular probability distribution used to generate the numbers fits decibels better than W/m^2.

The correlated pairs of numbers in dB were generated as follows. First 1000 values of x were generated from a rectangular distribution with preset mean of 40 dB and

range r of 40 dB. The values thus ranged from 20 to 60 dB. The standard deviation σ is related to the range by

$$\sigma = \frac{r}{2\sqrt{3}} \tag{24}$$

Next, for each value of x a value of y was calculated from (Kottegoda 1980)

$$y = \mu + \rho_d(x-\mu) + \sqrt{1-\rho_d^2}\,\epsilon \tag{25}$$

where ϵ is a random variable with zero mean and standard deviation σ given by equation (24). Equation (25) is a formula which gives a correlation coefficient between x and y of ρ_d and which gives y the same mean and standard deviation as x. The correlation coefficient was checked using (Paradine & Rivett 1964)

$$\rho_d = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{(n\Sigma x^2 - (\Sigma x)^2)\ (n\Sigma y^2 - (\Sigma y)^2)}} \tag{26}$$

The next step was to convert x to sound power in W/m² by the standard formula

$$p_x = 10^{x/10 - 12} \tag{27}$$

and p_y similarly obtained from y, and the correlation coefficient ρ_p calculated by using p_x and p_y in

equation (26).

The operation was repeated for a large number of values of ρ_d and for ranges r of 30, 20 and 10 dB. The results are plotted in Figure 8.8 together with the fitted equation

$$\rho_d = \rho_p + z_0 \frac{1 - \rho_p}{1 - e^{-c\left(a + \frac{\tan((\rho_p - 0.5)\pi)}{\pi}\right)}} \qquad (28)$$

where

$$z_0 = 1 - e^{-0.042\,r} \qquad (29)$$

$$a = 0.041 - e^{-0.094\,r} \qquad (30)$$

and

$$c = 0.161\,r + 5.5 \qquad (31)$$

Program: \lwork2\logrho.wk3.

8.5.4 Effect of fine correlation of excitation pattern

Armed with this information, we can now repeat the analysis of the variation of the information content of a sound PR vector with the length of the vector.

As described above, the information content of a sound PR vector cannot be calculated directly because of its internal correlation, and is taken instead to be the information content of its KL transform coefficients (Figure 8.6). The information content therefore depends directly on the correlation matrix of the PR vector.

The result of the repeated analysis is shown as the lowest curve on Figure 8.6. The information content now shows an interesting shift at a vector length of around 50, before continuing to rise with vector length, though not as steeply as before. This shift corresponds to the appearance of high-frequency basis functions (eigenvectors), first with such small variances (eigenvalues) as to have negative information content (equation for h in Figures 8.5 and 8.6), and then with negative eigenvalues.

The proper way to interpret this is probably to compare it with oversampling and aliasing effects in a Fourier analysis of the excitation pattern. If so, then it is not clear what is to be gained by increasing the number of specification points beyond the shift, even though the information content is shown to rise.

Programs: \cwork\progs\speckl.c, \lwork2\allspekl.wk3.

8.5.5 Sparse-spectrum sounds

Sparse-spectrum sounds, in which both frequency and loudness of a relatively small number of sinusoids are specified, were also examined in Chapter 6. The equivalent of Figure 8.6, showing the information content of full-spectrum sounds as a function of the specification interval, is Figure 8.9, where the corresponding variable is the number of sinusoids.

It would have been nice to find some sort of upper limit to the number of sinusoids as there is a lower limit to the specification interval. Unfortunately, sounds of this nature become increasingly more difficult to generate as the number of sinusoids increases (curve D, Figure 8.9). On the other hand, while 14 sinusoids may not sound much, they do become quite crowded if you try to imagine them in Figure 3.1.

8.5.6 Conclusion

While curve C in Figure 8.9 is a truer measure of the information content of a sparse spectrum than curve B, curve B may nevertheless be the appropriate curve to compare with the results of Figure 8.6 for full-spectrum sounds, since the latter involved no squeezing of variances (and neither was the validity of the generated

sounds checked).

Comparing curve B, with up to 28 numbers representing the sound, with the curve in Figure 8.6 up to 30 spectrum specification points, shows no great advantage of one method over the other.

Consideration of information content therefore points to use of a full spectrum with a vector length of some 50 numbers.

8.6 Argument based on music and speech synthesis

The four arguments so far presented have assumed that the sound representation consists of a set of loudnesses (the numbers) at a fixed set of frequencies (the names of the numbers).

In specifying music for synthesis, there is generally a limit on the number of notes that can be played at once. The reason for this is the computational load entailed by the large number of overtones (sinusoids at other frequencies than the fundamental) that are included in each note in order for it to sound like a violin or a trumpet. These overtones are often called harmonics, but in the general case (for example bells) are not necessarily all harmonic.

Thus there is a distinction between the number of notes and the number of different fundamental frequencies a note can be given. In simple systems the latter may be limited to the number of semitones in six or seven octaves, but they can also be made continuously variable. In either case, the frequency of a note is no longer a name but a number, and the number of these numbers is the number of notes that the system allows to be played simultaneously. The computational load is such that, unless special overtone-pruning measures are adopted (eg Haken 1992), the number of notes is invariably less than the number of players in a large orchestra.

The total number of numbers required to specify a spectrum in musical synthesis is not limited to one fundamental frequency per note. Each note also requires a loudness, and specification of its overtones and of the spectral profile of its noise content. (Temporal effects, specified by such things as attack and decay times and amounts of tremolo and vibrato, are irrelevant here since for the moment we are only considering the spectrum.) The overtones and noise profile are usually specified by only one number, namely an item in a list of instruments.

The total number of numbers required to specify a spectrum in musical synthesis is thus equal to 3N, where N is the number of simultaneous notes. If N is 16 (the

first violins, for example, count as one instrument), then 48 numbers are required.

In speech there is little interest in specifying two voices at once, which makes N above equal to 1. The numbers required to specify a phoneme are one loudness, one fundamental frequency (for voiced phonemes), one phoneme name, one dialect name, one language name, and one speaker name (such as "adult female" or "Margaret Thatcher"). Each of the above "names" is in fact an item in a list, and is thus a number. A total of 6 numbers are therefore required to specify a spectrum in speech.

8.7 Discussion

To recap, the different lines of thought described above give variously

	A	30	erbs and erds
	B	1150	sinusoid frequency DLs
	C	150	1-erb formant frequency DLs
	D	300	0.1-erb formant freqency DLs
	E	30	spectral modulation depth
	F	50	information theory
	G	48	music synthesis
and	H	6	speech synthesis

as the number of numbers required to specify a spectrum.

The differences between these numbers are of two kinds. First, B, C and D recognise that adjacent notes would interfere with each other and so don't allow all the notes to be played together, while A, E and F, with their wider initial spacing, impose no such restriction. The question then arises as to how A, E and F can convey the difference between two sinusoids one frequency DL apart. If not, then these methods are deficient, since they exclude sounds of this class. Baldi & Heiligenberg (1988) and Schorer (1989) attempt in different ways to show how this might be achieved.

Second, all except G and H are intended as methods of representing any possible steady sound. The existence in G and H of items from lists implies nonreproducible sounds, namely those not on the lists.

Two further factors will influence our final choice. First, we require our numbers to be in units of difference limens in order to apply KL transforms. It is hard to see how items from lists can be in units of difference limens, except in the special case of simply constraining a variable already in units of difference limens to be an integer.

Second, in order to maximise the rate of acquiring

information from the environment, it is expected that the sounds, after a longish learning period (Chapter 3), will be sounded very fast, leaving no time for the small DLs of Figures 8.1 and 8.2 (methods B and maybe D) to develop.

Pure-tone frequency difference limens

Replotted on erb scale from Moore (1989) p164

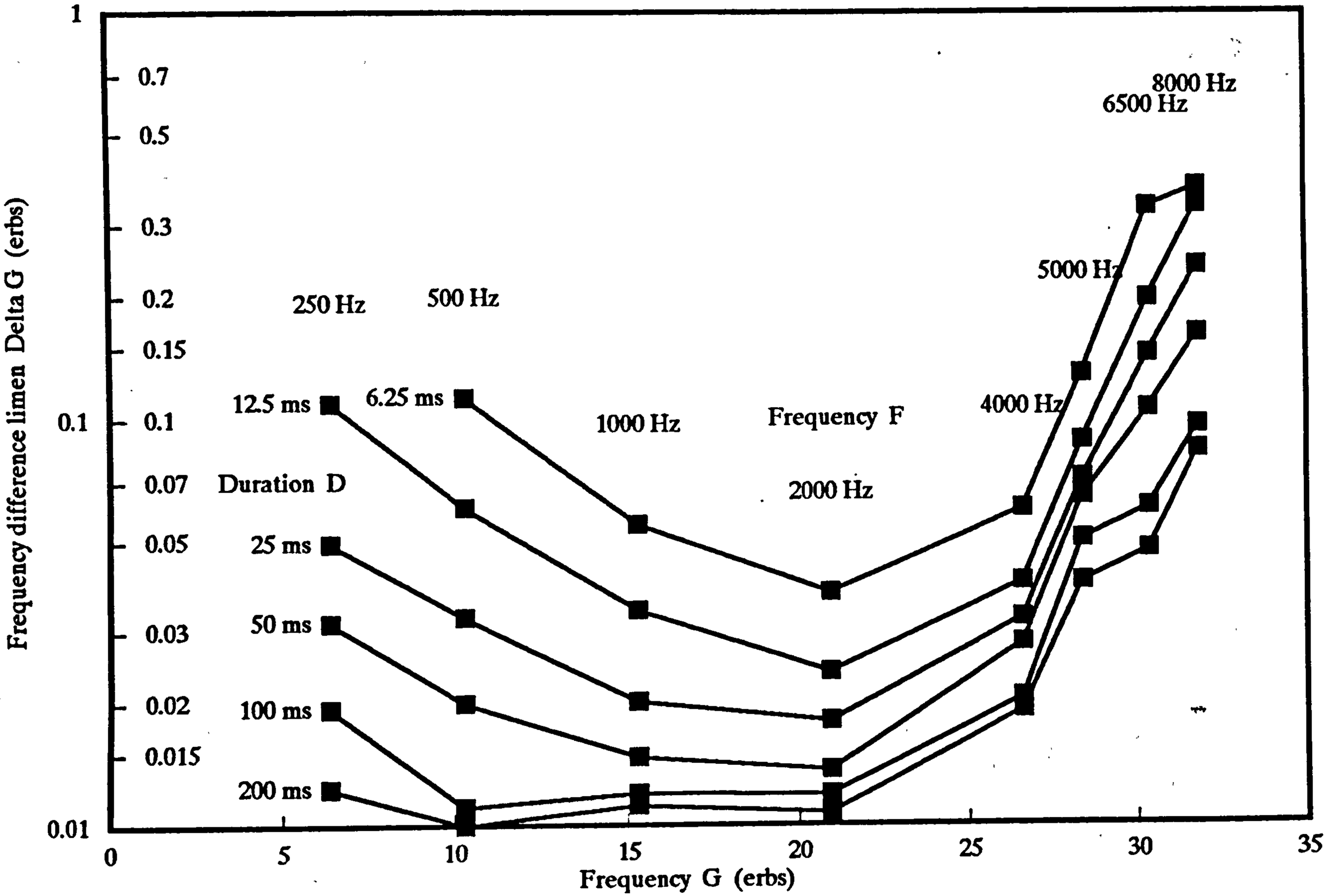


Figure 8.1

Figure 8.2

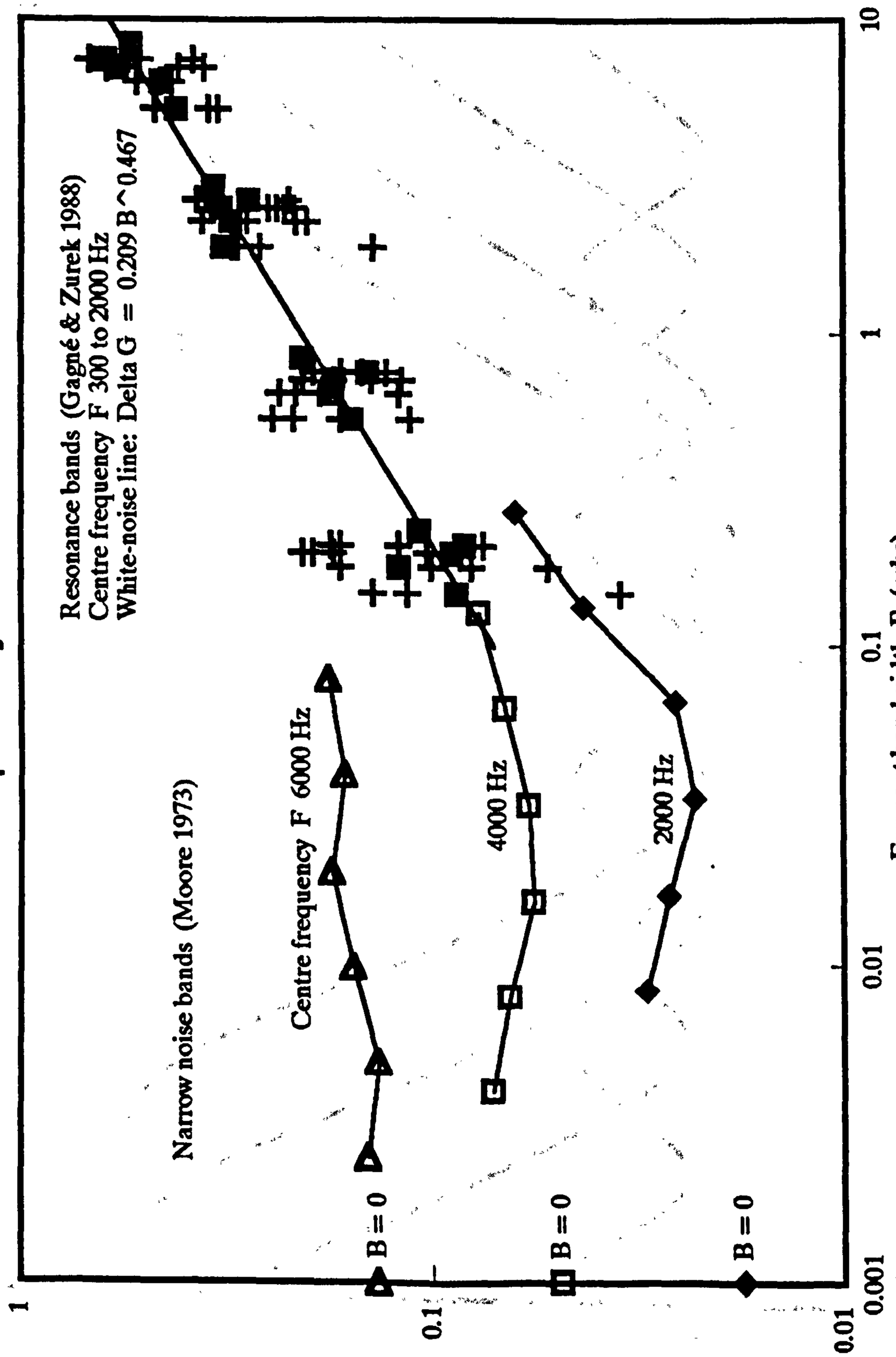

Formant frequency difference limens
Resonance bands (Gagné & Zurek 1988)
Centre frequency F 300 to 2000 Hz
White-noise line: Delta G = 0.209 B^0.467
Narrow noise bands (Moore 1973)
Centre frequency F 6000 Hz
4000 Hz
2000 Hz
B=0
B=0
B=0
10
1
0.1
0.01
1
0.1
0.01
0.001
Formant bandwidth B (erbs)
White noise source
Sawtooth source
Centre-frequency difference limen Delta G (erbs)

Figure 8.3

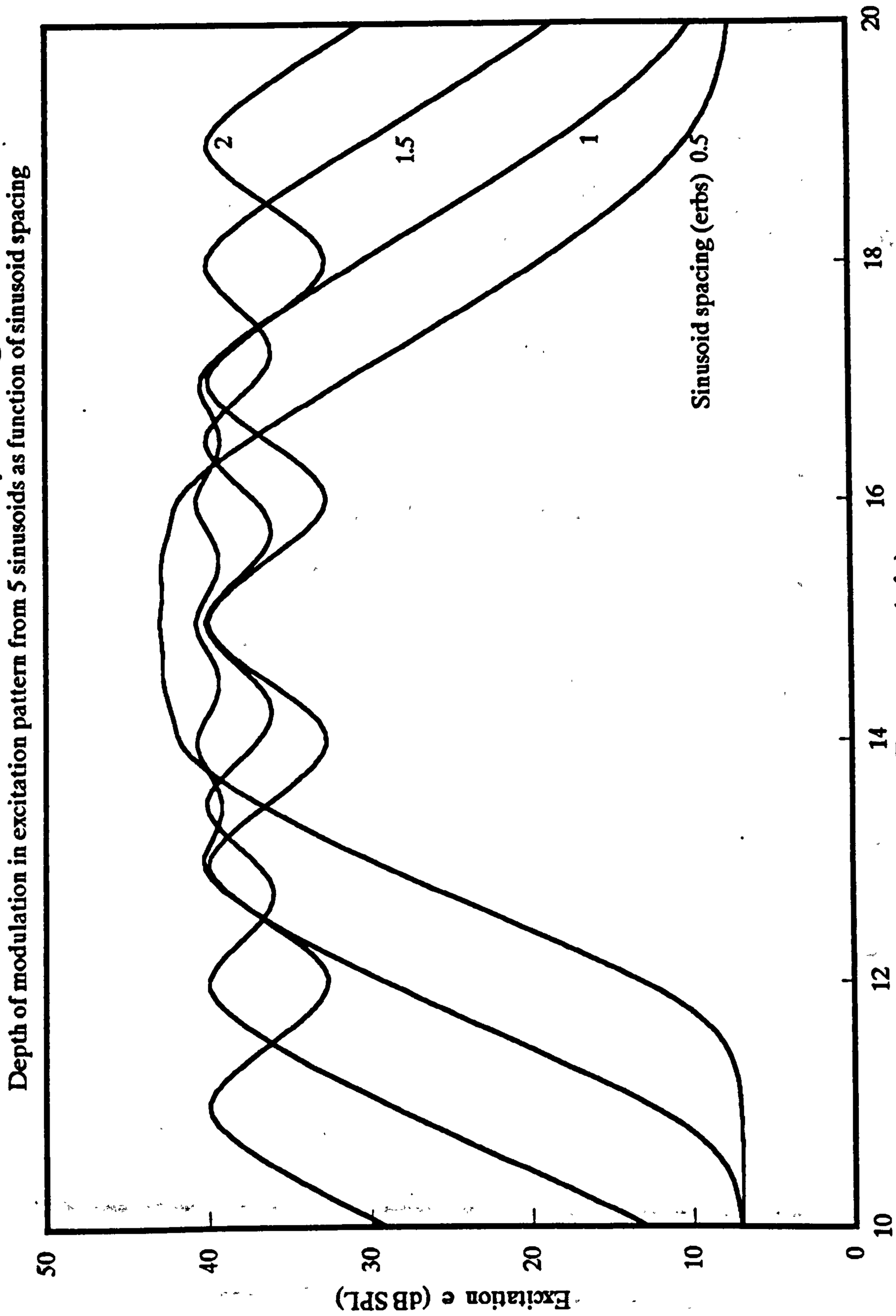
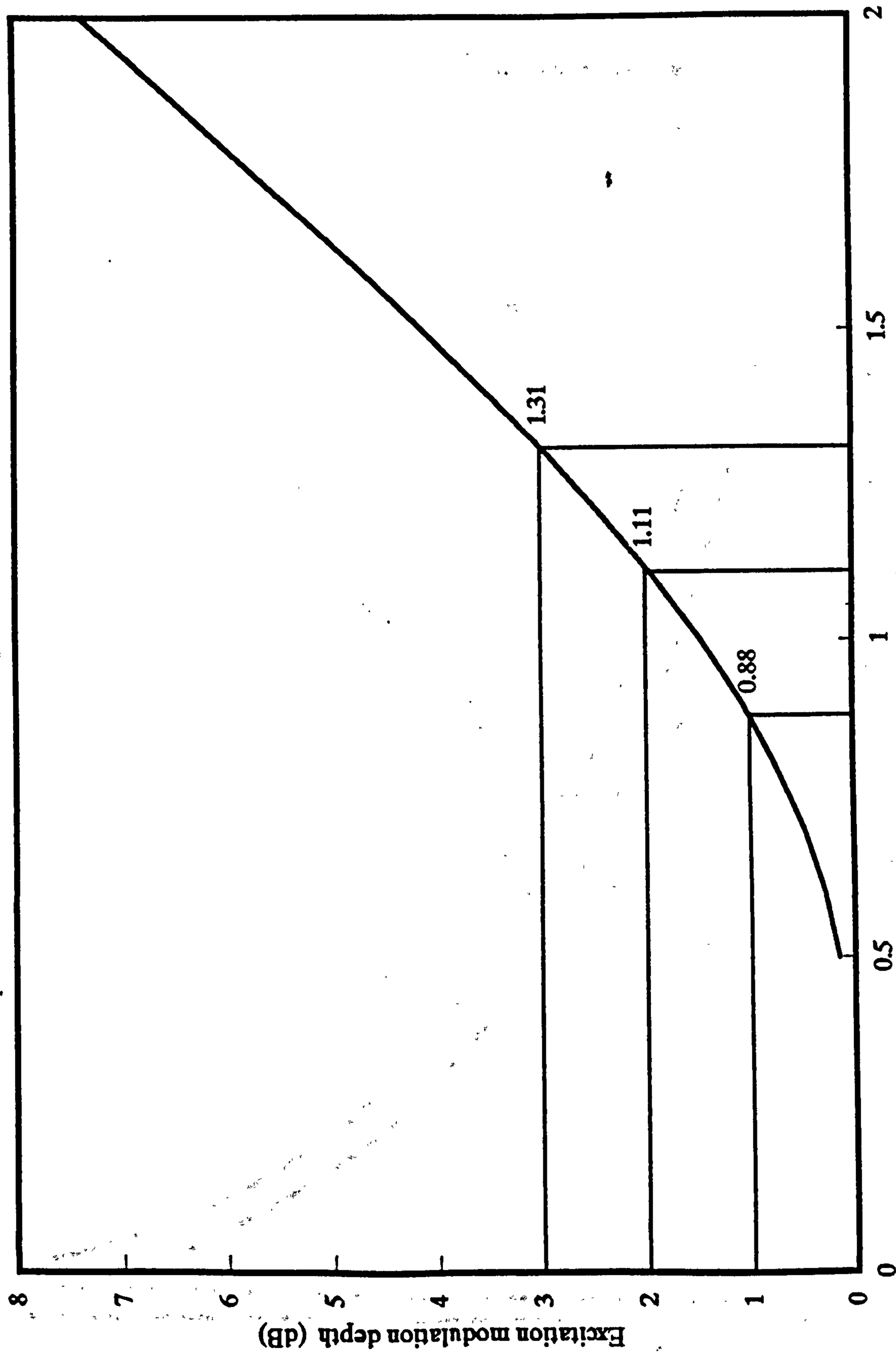

Minimum sinusoid spacing
Depth of modulation in excitation pattern from 5 sinusoids as function of sinusoid spacing
Excitation e (dB SPL)
50
40
30
20
10
0
Frequency g (erbs)
12
14
16
18
20
Sinusoid spacing (erbs) 0.5
1
1.5
2

Figure 8.4


Minimum sinusoid spacing
Depth of modulation in excitation pattern as function of sinusoid spacing
Excitation modulation depth (dB)
8
7
6
5
4
3
2
1
0
0
0.5
1
1.5
2
Sinusoid spacing (erbs)
1.31
1.11
0.88

Shannon's theory of information

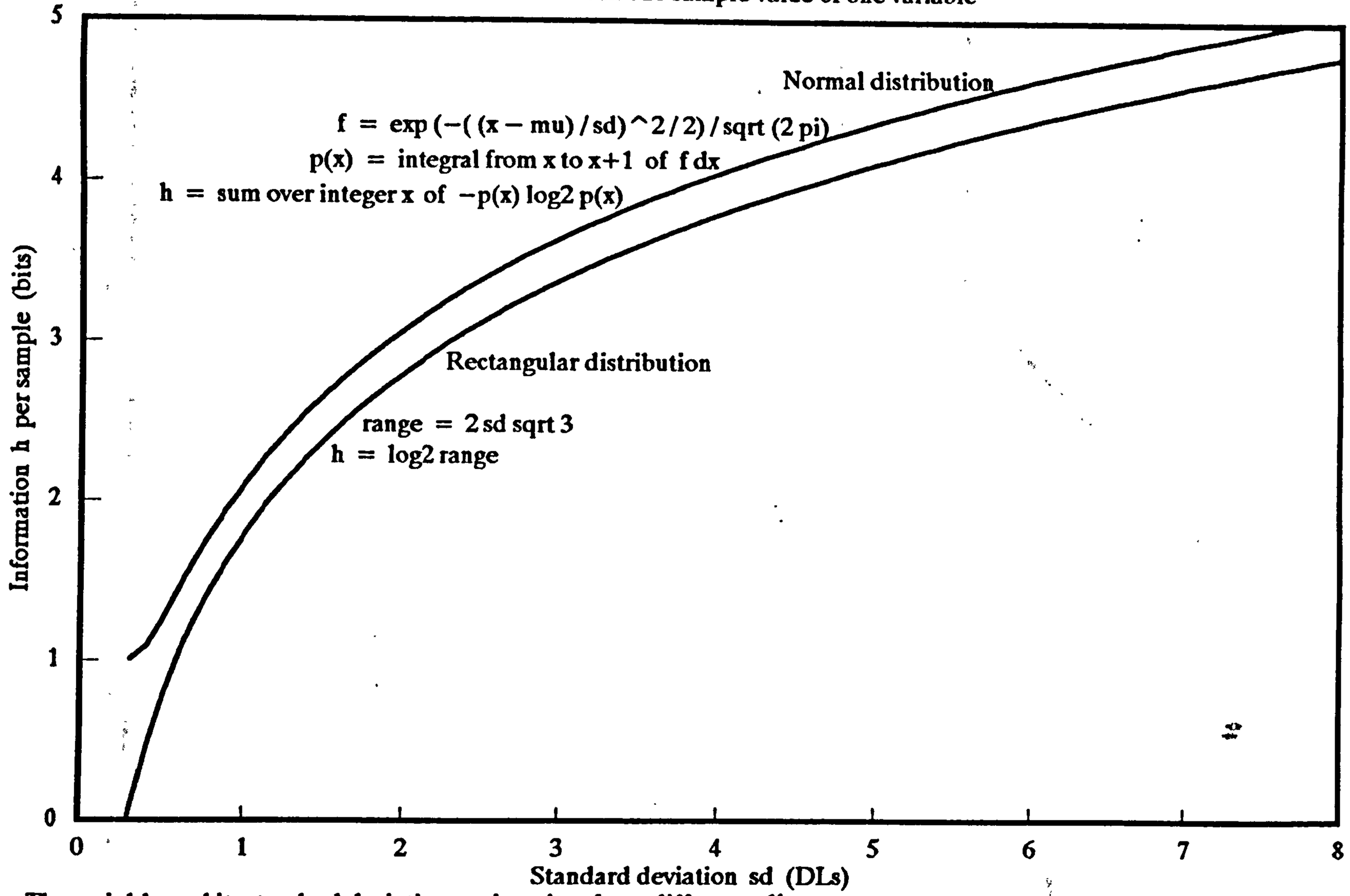
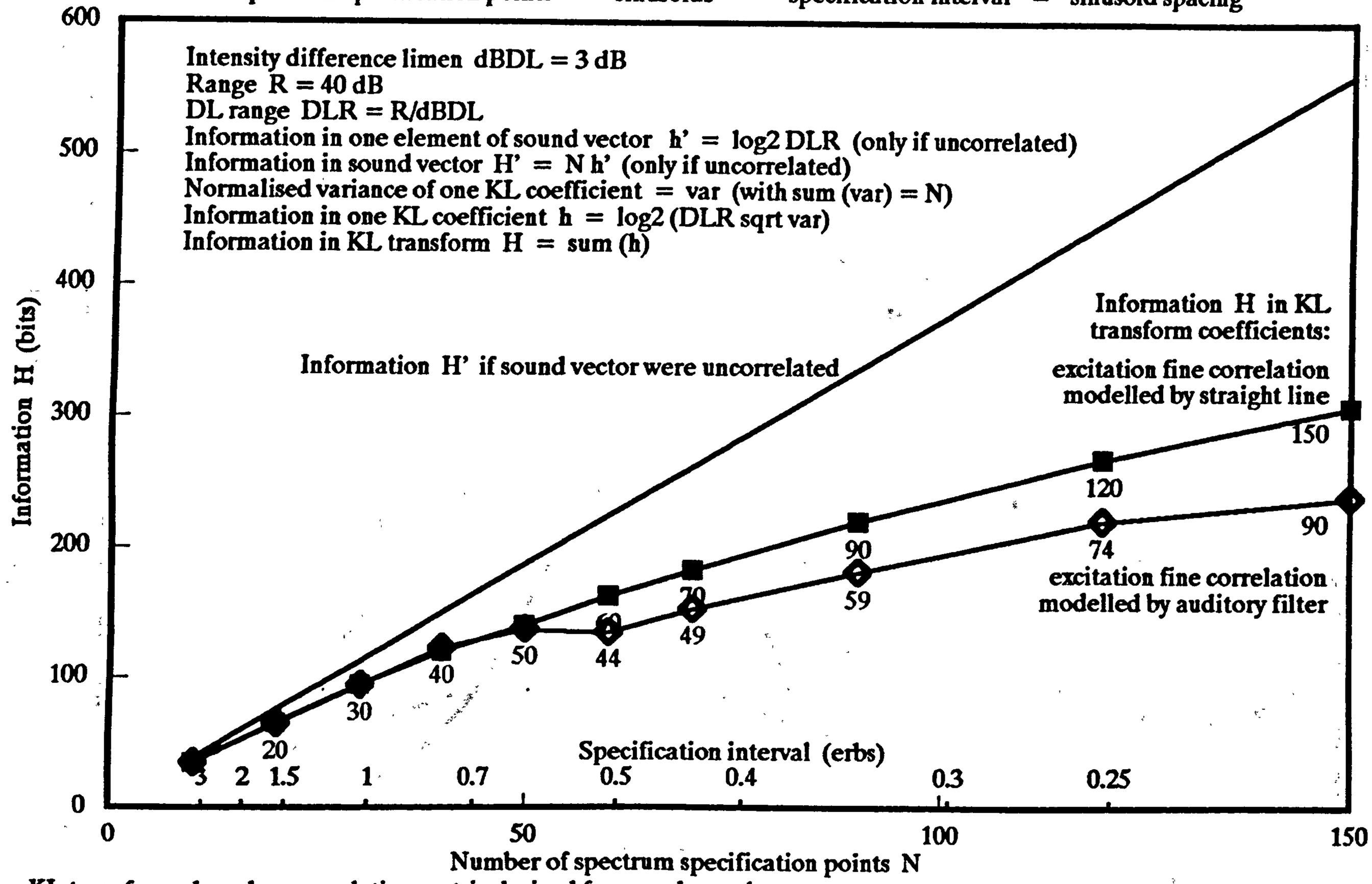


Figure 8.5

The variable and its standard deviation are in units of one difference limen.

Spectrum information versus specification interval

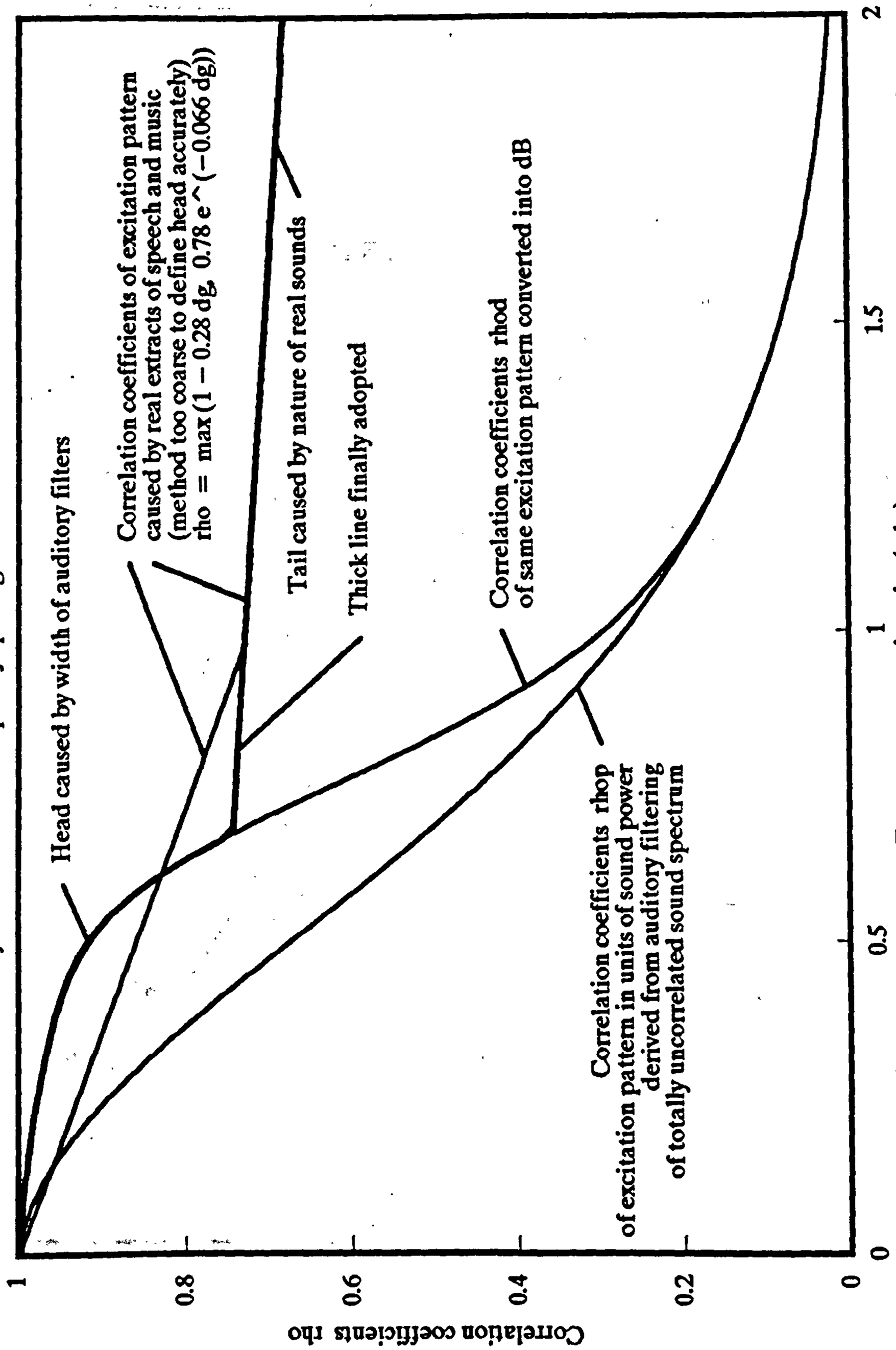
"spectrum specification points" = "sinusoids" "specification interval" = "sinusoid spacing"



KL transforms based on correlation matrix derived from real sounds.
Below each point is shown number of KL transform coefficients used (remainder supply no further information).

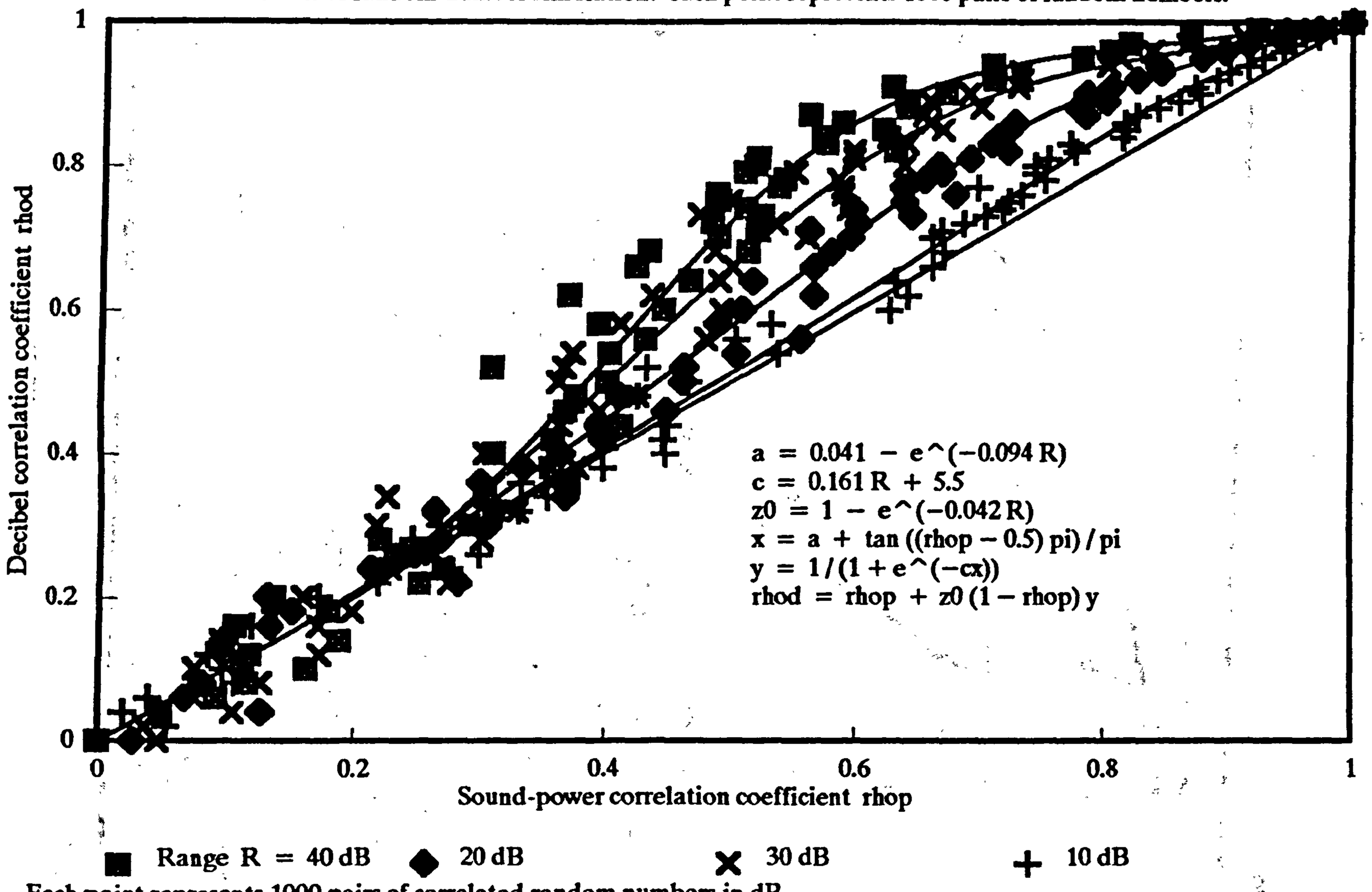
Figure 8.6

Figure 8.7


Simultaneous excitation-level correlation
Theoretical adjustment at close frequency spacing of measured correlation of real sounds
Correlation coefficients of excitation pattern caused by real extracts of speech and music (method too coarse to define head accurately) rho = max (1 – 0.28 dg, 0.78 e^(–0.066 dg))
Tail caused by nature of real sounds
Thick line finally adopted
Correlation coefficients rhod of same excitation pattern converted into dB
Head caused by width of excitation pattern in units of sound power derived from auditory filtering
Correlation coefficients rhop of totally uncorrelated sound spectrum
Correlation coefficients rho
Frequency separation dg (erbs)
0
0.5
1
1.5
2
1
0.8
0.6
0.4
0.2
0

Sound-power versus decibel correlation coefficients

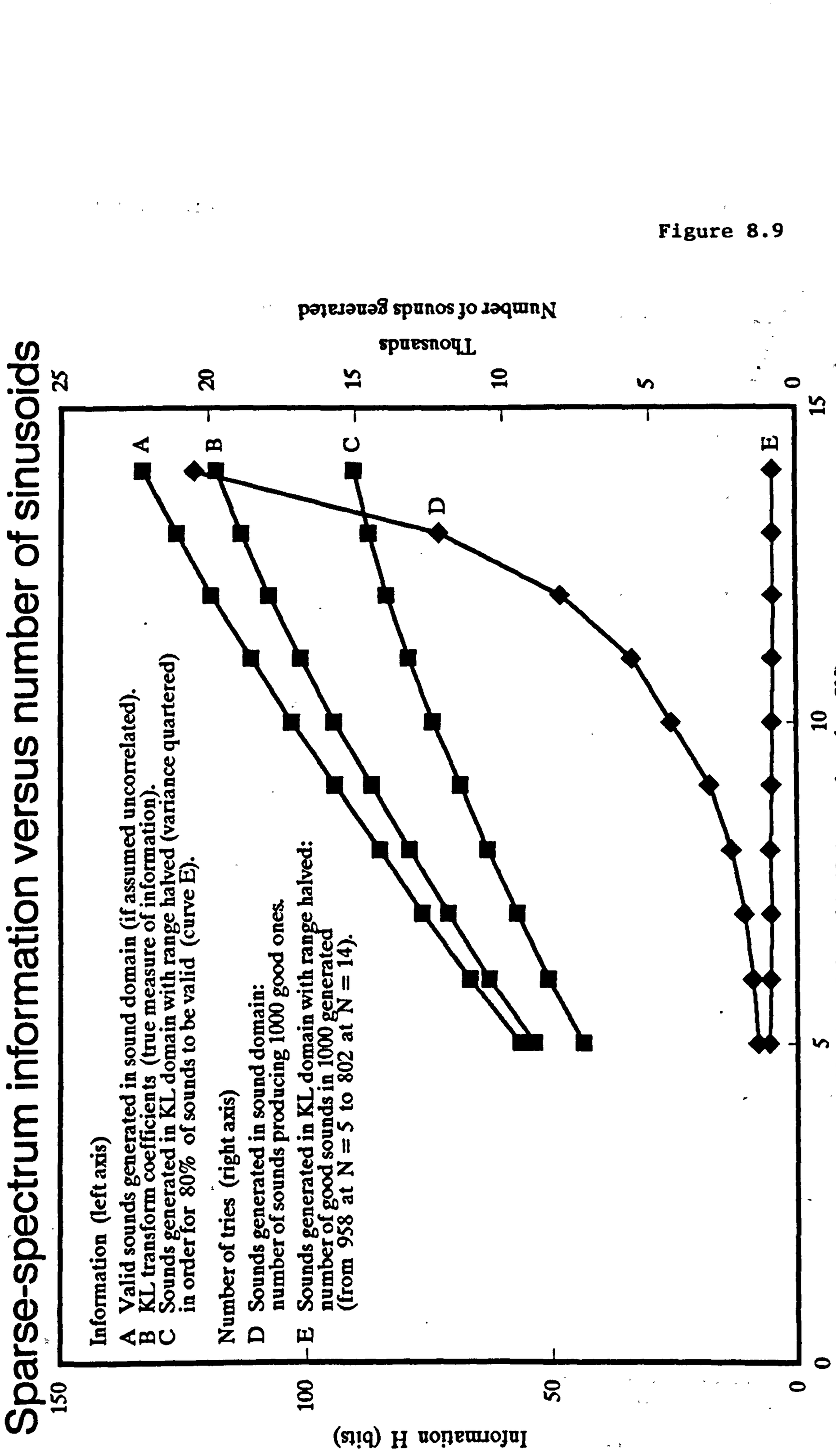
Results of random-number simulation: each point represents 1000 pairs of random numbers.



Each point represents 1000 pairs of correlated random numbers in dB, and 1000 corresponding pairs of numbers in sound power (W/m^2).

Figure 8.8

Figure 8.9


Sparse-spectrum information versus number of sinusoids
Information H (bits)
Thousands
Number of sounds generated
Information (left axis)
A Valid sounds generated in sound domain (if assumed uncorrelated).
B KL transform coefficients (true measure of information).
C Sounds generated in KL domain with range halved (variance quartered)
in order for 80% of sounds to be valid (curve E).
Number of tries (right axis)
D Sounds generated in sound domain:
number of sounds producing 1000 good ones.
E Sounds generated in KL domain with range halved:
number of good sounds in 1000 generated
(from 958 at N = 5 to 802 at N = 14).
Number of sinusoids N (vector length = 2N)
150
100
50
0
25
20
15
10
5
0
5
10
15
A
B
C
D
E

CHAPTER 9 SCHEME 4 - FREE-FIELD PATCH TRANSFORM

9.1 Motivation

All transforms considered so far have been slot transforms, meaning that the sound at any time is determined by the contents of a slot masking the scene. It is undeniable that a square or round patch from a scene is subjectively more meaningful than a long narrow strip of the scene, which is all that shows through a slot.

On the sound side, a meaningful chunk of sound that holds together is contained between two time limits and takes up the whole frequency spectrum, with some sort of spectral continuity from one instant to the next (see references under **PSYCHOPHYSICS - HEARING - AUDITORY STREAMING**). Typical examples are the phonemes of speech (chosen as the meaning of letters when alphabets are invented), the clang or thud of a struck object, and the chords in music (the bits musicians choose to write down one above the other as happening at the same time). Thus speech, music and many natural noises are all naturally divisible into short periods of similar spectral content. In keeping with common parlance, we shall simply refer to any such period as "a sound".

The free-field patch transform is thus an attempt to match sounds, as defined here, to shapes (small areas of a scene).

There remains the question of conveying the position of the patch in the scene. This cannot involve the spectral content of the sound, since that is concerned with the shape in the patch. Two methods come to mind. First, to sound the sound in such a way that it appears to come from the direction of the position of the patch in the scene. This is called simulated free-field listening (or presentation). Second, to sound the patches in some predetermined order. These two methods have very different consequences. The title of the chapter gives a clue as to which is chosen here.

9.2 Matching patches and sounds

9.2.1 General

It was decided to have another go at the KL method abandoned at the end of Chapter 6. It was thought that the method might turn out to be tractable because of the small number of pixels involved, not just computationally but also from the point of view of matching the basis functions and choosing their signs.

9.2.2 Psychophysical representation of patch

Let us now try to match patches and sounds. How does this differ from what we've been doing already? For one thing, we can now choose the patch size, in terms of pixels, to match the number of numbers needed to specify a spectrum for human consumption. This is very many times less (several thousand times less) than the number of pixels in the usual digital image.

Chapter 8 was devoted to the question of how many numbers are needed to describe a spectrum, and gave an answer in the region of 50. Accordingly, let us take a patch to be described by the following 53 numbers:

- 25 slopes in the x direction
- 25 slopes in the y direction
- one overall brightness
- one Oleari x
- one Oleari y.

The relative weight given to the colour of the patch (three out of 53) is in accordance with the coarse colour resolution of human vision, discussed in Section 2.6.2. For the meaning of the Oleari x and y colour coordinates, see Section 2.6.1. Note that they have nothing to do with the x and y directions.

Slopes (brightness gradients) are chosen rather than brightnesses because of the psychophysical importance of edges and their orientation. This importance has even been demonstrated physiologically: famous experiments by Hubel & Wiesel (1962, 1968) showed different areas of the visual cortex of cats and monkeys to be sensitive to different edge orientations. The choice of slopes over brightnesses is equivalent to the linear weighting on the "best the eye wants to do" side of Figure 6.2.

9.2.3 Statistics of patch PR (first go)

The statistics of the patch psychophysical representation were examined in the by now usual way (see sections 6.2.4 and 6.3.3). First, the 53 numbers were extracted from patches from two real scenes, \pics\bike.q and \pics\shanti.q, by program \cwork\progs\patstat.c, and various correlations calculated.

Figure 9.1 is divided into seven panels. Bottom right is the scene in question. Top left and top right are x and y slopes (brightness gradients) respectively, calculated from the four surrounding pixels. These two panels are reproduced in Figures 9.2 and 9.3.

The problem now arises as to how to interpret the correlation coefficients derived from the scene. The

straightforward way would be a 53 by 53 matrix of correlation coefficients. However, we would expect many of the numbers in this matrix to be the same, since the important factor is clearly relative and not absolute position.

Bottom left in Figure 9.1 are four panels of correlation coefficients between the first 50 of the 53 variables, that is the slopes only. The purpose of these four panels is to examine the variation of correlation coefficient with relative position, as follows below. Each of the four panels has as abscissa separation in the x direction and as ordinate separation in the y direction. The origin (zero separation) is in the centre of each panel.

Top left of the four is x slope versus x slope. Top right is y slope versus x slope (y slope considered movable and x slope fixed). Bottom left is x slope versus y slope. Bottom right is y slope versus y slope.

We would expect the top left and bottom right panels to be symmetrical about both axes, and the bottom right panel to be a 90° rotation of the top left panel. Also, we would expect the top right and bottom left panels to be symmetrical about lines at 45° to the major axes, and both these panels to be identical. Because the correlation coefficients were extracted from a real

scene, these symmetries are only approximately true. An additional problem was the different high spatial frequency noise in the x and y directions, easily visible in comparing Figures 9.2 and 9.3, caused by the fact that the scene is digitised from a raster scan.

Figures 9.4 to 9.6 show the same information for a second scene, \pics\shanti.q. Any directional statistics in this scene are heavily influenced, and biased, by Shanti's shoulder straps. The expected symmetries hold even less for this scene.

In order to derive correlation coefficients with the correct symmetries, some artificial patches were generated at random, and their statistics examined as for the natural scenes, by program \cwork\progs\edg1stat.c and edg2stat.c. The 6 by 6 pixel artificial patches each contained just one straight edge positioned at random. Figure 9.7 shows one such patch. The edge is positioned at a random angle from 0 to 360° and at a random radius from the patch centre. The brightnesses of pixels straddling the edge are calculated as shown according to how much of the pixel is on each side of the edge.

The first 12 (of 1000) of these patches are shown in the top row of Figure 9.8. The second and third rows show the corresponding 5 by 5 x and y slopes, with grey zero, black negative and white positive.

Bottom left is a large panel of 50 by 50 correlation coefficients relating all 50 slopes in a patch to each other. To produce this correlation matrix the patch vector was ordered as in section 9.2.2.

The same numbers as in this correlation matrix are presented in the top two bottom centre panels, this time ordered by lag as in Figures 9.1 and 9.4. Again, the expected symmetries are only approximate, being derived from a sample of 1000, not the whole population. Nevertheless, the resemblance to the four bottom left panels of Figure 9.1 is striking.

The bottom centre panel of Figure 9.8 does have the expected symmetries. This was achieved by classifying the relative positions differently into a smaller number of categories, namely one quadrant only of the **xx** panel for the **xx** and **yy** cases, and one octant only of the **xy** panel for the **xy** and **yx** cases.

The numbers thus obtained are redisplayed in the bottom right panel of Figure 9.8 as the final correlation matrix to be adopted (at least for 50 out of the 53 elements of the patch vector).

Finally, the correlation matrix was turned into a covariance matrix by defining the slope difference limen to be one fortieth of the maximum slope, and the

brightness difference limen to be one fortieth of the maximum brightness. The Oleari x and y colour coordinates are in difference limen units by definition.

9.2.4 Decorrelation of patch PR (first go)

Karhunen-Loève basis functions were extracted from the 53 by 53 covariance matrix obtained as described above. These (or every seventh one) are shown in Figure 9.9.

Figure 9.10 shows the KL transform coefficient variance.

Figure 9.11 shows a randomly generated patch using the slope basis functions of Figure 9.9 and the inverse KL transform. The 50 slopes are then turned into the 36 brightnesses shown in the left half of the figure by means of a minimum-twist algorithm. Note that what is achieved is a patch with a clearly coherent edge, something lost when working directly with brightness statistics.

9.2.5 Psychophysical representation of sound

This has been dealt with at length in Sections 3.2 and 6.3.2.

9.2.6 Statistics of sound PR

This has been dealt with at length in Section 6.3.3.

9.2.7 Decorrelation of sound PR

This has been dealt with at length in Section 6.3.4.

9.2.8 Matching of patch and sound basis functions (first go)

As a first stab, patch and sound basis functions were simply matched in order of decreasing variance, that is Figure 9.10 with Figure 6.26. The results of this mapping are shown in Figures 9.12 to 9.15. A similar exercise, but using only the shape information (without the three colour coordinates), results in Figures 9.16 to 9.19.

These two mappings are immediately seen to be deficient in two respects. First, as feared, a nice strong feature such as a straight edge does not correspond to any similarly namable sound (although it might help to actually listen to the sounds shown).

Second, there is no translation invariance. If it is

intended to go through the basis functions and match them up two by two in order to get a subjectively more meaningful or otherwise satisfactory mapping, then translation invariance is essential, first in order to reduce the number of permutations, and second because we don't want an off-centre patch to produce a very different sound. Again there are two reasons for this. First, it is subjectively unsatisfactory. Second, the patch location, if conveyed by free-field listening effects, is only approximately known, with an accuracy of something like the patch size itself.

9.2.9 Translation invariance

It is possible that a suitable solution would be to work with the Fourier transform magnitudes of a patch (either the Figure-6.2-weighted magnitudes of the FT of the patch brightnesses or the magnitudes of the FT of the patch brightness gradients).

While such a transform is ambiguous (different brightness patterns can produce the same FT magnitude coefficients), it can be argued that such patterns do not occur naturally and would therefore not be considered by the brain as possible originators of the sound. Not only that, but there exist algorithms for automatic reconstruction of signals from Fourier magnitude only

(see references under that heading).

From the point of view of the general problem of optophonics (GPO - what is the best scene to sound mapping given that the variety of sounds is a constraint), it can in fact be argued that this feature (the property of inconsequential ambiguity - PIA) is a positive virtue, because it greatly reduces the number of brightness patterns that need to have a mapping into sound. Unfortunately, I've only just thought of this, and so haven't looked into it.

Instead, an unambiguous type of translation invariance has been investigated, namely automatic centring. The idea is to define a measure of interest, and, starting from a randomly chosen point in the scene, to find the locally most interesting patch and then sound it, in much the same way as the eye centres (fixates) on interesting features before passing on. Both the position and size of the patch are free in the search.

A measure of interest which successfully frames in a patch such items as eyes and mouths, and at a different scale faces or heads, is as follows. A two-dimensional weighting function, zero everywhere outside the patch, is used to sum the slopes (brightness gradients) in the patch. Patch size is one of the variables, and the sum must be divided by the sum of the weightings before the

interest in two patches of different size is compared.

Unfortunately, there is an additional scale effect. Suppose two patches of different size are being compared, each containing nothing but an edge in the same position. The required answer is that they are both equally interesting. However, simple summing of the slope values will not give this answer. Suppose for simplicity that the weighting function is 1 over the whole patch, that the small patch covers 5 by 5 slope values and the big patch 10 by 10, and that the edge is horizontal and falls exactly between two rows of pixels and gives a slope of 1 when a slope value is situated on it. Our interest measure is then 5/25 = 0.2 for the small patch and 10/100 = 0.1 for the large patch.

The size bias applied is shown in Figure 9.20, together with suitable values for the parameters as far as can be ascertained at present.

The shape of weighting function that seems to work best is a raised radial negative cosine. This is a circular bell shape with the centre of the bell depressed back down to zero (k = 1 in Figure 9.21). This shape tends to favour patches centred on an area of constant brightness surrounded by an edge - the simplest type of object. Figures 9.22 and 9.23 show some patches found automatically in this way. In each case a cross-shaped

cursor marks the patch centre at the start of the search.

The program finding these patches is \cwork\progs\findpat.c.

Unfortunately, the statistics of patches chosen in this way are of course different from those of patches chosen or generated at random or in some other way. It is therefore necessary to rederive the patch covariance matrix for such patches. To do this, a robust program in the nature of findpat.c is required. A present findpat.c is an interactive program to examine the effects of different parameters in the interest and bias equations.

9.3 Next step

After the statistics of the new type of patch have been found in the by now usual way, the interesting step will be what was first attempted in Figures 9.12 to 9.15, only listening to the sounds too. All the rows of those figures will now give much the same sound, since a change of row merely involved a translation of the patch over the scene. Instead, in the same space, it will be possible to examine the sound made by other fundamental shapes such as angles and curved or occluded edges (one edge disappearing behind another is a fundamental feature of the 2D vision of 3D scenes).

Then will come the WORK - changing the sign (and sometimes within limits the order) of the patch and sound basis functions being matched until some sensible results are obtained, namely distinctive features matching distinctive sounds.

For further recommendations as to Scheme 4, see Section 10.4.3.

Figure 9.1

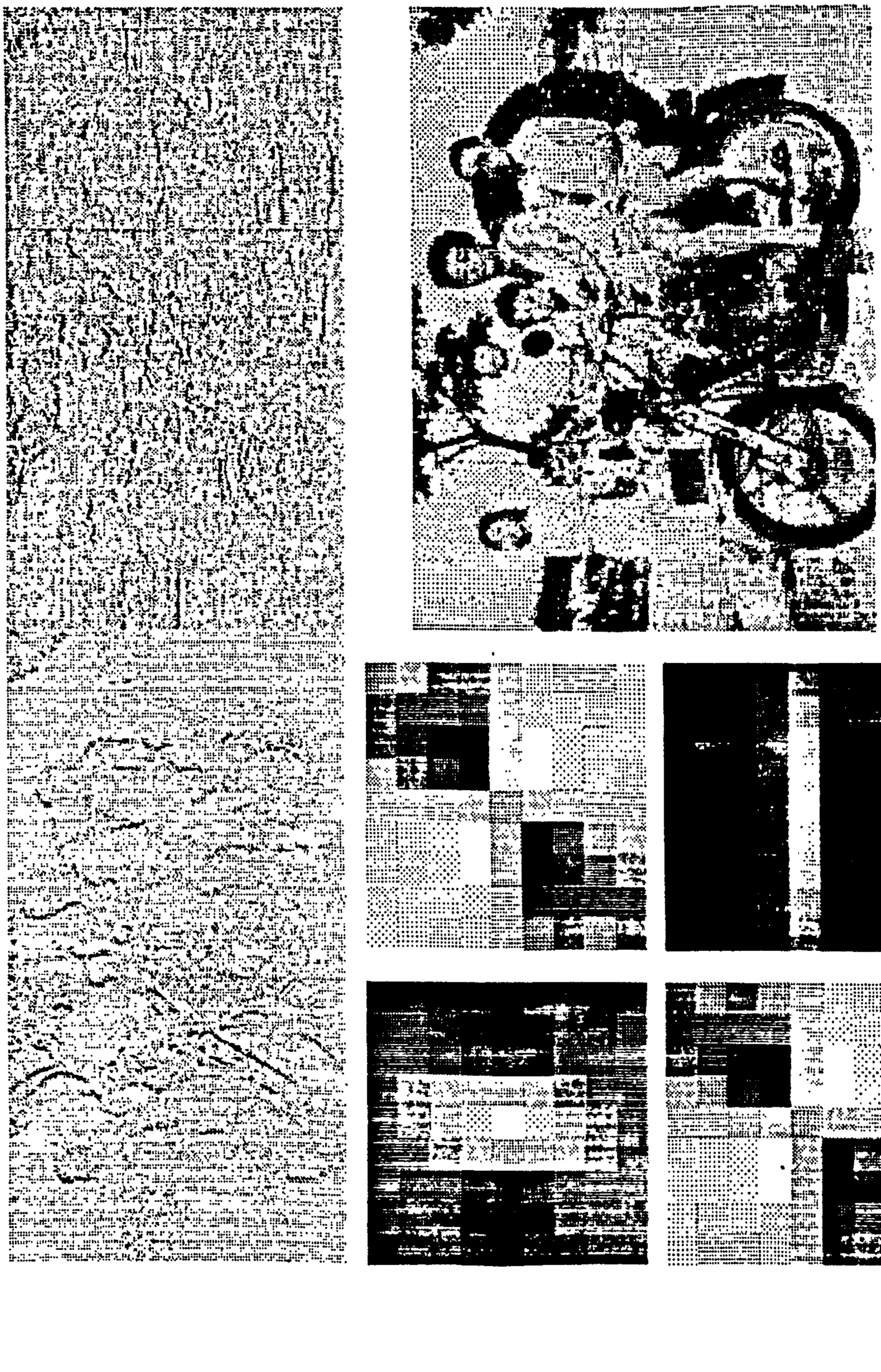


Figure 9.2

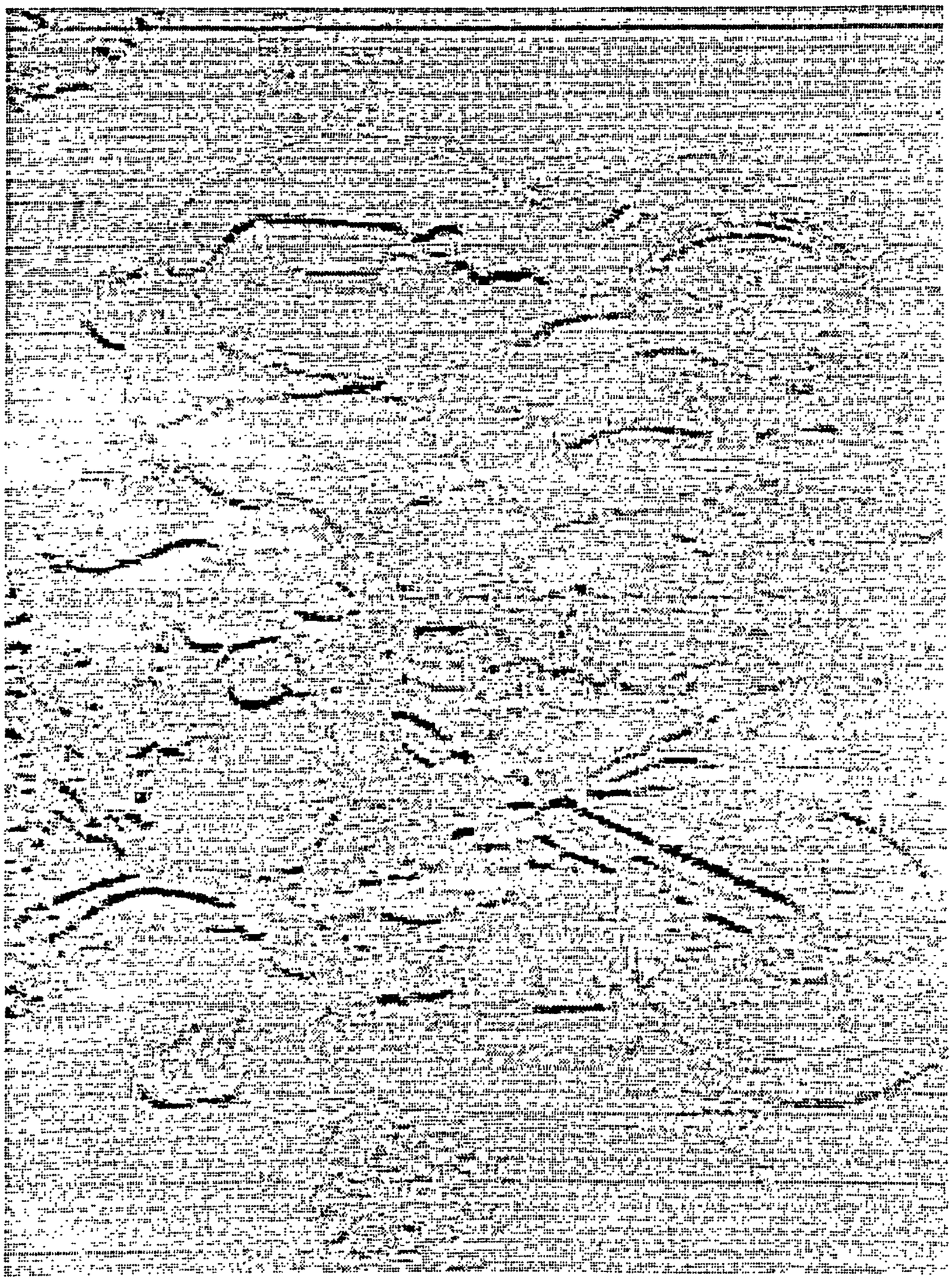


Figure 9.3

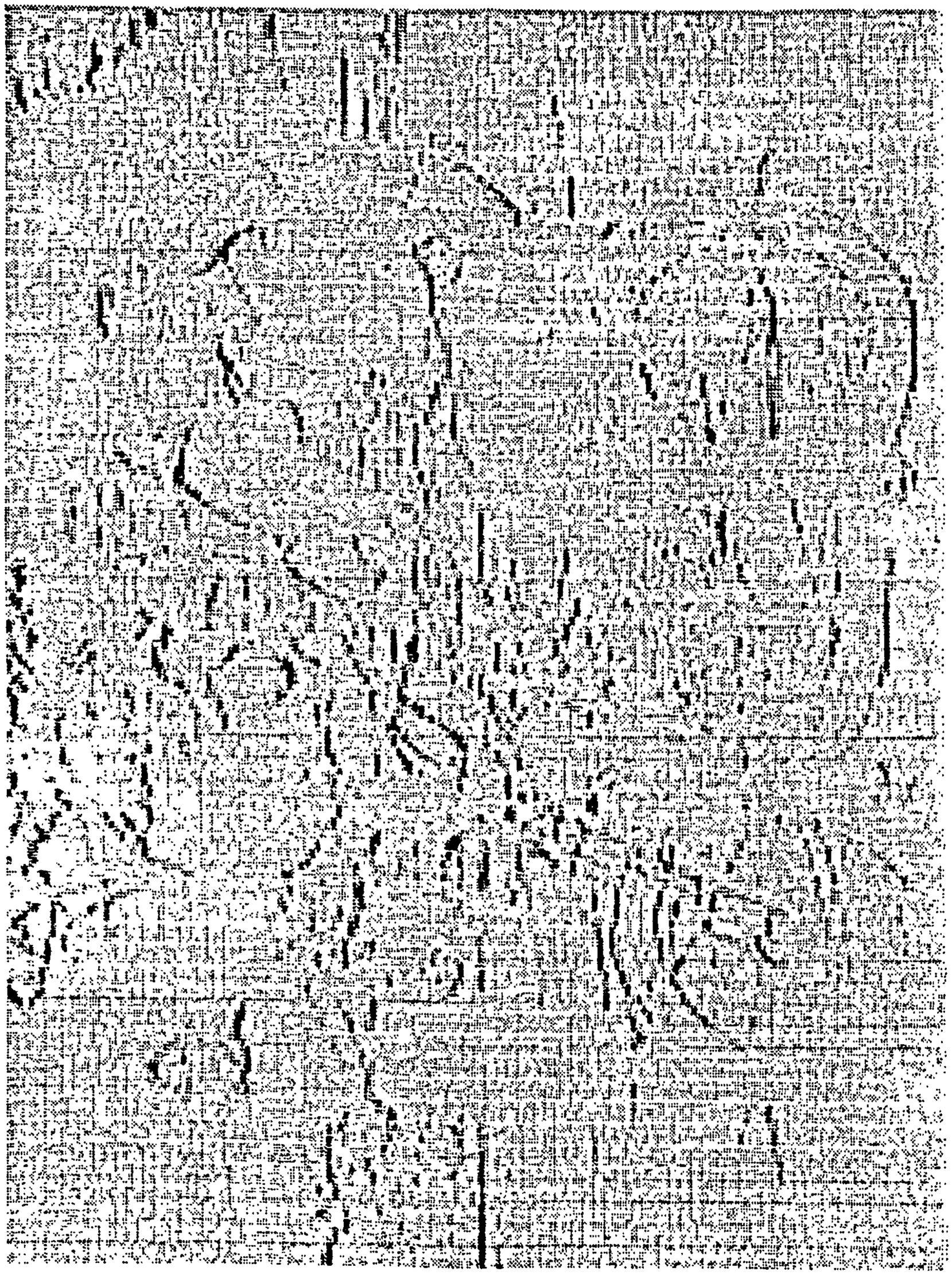


Figure 9.4

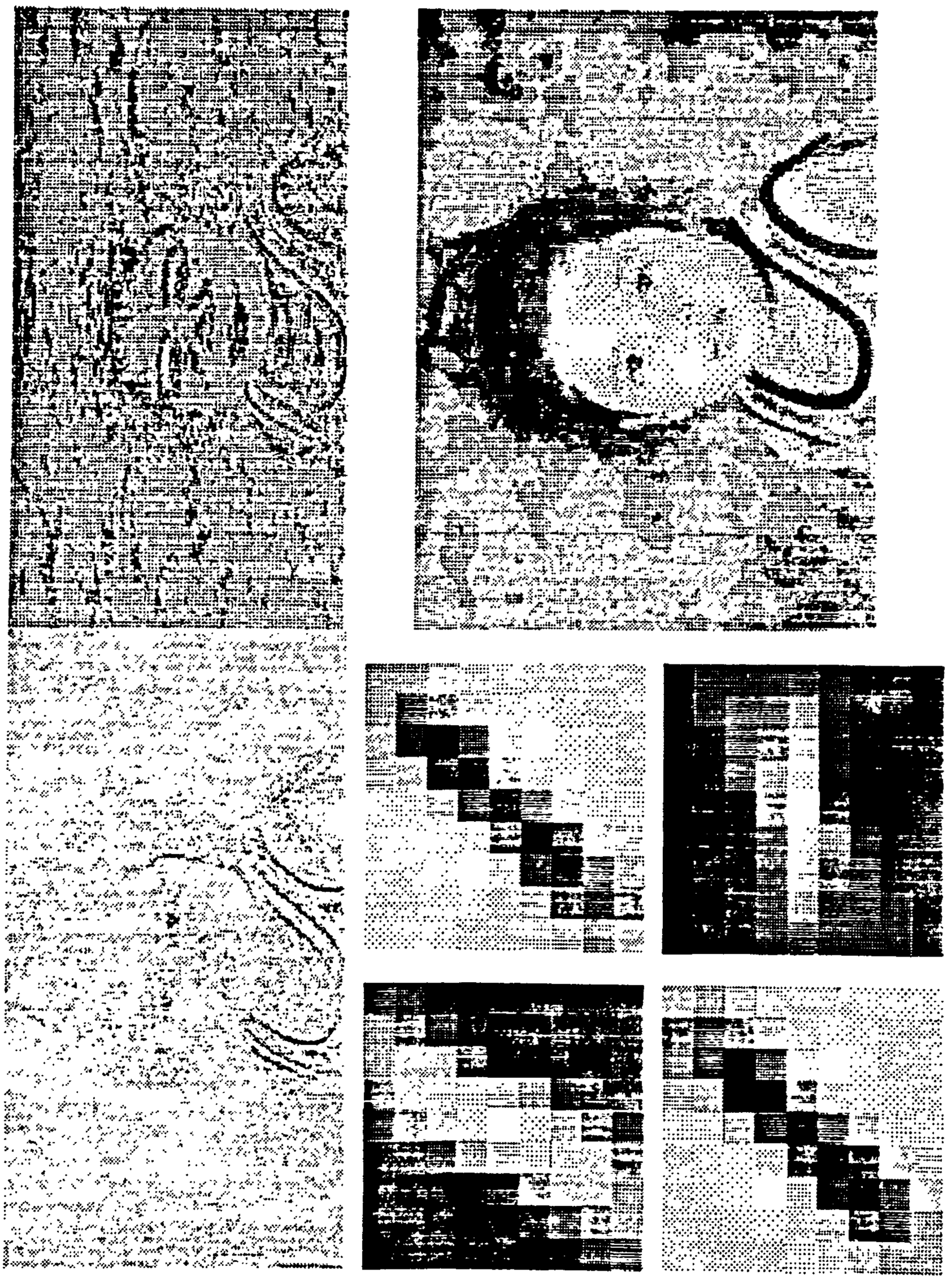


Figure 9.5

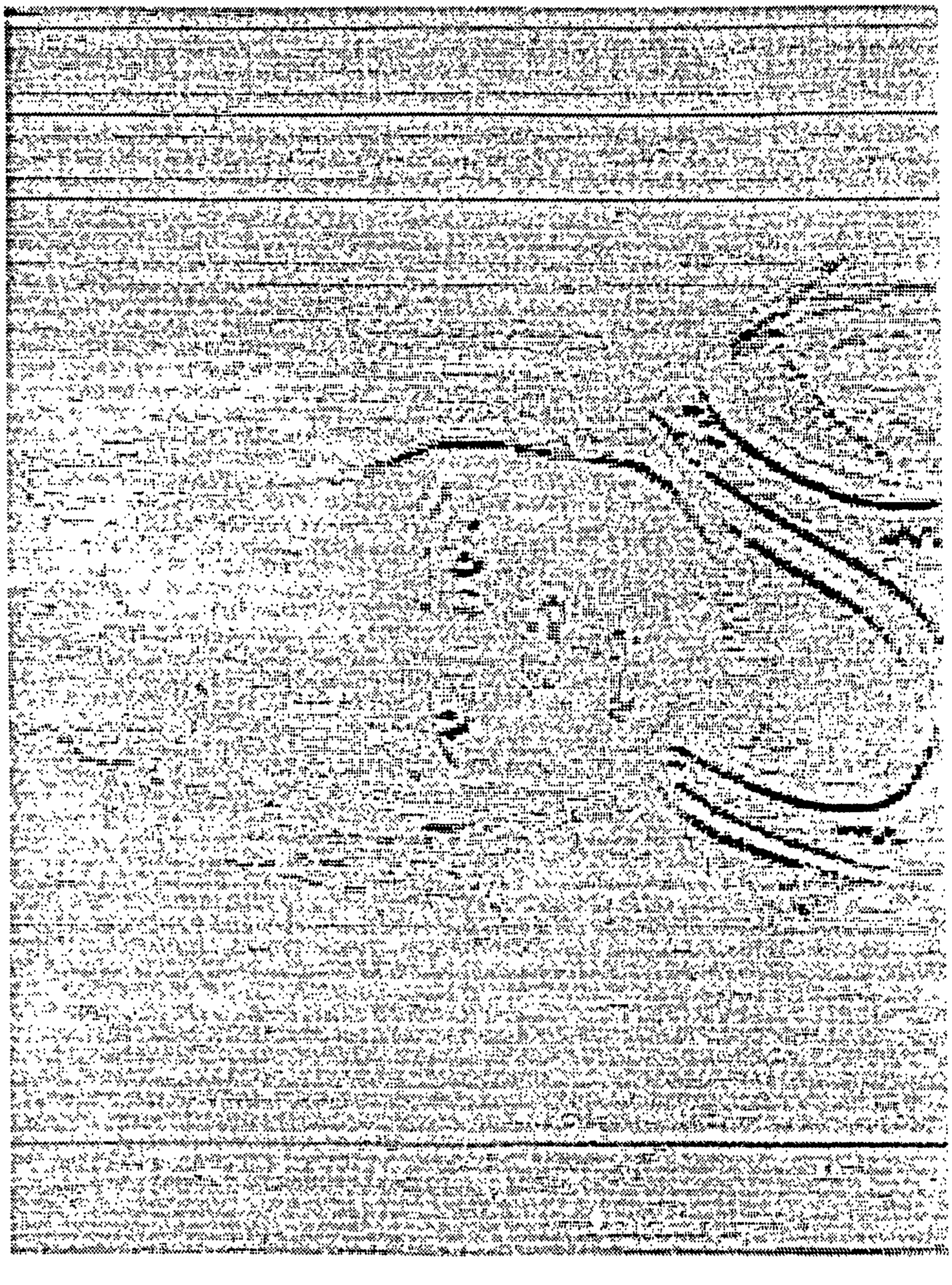
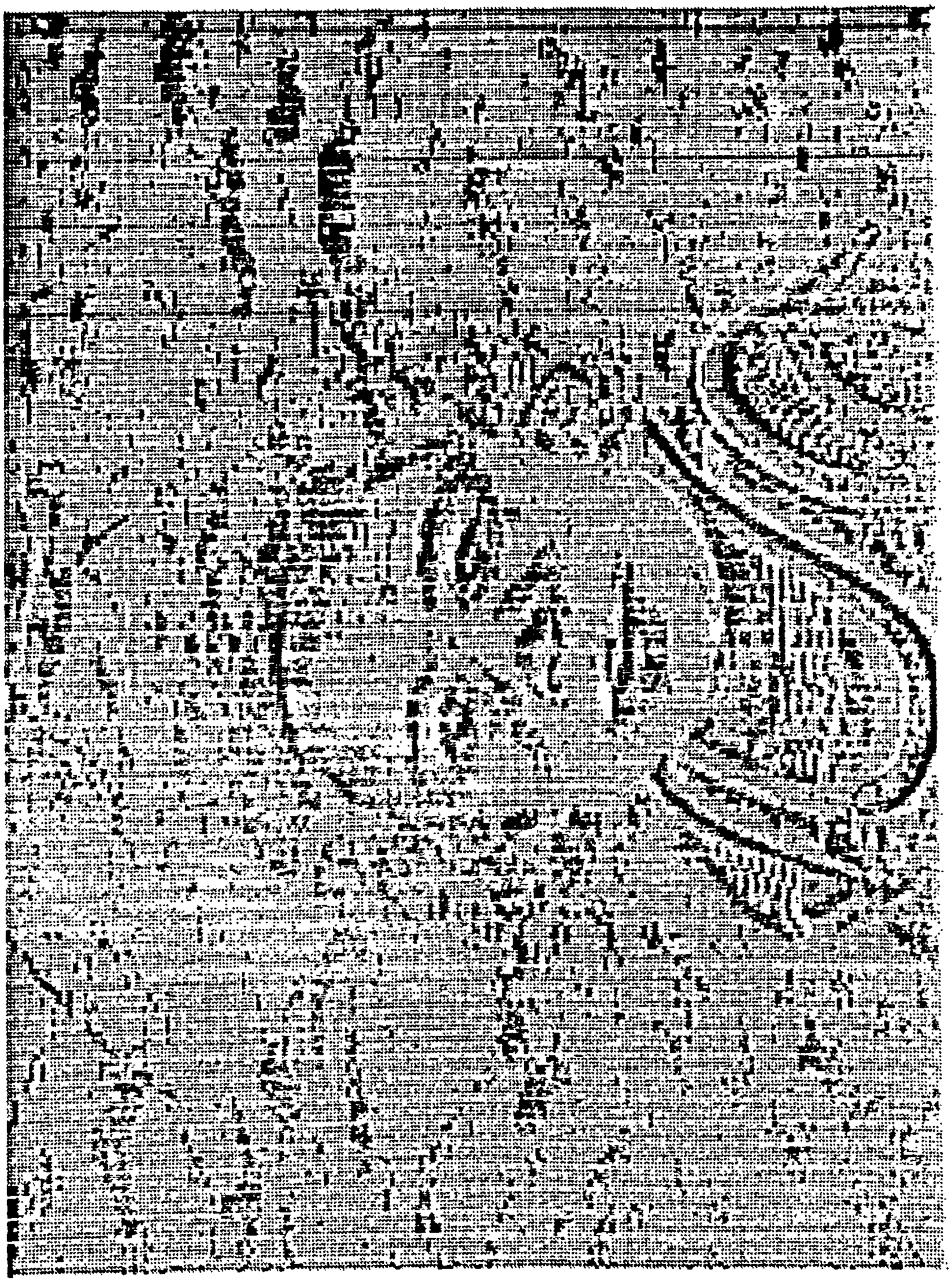
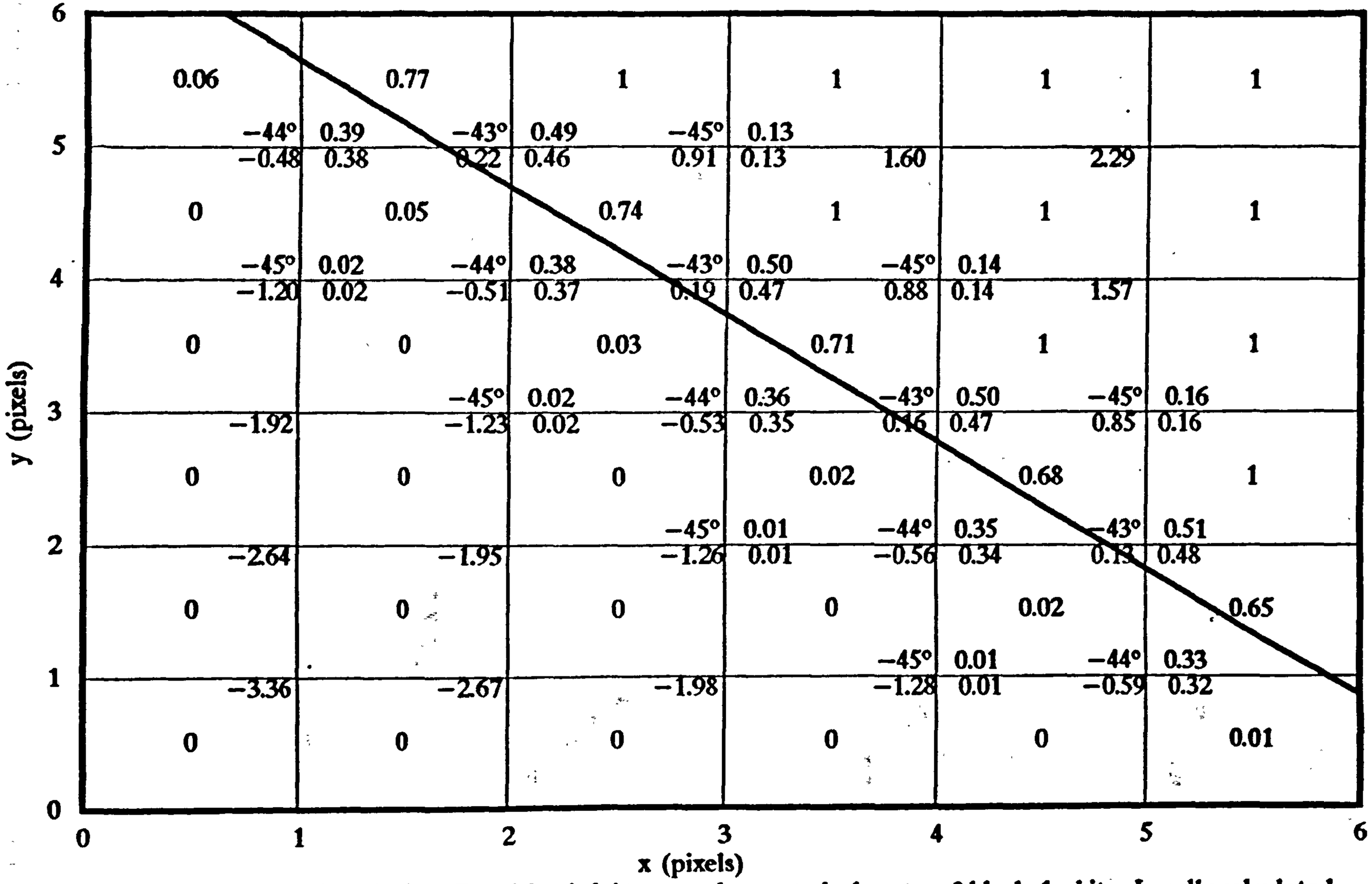


Figure 9.6



Brightness gradients from four surrounding pixels

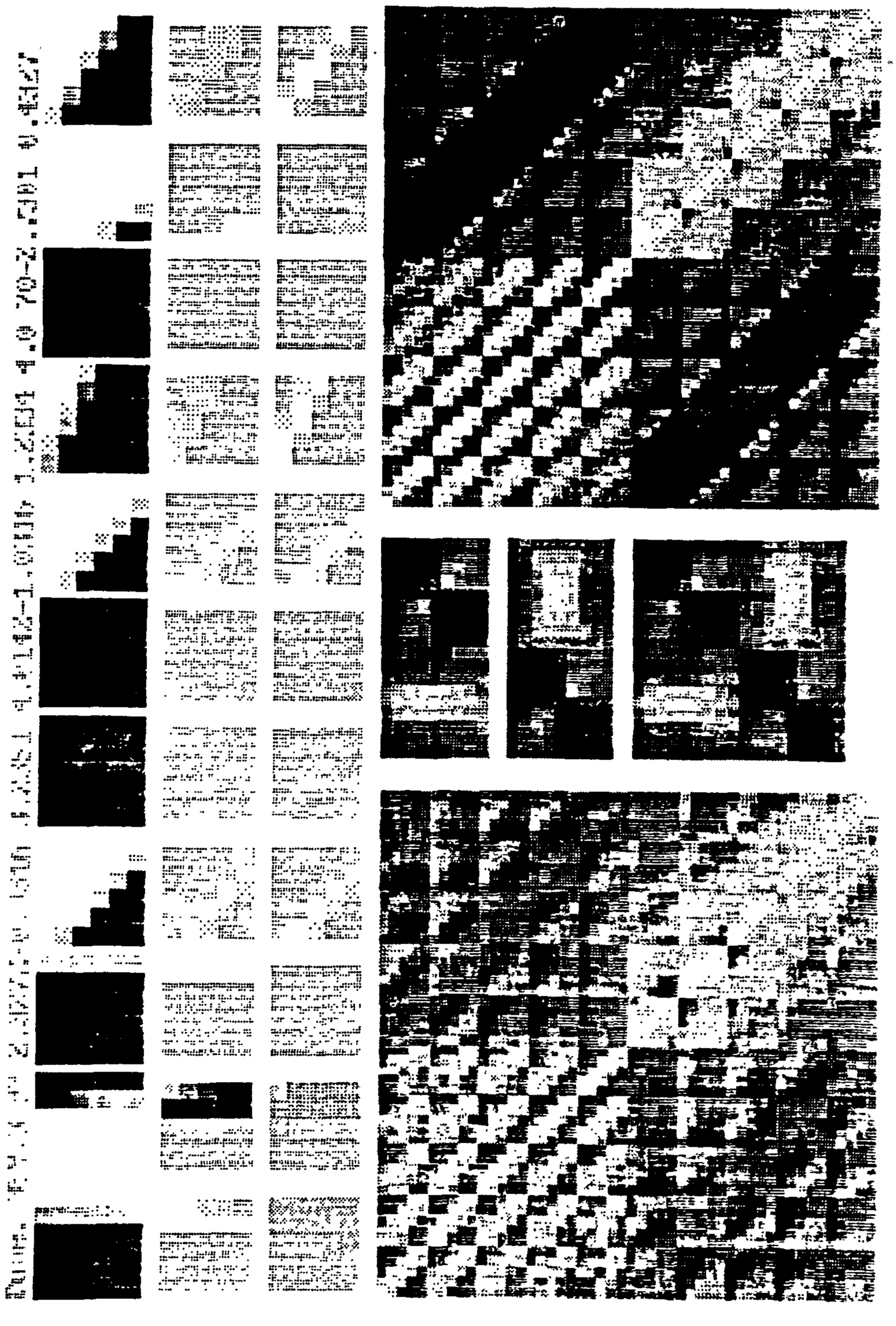
Derivation of input to correlation analysis



Line shows randomly positioned edge. Resulting brightnesses shown at pixel centre: 0 black 1 white. Locally calculated numbers clockwise from top right: 1 slope in y direction 2 slope in x direction 3 distance from edge 4 edge angle to x axis.

Figure 9.7

Figure 9.8



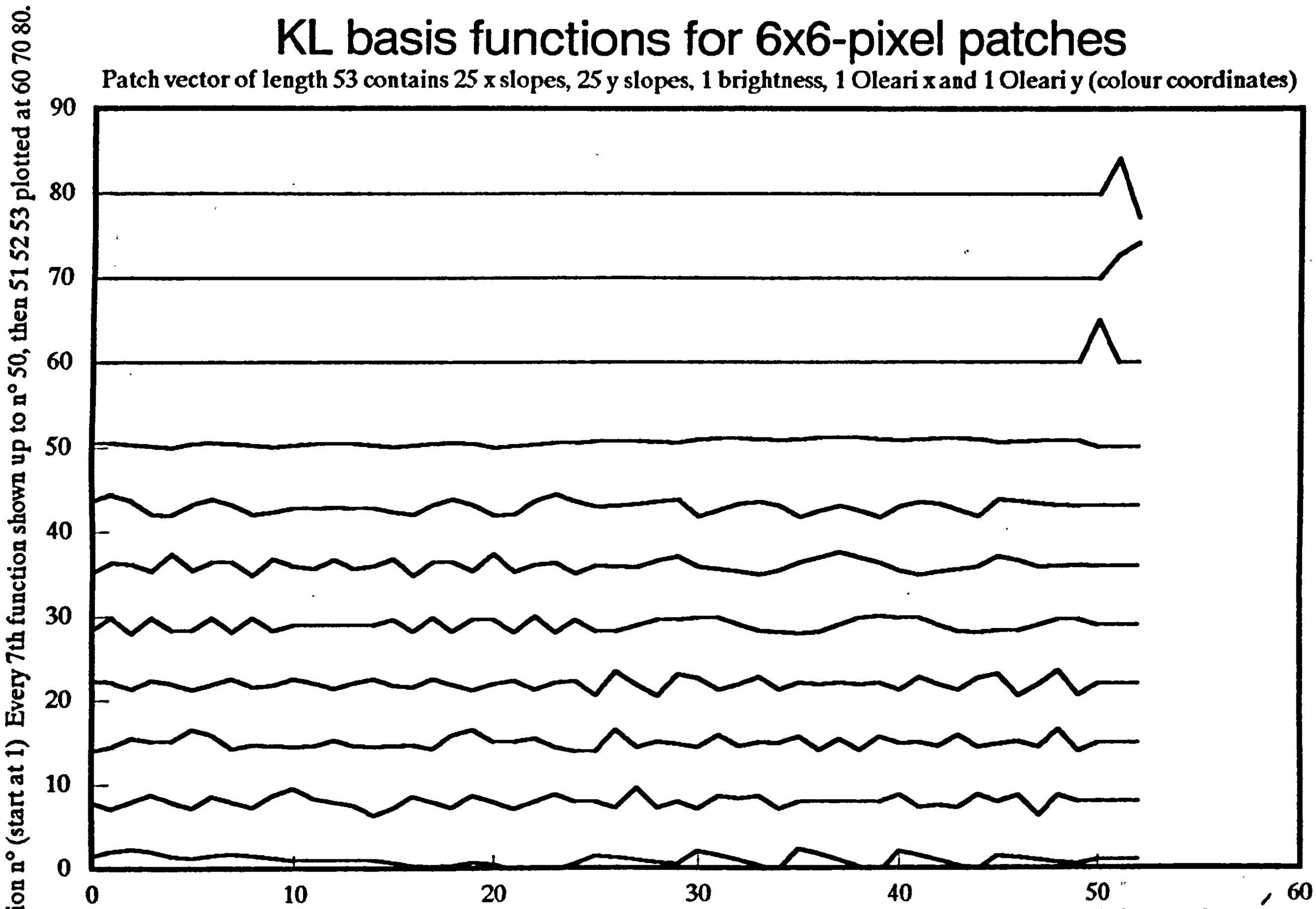
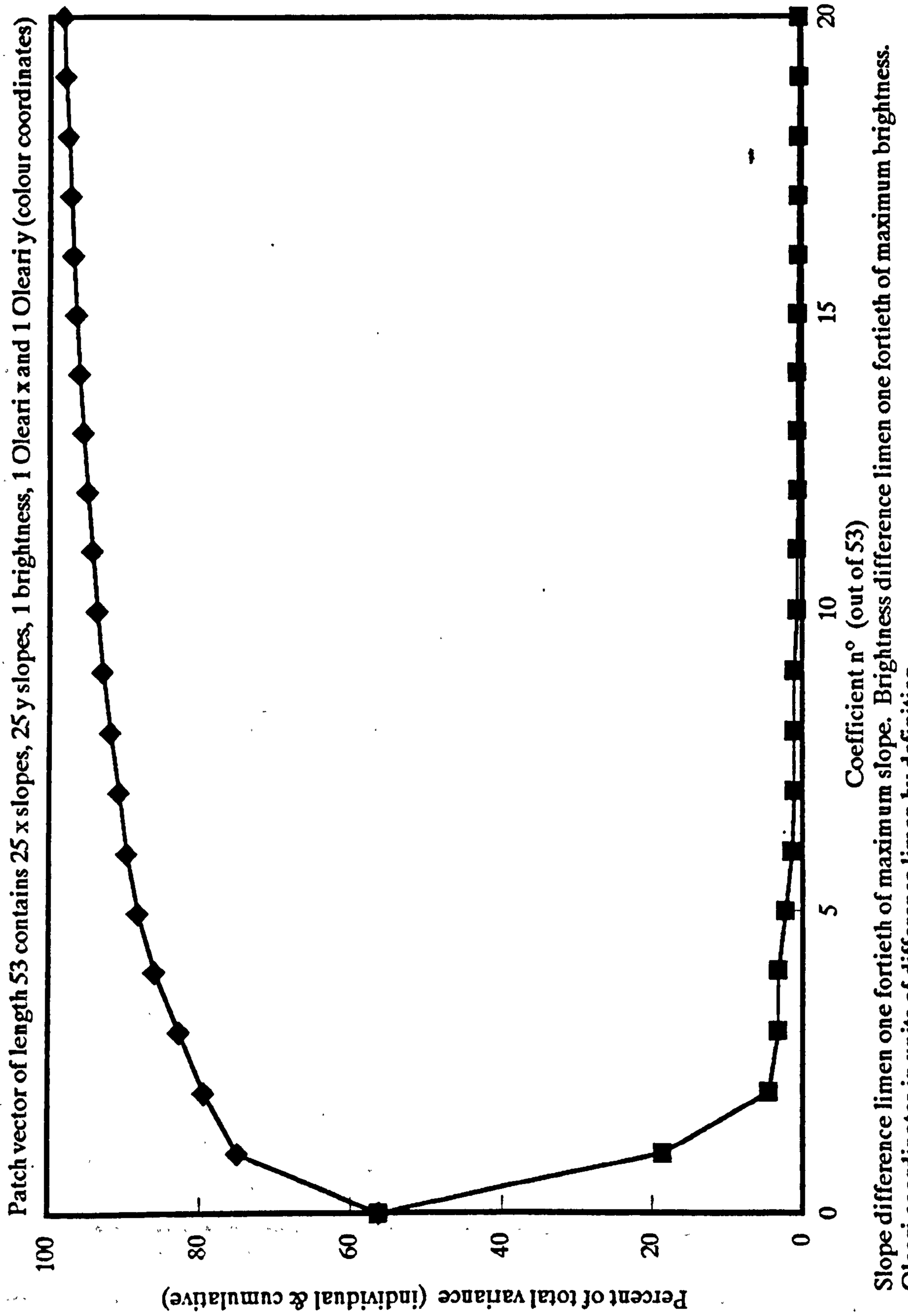

KL basis functions for 6x6-pixel patches
Patch vector of length 53 contains 25 x slopes, 25 y slopes, 1 brightness, 1 Oleari x and 1 Oleari y (colour coordinates)
Function n° (start at 1) Every 7th function shown up to n° 50, then 51 52 53 plotted at 60 70 80.
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
x-slope number (5 row + col) y-slope number + 25 N° 50 is brightness. N° 51 is Oleari x. N° 52 is Oleari y.
Slope difference limen one fortieth of maximum slope. Brightness difference limen one fortieth of maximum brightness.
Oleari coordinates in units of difference limen by definition.

Figure 9.9

Figure 9.10


KL transform coefficient variance
Patch vector of length 53 contains 25 x slopes, 25 y slopes, 1 brightness, 1 Oleari x and 1 Oleari y (colour coordinates)
Percent of total variance (individual & cumulative)
100
80
60
40
20
0
0
5
10
15
20
Coefficient n° (out of 53)
Slope difference limen one fortieth of maximum slope. Brightness difference limen one fortieth of maximum brightness.
Oleari coordinates in units of difference limen by definition.

Random 6x6-pixel patch

Generated from 50 random numbers and inverse KL transform giving 25 x and 25 y brightness gradients

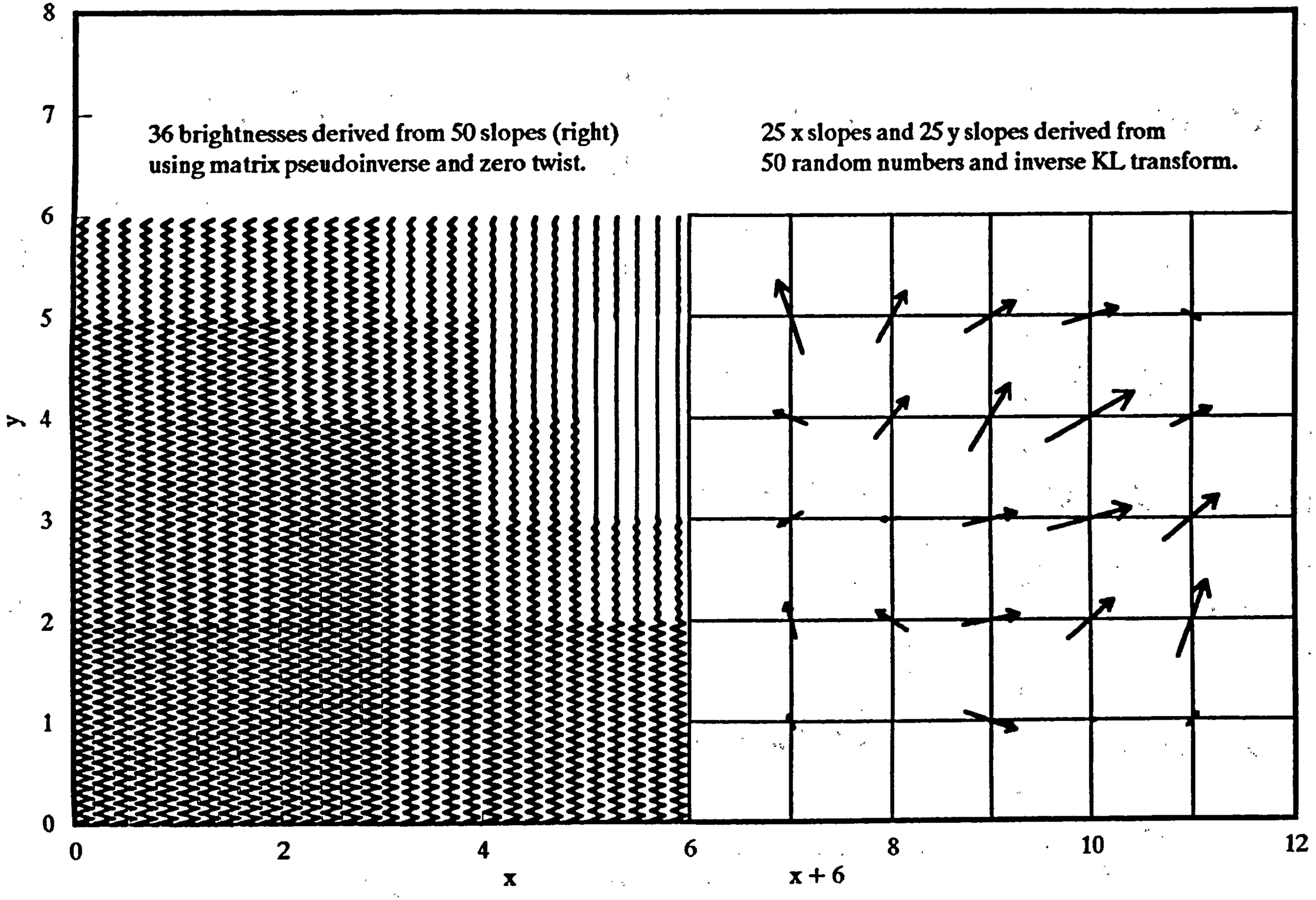
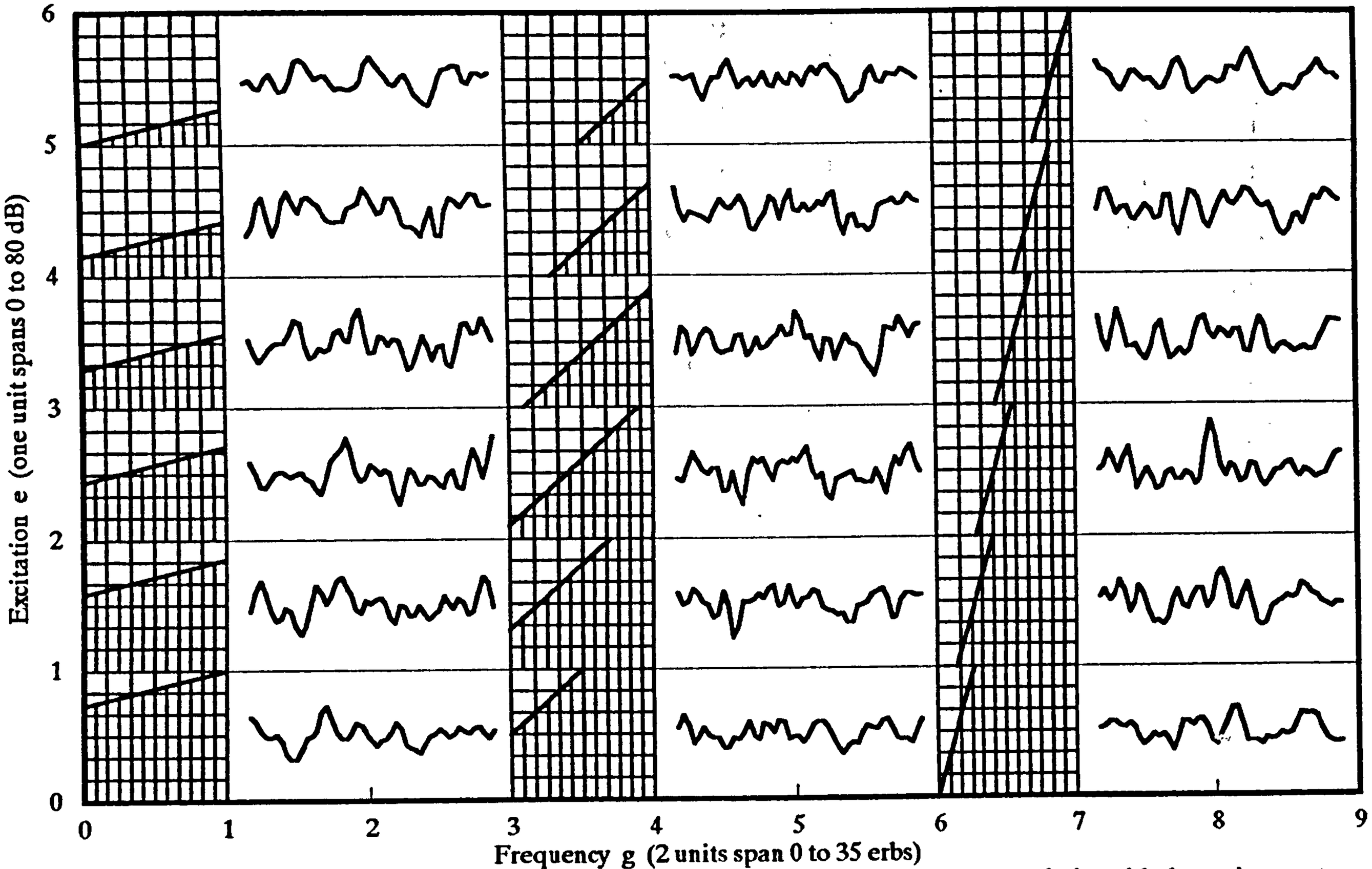


Figure 9.11

Spectrums resulting from straight edges

Calculated in turn: 36 brightnesses, 25 x and 25 y slopes, forward slope KL, inverse spectrum KL



1st three weightiest spectrum eigenvectors aligned with three overall colour coordinates, remainder with slope eigenvectors. Overall brightness = 17.0, Oleari x = −13.0, Oleari y = −34.0, all mean values and in DL units.

Figure 9.12

Figure 9.13

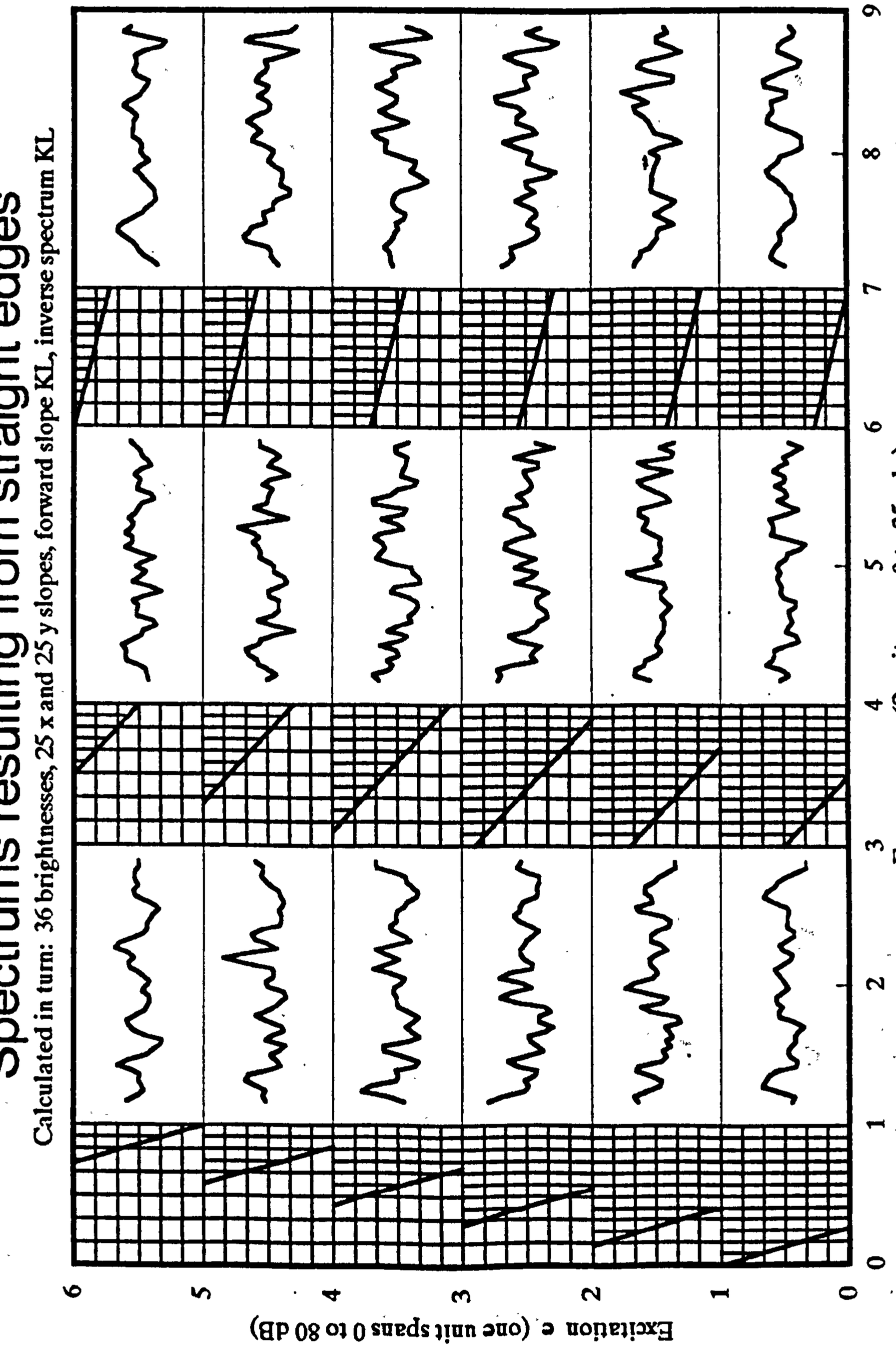
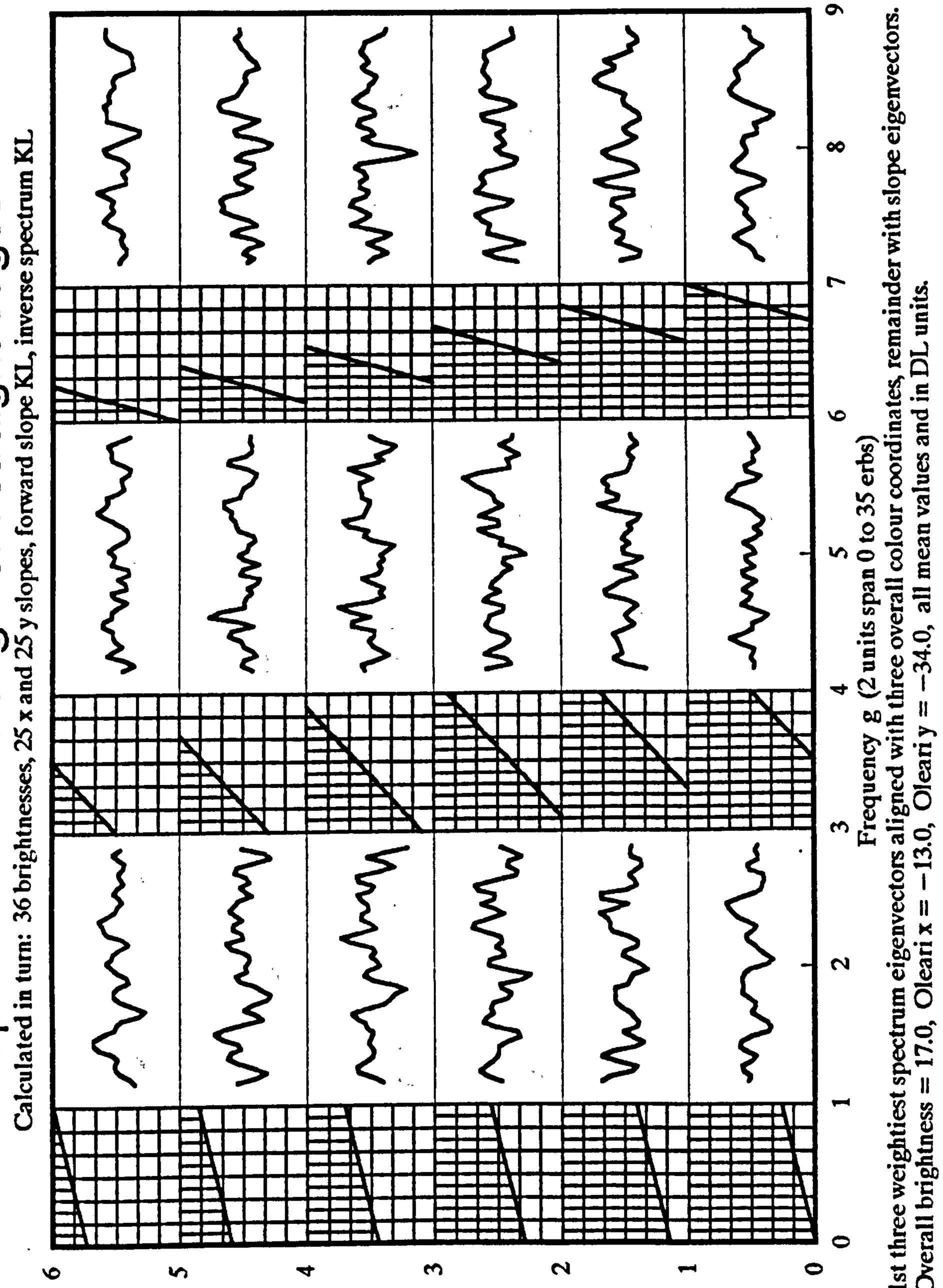
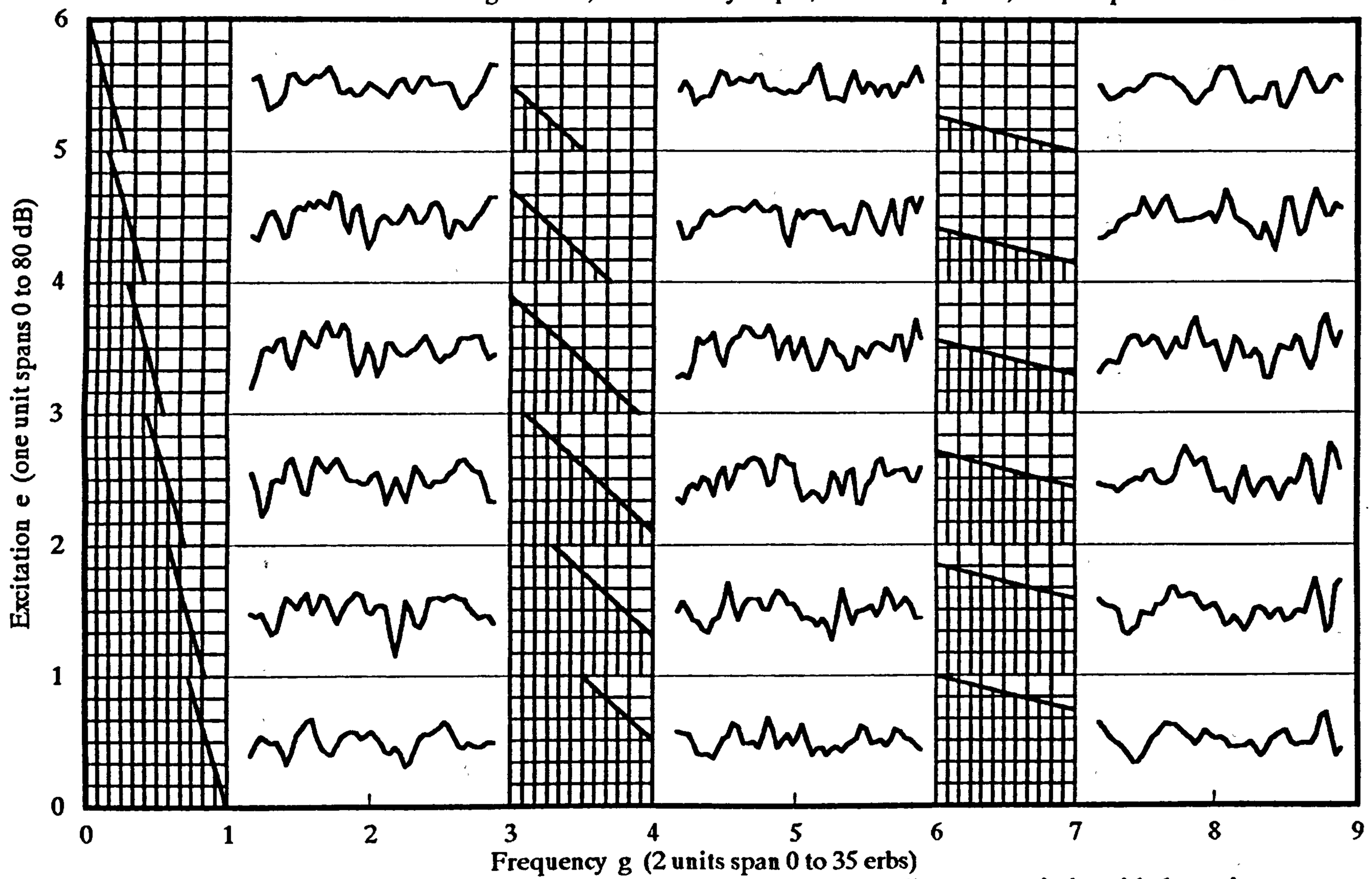

Spectrums resulting from straight edges
Calculated in turn: 36 brightnesses, 25 x and 25 y slopes, forward slope KL, inverse spectrum KL
Excitation e (one unit spans 0 to 80 dB)
Frequency g (2 units span 0 to 35 erbs)
1st three weightiest spectrum eigenvectors aligned with three overall colour coordinates, remainder with slope
Overall brightness = 17.0, Oleari x = −13.0, Oleari y = −34.0, all mean values and in DL units.

Figure 9.14


Spectrums resulting from straight edges
Calculated in turn: 36 brightnesses, 25 x and 25 y slopes, forward slope KL, inverse spectrum KL
Excitation e (one unit spans 0 to 80 dB)
Frequency g (2 units span 0 to 35 erbs)
1st three weightiest spectrum eigenvectors aligned with three overall colour coordinates, remainder with slope eigenvectors.
Overall brightness = 17.0, Oleari x = −13.0, Oleari y = −34.0, all mean values and in DL units.

Spectrums resulting from straight edges

Calculated in turn: 36 brightnesses, 25 x and 25 y slopes, forward slope KL, inverse spectrum KL



1st three weightiest spectrum eigenvectors aligned with three overall colour coordinates, remainder with slope eigenvectors. Overall brightness = 17.0, Oleari x = −13.0, Oleari y = −34.0. all mean values and in DL units.

Figure 9.15

Figure 9.16

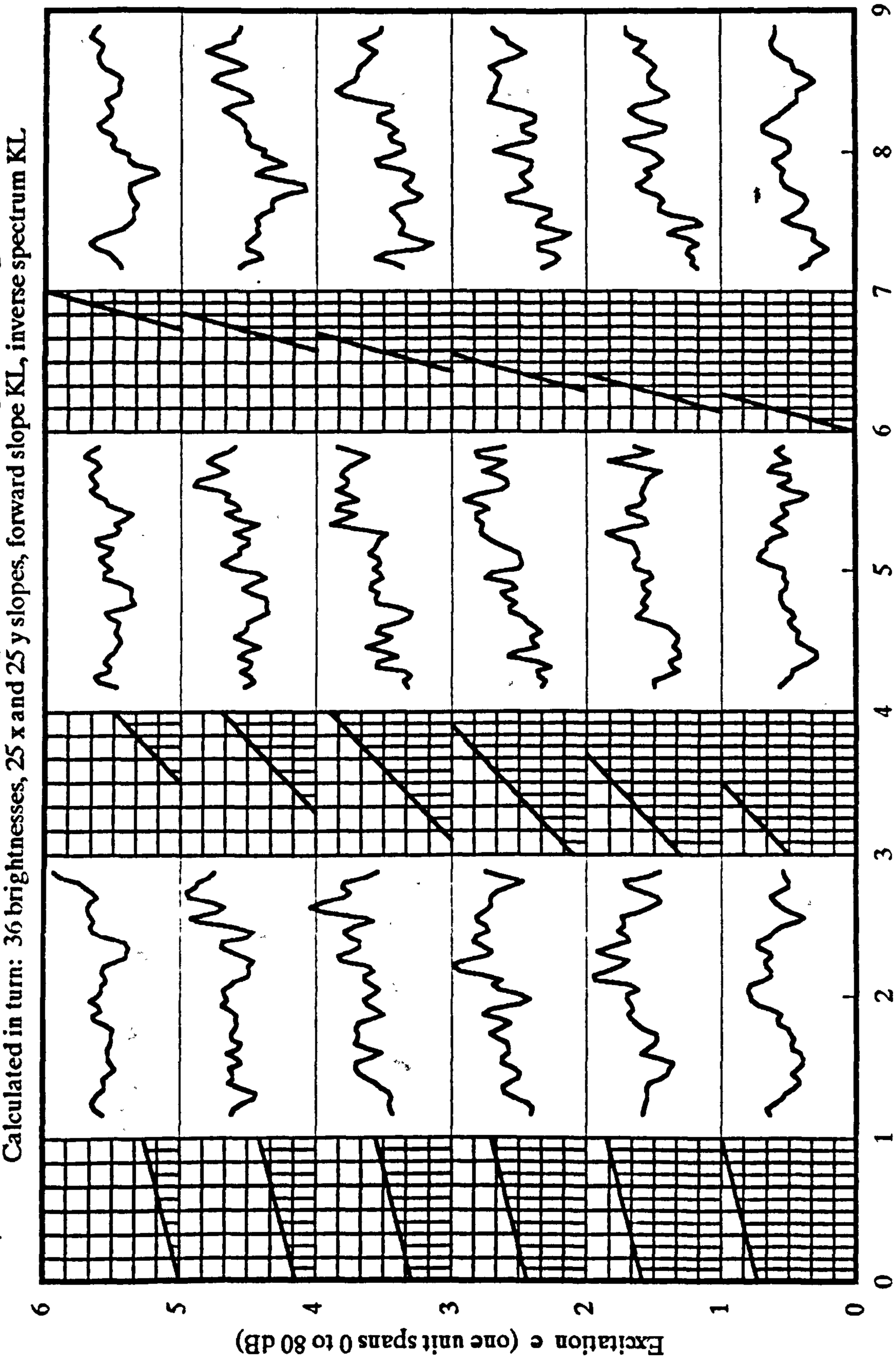

Spectrums resulting from straight edges
Calculated in turn: 36 brightnesses, 25 x and 25 y slopes, forward slope KL, inverse spectrum KL
Excitation e (one unit spans 0 to 80 dB)
Frequency g (2 units span 0 to 35 erbs)
Spectrum eigenvectors aligned with slope eigenvectors, colour coordinates ignored.

Figure 9.17

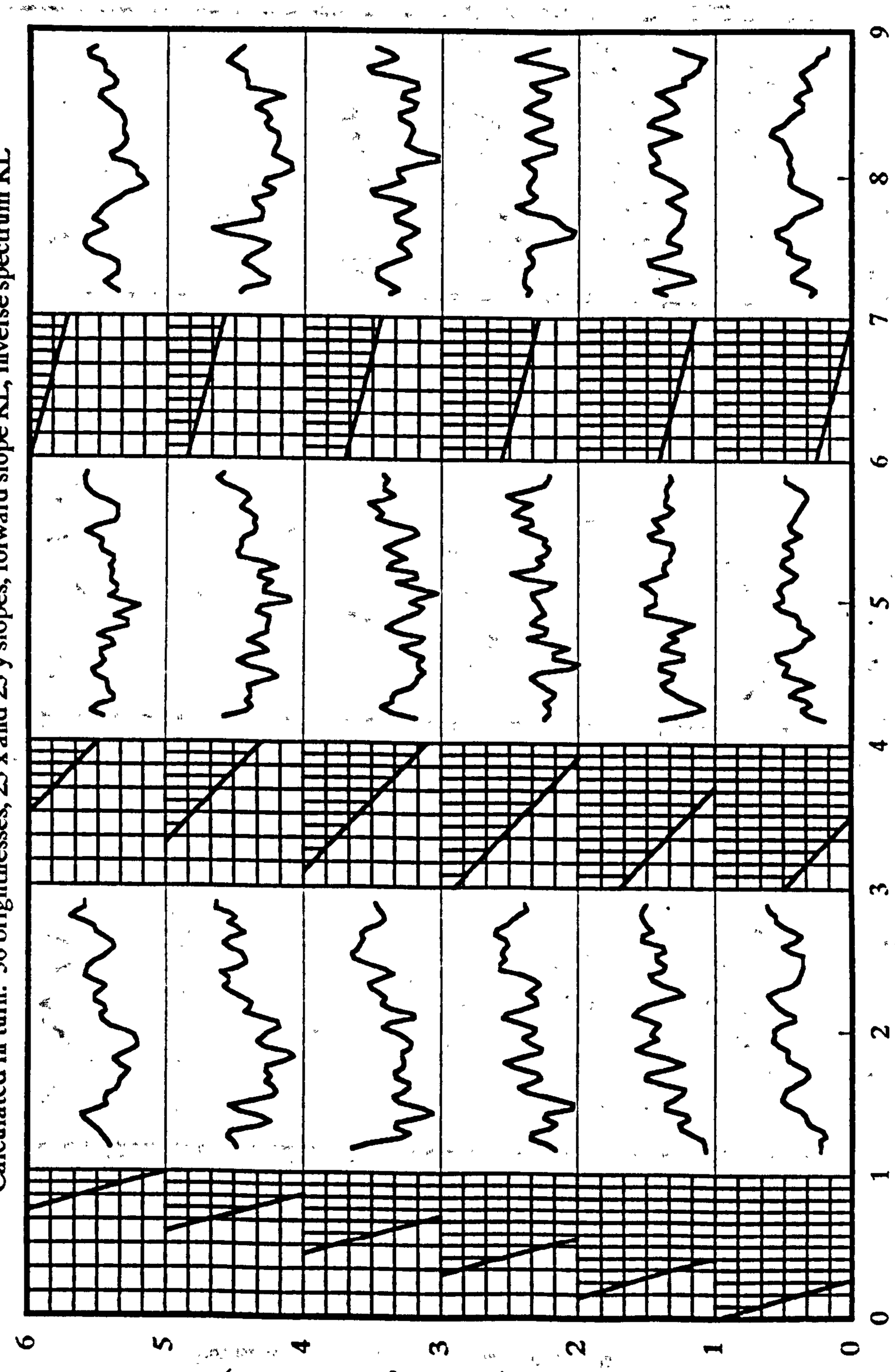

Spectrums resulting from straight edges
Calculated in turn: 36 brightnesses, 25 x and 25 y slopes, forward slope KL, inverse spectrum KL
Frequency g (2 units span 0 to 35 erbs)
Spectrum eigenvectors aligned with slope eigenvectors, colour coordinates ignored.
Excitation e (one unit spans 0 to 80 dB)

Figure 9.18

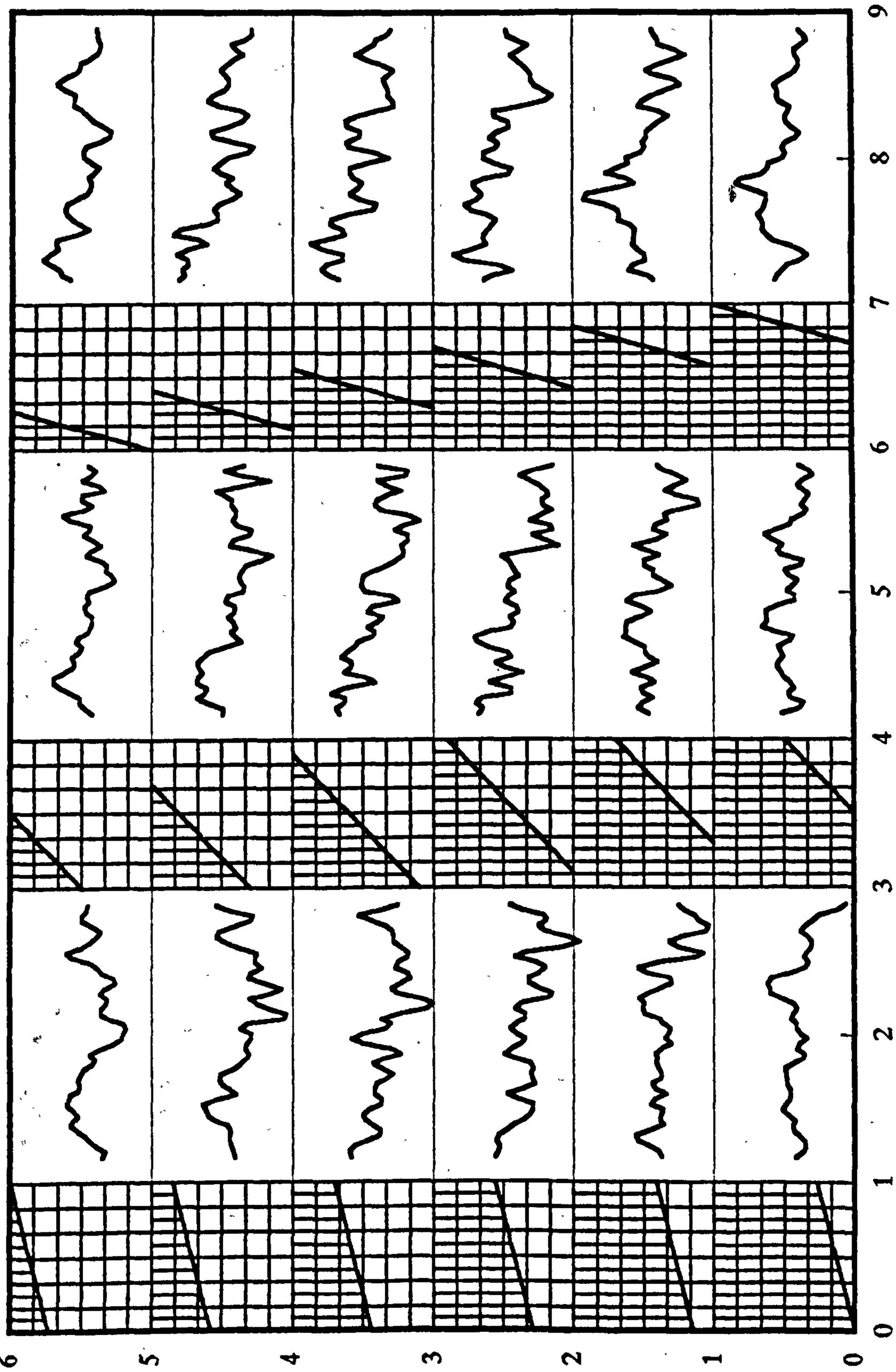

Spectrums resulting from straight edges
Calculated in turn: 36 brightnesses, 25 x and 25 y slopes, forward slope KL, inverse spectrum KL
Spectrum eigenvectors aligned with slope eigenvectors, colour coordinates ignored.
Frequency g (2 units span 0 to 35 erbs)
Excitation e (one unit spans 0 to 80 dB)

Figure 9.19

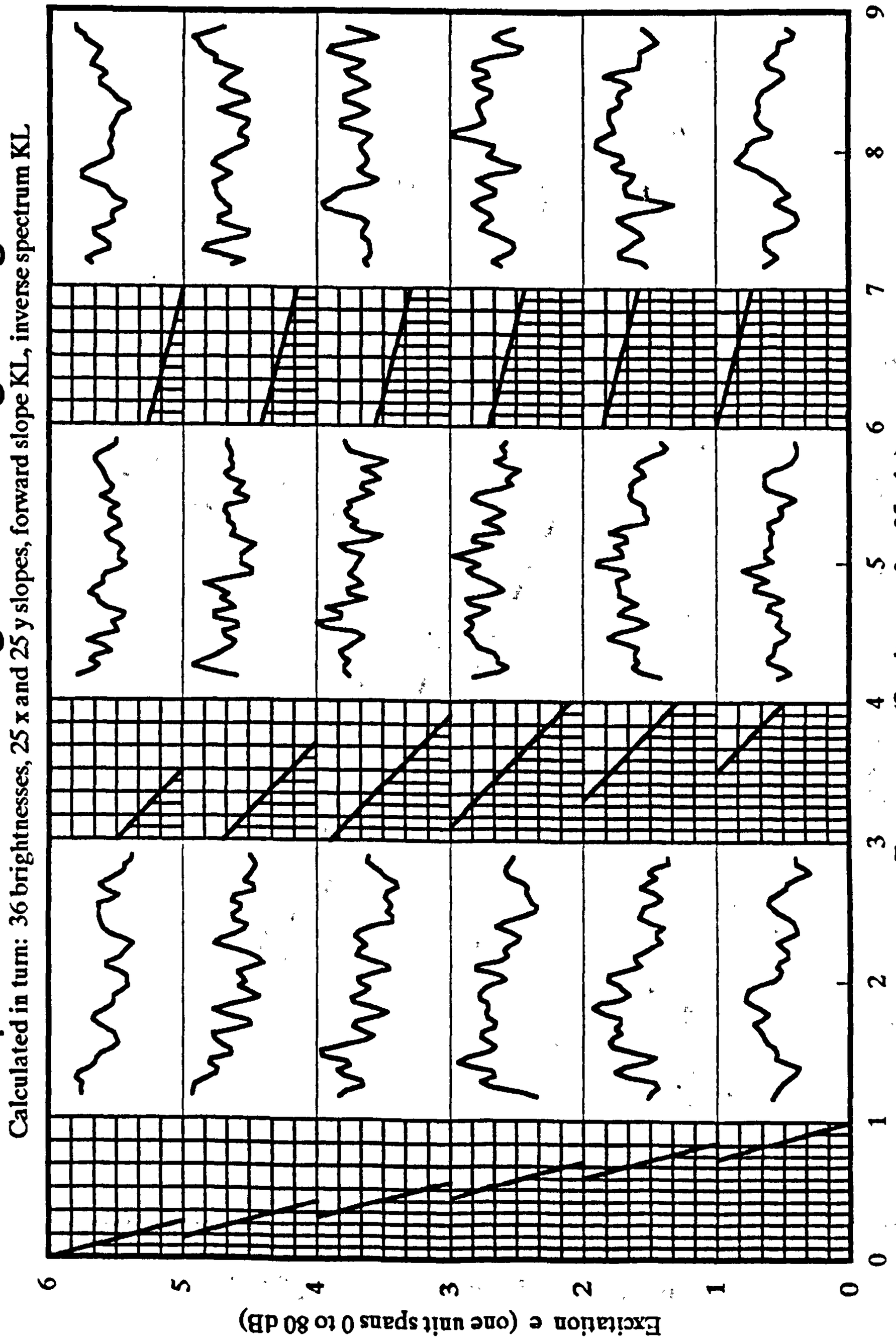
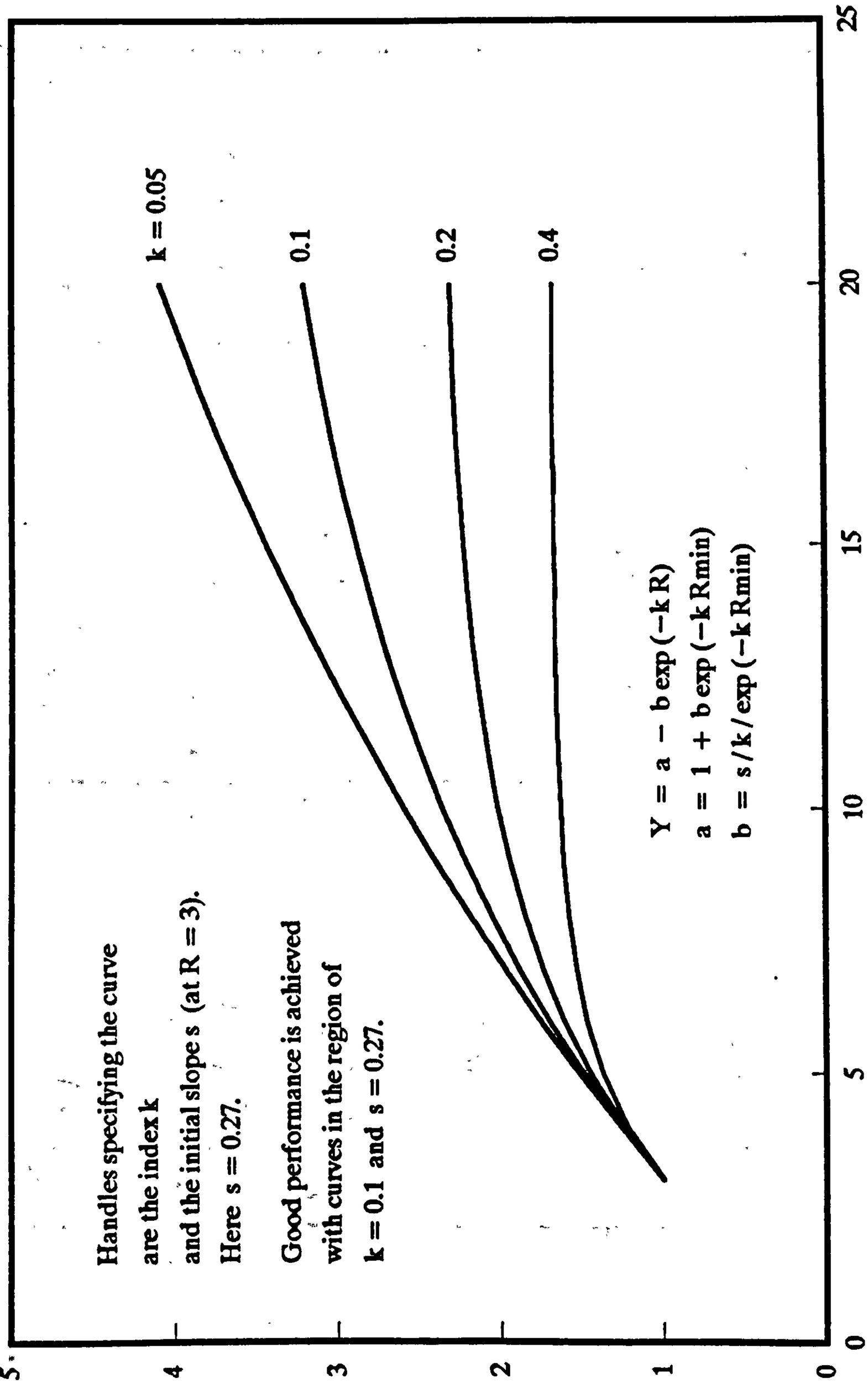

Spectrums resulting from straight edges
Calculated in turn: 36 brightnesses, 25 x and 25 y slopes, forward slope KL, inverse spectrum KL
Excitation e (one unit spans 0 to 80 dB)
Spectrum eigenvectors aligned with slope eigenvectors, colour coordinates ignored.
Frequency g (2 units span 0 to 35 erbs)
Spectrum eigenvectors
0
1
2
3
4
5
6
7
8
9

Figure 9.20


Patch-interest size bias
Sample bias curves to correct for effect of finite pixel size
Handles specifying the curve
are the index k
and the initial slope s (at R = 3).
Here s = 0.27.
Good performance is achieved
with curves in the region of
k = 0.1 and s = 0.27.
k = 0.05
0.1
0.2
0.4
Y = a − b exp (−k R)
a = 1 + b exp (−k Rmin)
b = s / k / exp (−k Rmin)
Patch radius R (pixels)
(Rmin = 3).
ghting bell.
Weighting arbitrarily set to 1 at minimum patch size (R
eights refers to the weights under the slope-weigh
Weighting Y applied to sumslopes/sumweights
Weightin
Sumweigh
0
5
10
15
20
25
0
1
2
3
4
5

One-parameter interest-weighting window

Window is circular, graph shows vertical section.

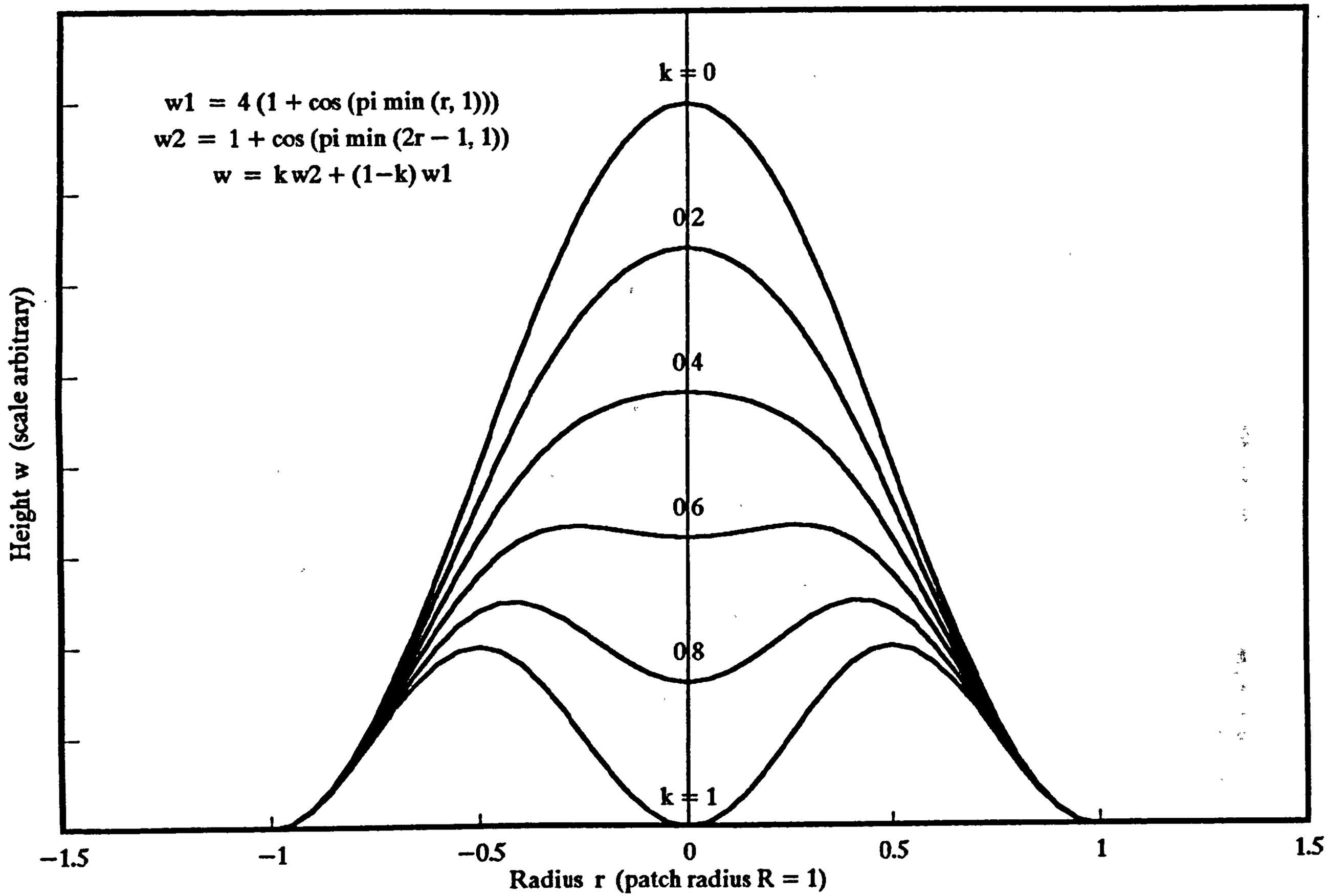


Figure 9.21

Figure 9.22

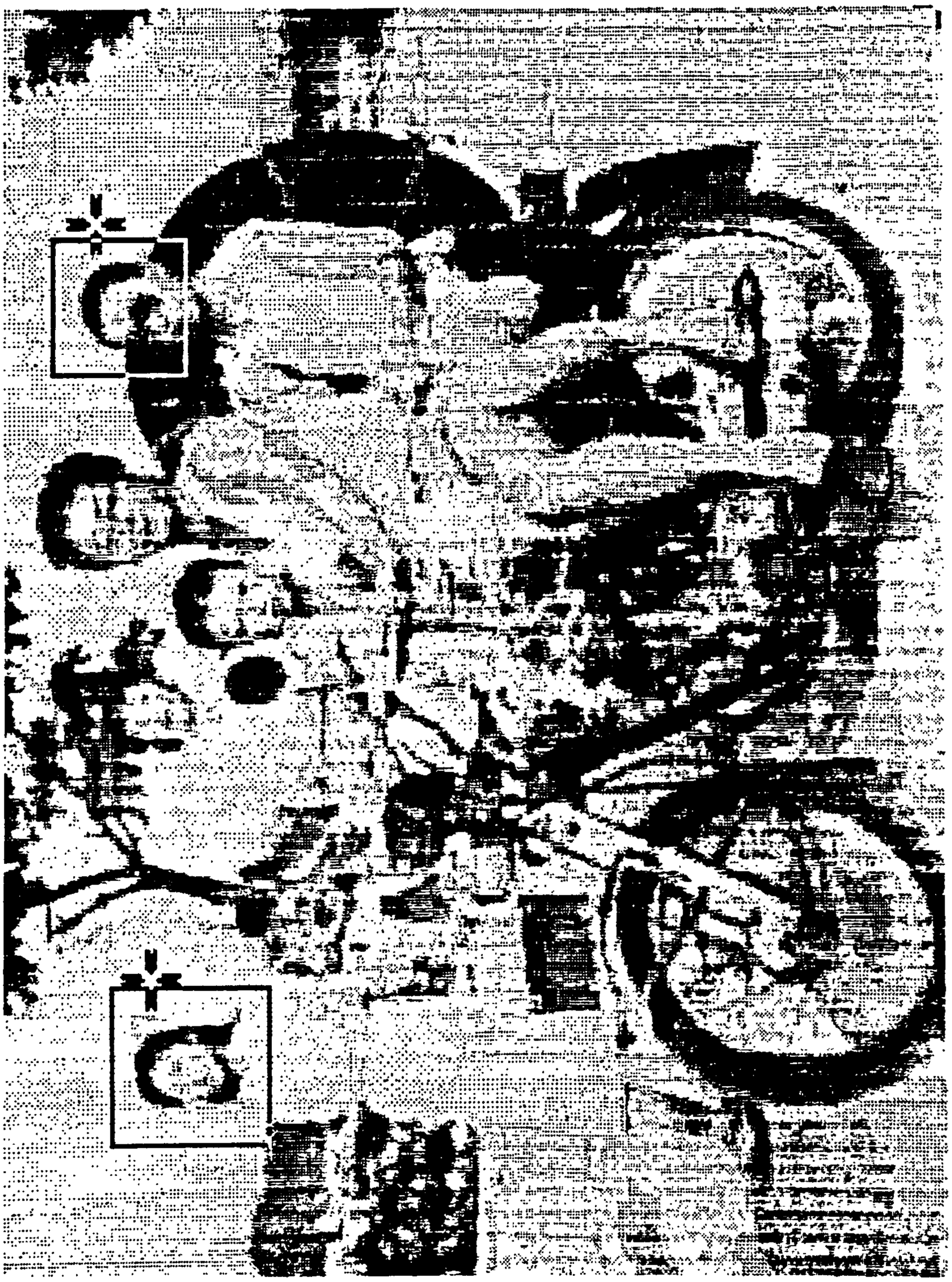
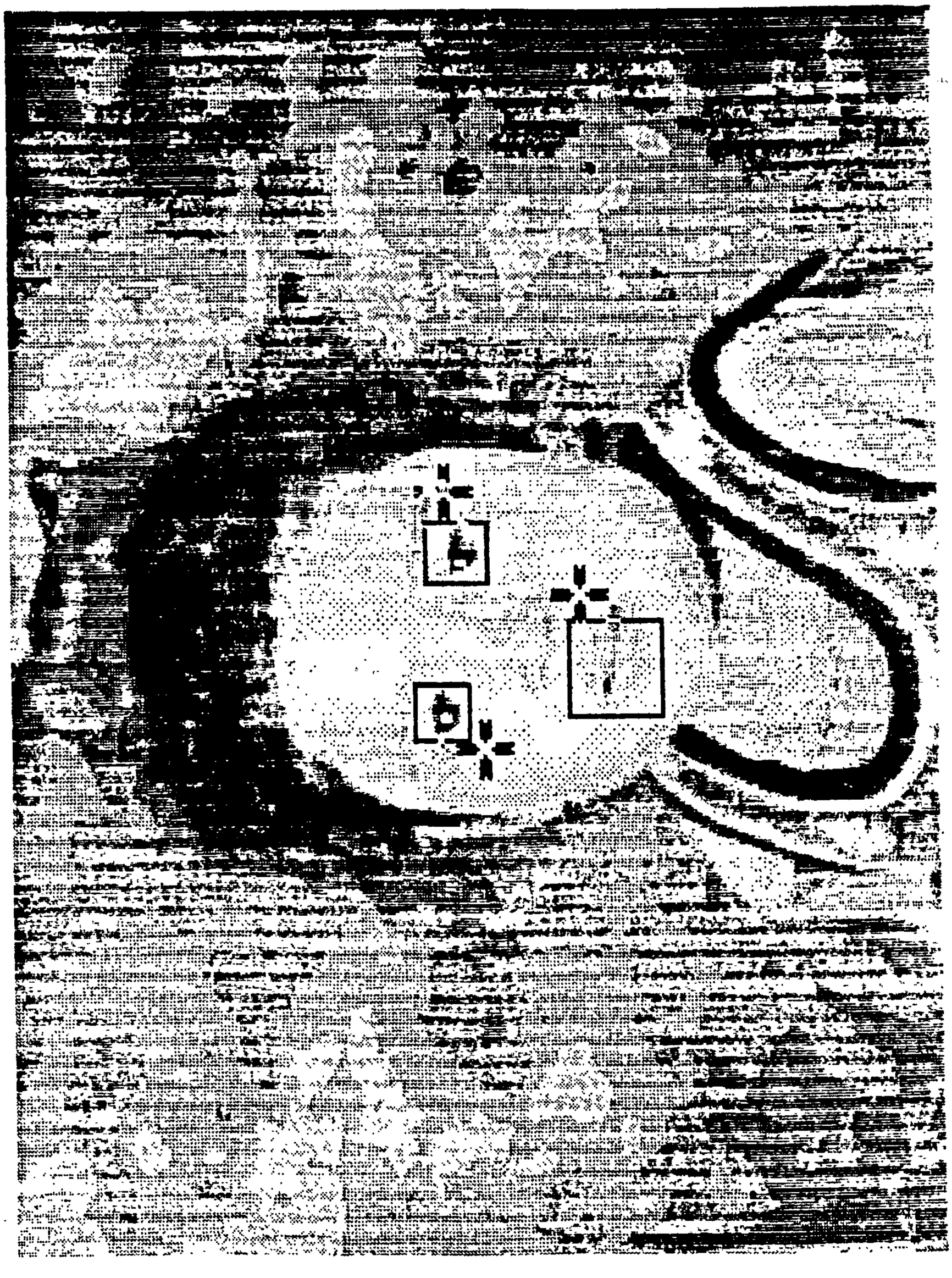


Figure 9.23



CHAPTER 10 CONCLUSIONS AND RECOMMENDATIONS

10.1 Recall of objectives

Before summarising what has been achieved and making recommendations for the next steps to take on the road to producing a marketable product, it is helpful to recall what we were trying to do.

The success of Fish's flying-spot scanner systems (Fish 1976) shows that there exist schemes allowing people to discern shapes and objects by means of an auditory signal derived from information captured by a TV camera. Two questions arise. First, if Fish did it, what are we trying to do? Second, if Fish did it, why is his system not commonplace?

The success and failure of the flying-spot scanner both have the same cause. That cause may be described as underambition and overexplicitness. It is a characteristic of Fish's system that is important to understand but difficult to explain.

Basically the system works because it laboriously lists the positions of the edges in the scene in such a way that with a bit of effort you can't go wrong.

On the other hand, such a method is only possible for simple shapes or objects rather than whole scenes, and it is this lack of usefulness, together with the limitations of hardware in 1976, that led to its failure.

Nevertheless, I should state that in my opinion a cheap well turned-out version of Fish's flying-spot scanner in modern hardware would probably sell.

Any improvement on Fish's flying-spot scanner must therefore be more ambitious while at the same time not scorning its qualities. The flying-spot scanner is limited by the fact that it is a point mapping. The difference between a point mapping, a slot mapping and a patch mapping was explained in Chapter 2. The distinction refers to which part of the scene generates the sound (the spectral content of the sound) at any instant.

The dimensionality of a sound spectrum as experienced by the human ear is of the order of 50 (Chapter 8 was exclusively devoted to this question). The dimensionality of the information available at a point in a scene depends on the scheme - from one for brightness in a black and white scheme to maybe six in a colour scheme using edge strength and orientation - in any case nowhere near 50. Thus a point mapping inherently underuses human hearing.

The objective of the present research is to look at schemes that are not inherently limited in any way. That is to say, any limitations must be those of human hearing only, and no more. Such limitations are well documented, while limitations on the ability to learn to recognise codes that <u>are</u> audible are speculative and controversial. There is an argument, amazingly, in favour of censorship of the scene in order not to overload the poor user's senses. This argument, originally doubtless a mistaken reaction to past failures, unfortunately at one time gained a certain political correctness. The research avenues rejected on this basis are of course not known.

Both slots and patches can have any chosen dimensionality. The objective of the present research was thus to look at slot and patch mappings - but is there anything else? Well, there's Mead's SeeHear (Nielsen et al 1989), which produces a continuous sound based on the time differential of brightness. It could be classed as a patch transform with only one patch, and that patch as big as the scene. However, this doesn't do justice to the very different nature of the resulting scheme as a whole: no worries as to patch size or as to which patch to sound next. Better introduce a new class and call it a synchronous mapping.

Synchronous mappings are at the opposite extreme of the range from point mappings. At first glance they appear

to offer the perfect solution: just put it on and listen to a continuous sound, which will only change if something in the scene changes. Then difficulties rapidly appear. The whole scene can have a dimensionality of order 50 only. Massive data compression is then applied to the captured scene in an attempt to achieve this. Unfortunately, although only 50 numbers result, they are highly decorrelated and not matchable to the 50 correlated numbers describing a spectrum. Synchronous mappings turn out to be as limited as point mappings.

Most of the present research concerns slot mappings. It would have been nice to spend more of the time looking at patch mappings - unfortunately these are much more difficult to invent. Whether a slot mapping or a patch mapping turns out to be best is something that remains to be proved - my bet is on a patch mapping.

10.2 Achievements

10.2.1 General

What's been achieved is mainly a greater insight into the general problem of optophonics (GPO), to an extent that it is now possible with considerable confidence to say (in the section on recommendations) what should be done

next in the way of developing further mappings and testing them, in order to arrive at the stage of having a marketable product.

10.2.2 Theoretical performance test (TPT)

An early achievement was a theoretical performance test to which invertible mappings can be subjected. Despite its limitations, discussed at length in Section 3.2.6, the TPT has the merit of guaranteeing an upper bound on the performance of a mapping. That is to say, if the TPT says that certain detail or information will be lost, then it will be. This is because the TPT is based on the phenomenon of masking to be found in human hearing and because this phenomenon is very well documented.

10.2.3 Statistics of sounds and scenes

Various statistics of sounds and scenes, as perceived, have been measured. This has been done using no great data base, the purpose being to get a feel for the statistics rather than superexact results. The exercise may be repeated using a large data base at a future date if thought necessary. However, this is not one of the recommendations.

A more important question is what statistics are measured. It was shown that the statistics of brightness gradients are more informative than those of brightnesses, in which such characteristics of natural scenes as the coherence of edges are lost (Section 9.2.4). Also important is how the scenes (or rather patches therefrom) are selected, centred and sized, since that affects both their content and relative frequency, and thus their statistics.

10.2.4 Schemes 1 and 2

The four schemes examined were not all comparable. Schemes 1 and 2, the cartesian piano and cosine transforms, were not even sounded and only subjected to the theoretical performance test (TPT). The TPT showed the cosine transform to be worse than the piano transform, so Scheme 2 is discarded.

Scheme 1, the cartesian piano transform, is discarded for the four reasons given in Section 7.1. Should anyone nevertheless wish to continue with it, it is worth pointing out that it has been implemented by Meijer (personal communication) in semiportable form.

10.2.5 Scheme 3 - polar piano transform

Scheme 3, the polar piano transform, has several attractive features, not all of which yet have an equivalent in Scheme 4. The faults of Scheme 3 were

1 Colour wrongly mapped. Hue was mapped to musical key, and saturation to musicality, where musicality represents the local relative loudness of the pure tones (Figure 7.7) and the rest of the sound (Figure 7.5). The problem comes with colours that are nearly black, like Shanti's hair (Figure 7.7). Suppose the colour is (r, g, b) = (0, 0, 1), where white is (255, 255, 255). The formula for saturation causes this to be treated as a highly saturated colour, whereas subjectively it is very nearly unsaturated (black). The effect is that (0, 0, 1) sounds very different from (0, 0, 0), a violation of our continuity criterion (Section 1.8). A small fault, but one that needs correcting.

2 Limited invariance to scene translation. This is a fault inherent in Scheme 3, and can't be changed. What isn't known is the long-term ability of a user to recognise the shape of a wholly off-centre object (centre of scene

wholly outside object boundary). The short-term ability can be assumed zero, but the long-term ability is the one that matters.

3 Scheme 3 is another slot transform, meaning that the sound at any instant is related to the brightness distribution along some line drawn across the scene. It would be surprising if this turned out to be the best arrangement, better than relating the sound at any instant to the brightness distribution in a 2-D window on to the scene.

10.2.6 Scheme 4

Scheme 4 was only partly examined. It is in particular not yet clear whether the KL method can be used successfully to map brightness patterns in a patch or window on the scene to the spectral patterns of a steady sound. Recall that the problem with the KL method is that each basis function is arbitrarily signed, which means that it works just as well if it and its coefficient are both multiplied by -1. The small size of the patch, around 6x6 pixels, suggests that, with some effort, the problem may be soluble by trial and error.

10.3 Recommendations - mappings

10.3.1 General

Two promising and fundamentally different mappings (Schemes 3 and 4) have been developed in the present research. Both of these should be tested in prototype optophones, both for their own sake and because two is the minimum number of mappings necessary to develop the tests on. It will be remembered from Chapter 3 that no such general tests at present exist, although Fish (1976) developed ad-hoc tests which do adequately describe the performance of his mappings.

Although it is recommended that Schemes 3 and 4, as developed here, be tested, that should not be taken to preclude the testing of any other schemes that might be thought up, provided there is prima-facie evidence of superiority over those schemes. Indeed it is intended that the testing of Schemes 3 and 4 should suggest improvements to them. The point is that even if the changes turn out to be so great as to warrant a change of name, then that's fine too.

10.3.2 Tackling PIA

For instance, how should PIA be best investigated? Would

it be possible to adapt one of the other two or is a whole new scheme called for? Remember PIA, the property of inconsequential ambiguity? It is tempting to sweep it under the carpet, but this is only allowable after it is shown (and if it can be shown) that any scheme with it will be worse than Schemes 3 or 4. The first thing to do then is to try to show that there's nothing in it. There's nothing wrong in that, provided it's done conscientiously. If that fails, then there is no choice but to investigate it.

The example of PIA we came across (Section 9.2.8) concerned the retention of the information in Fourier magnitude and the abandonment of the information in Fourier phase, the object being to obtain a sound locally invariant to lateral displacement of the patch. It was then realised that if the resulting ambiguity only occurred in theory but not in practice, there would be a secondary advantage, namely the exclusion of some theoretically possible but in practice nonexistent scenes from the domain of the scene-to-sound mapping, thus freeing the sounds that those scenes would otherwise have mapped to but would never in practice have used.

Unfortunately (or not, depending on whether you hope PIA works), this particular example is not at first sight encouraging. Oppenheim & Lim (1981) showed convincingly that a scene reconstructed from its Fourier transform

with the magnitudes unchanged but random numbers used for the phases is unrecognisable, but that if the correct phases are used and random numbers or a constant value given to the magnitudes then a recognisable reconstruction of the scene results.

On the other hand, this is not exactly what we are trying to do, and the question needs looking at in more detail. Nawab et al (1983) show that signals containing not too many adjacent zeros can be reconstructed completely from the magnitude of their short-time Fourier transform, and that the results can be extended to two-dimensional images. This seems to correspond to where we first encountered PIA in Scheme 4, namely magnitudes of Fourier transforms of patches in scenes.

Here the researcher's general approach to optophonics becomes critical. What is his immediate conclusion from the above statements? Remember the situation: the user is supposed to hear a sound derived from the Fourier transform magnitude of patches taken from the scene, and from that to reconstruct the scene in his head. Does the researcher try to decypher Nawab's algorithm, and throw up his hands in horror at the thought of asking the user to do it in real time in his head? Or does he breezily announce that the user will soon get used to it?

The user is not being presented with the Fourier

transform magnitude but with "a sound derived from the Fourier transform magnitude...". This derivation, based on the KL method developed in Chapter 6, is designed to produce a sensible sound.

The only sensible approach is to try it and see. Unfortunately, this means implementing it in one of the prototypes. I think the method will probably work.

However, even if the method works, what will it tell us about PIA? How will the benefits of using sounds otherwise abandoned (by being matched to nonexistent scenes) manifest itself? Less scenes are being mapped on to the same number of sounds. How many times less?

In order to assess a scheme with PIA, we not only need to show that it works but that it is better. If the tests are properly designed then any improvement will show up in the tests. Nevertheless, in order to help decisions at a much earlier stage, it would be nice to have a theoretical measure of the benefits of a scheme with PIA.

When we look at the negative of a photograph, we know we are looking at the negative of a photograph, and not at a photograph of something else. Suppose we decided to exploit this example of PIA in an optophone by arranging the sound produced by a scene to be the same as the sound produced by the negative of the scene. Since every

natural scene has exactly one negative, we would be halving the number of possible scenes. Each scene would therefore be able to take up (map on to) twice as much sound. Is there therefore the notion of the amount of sound required by, or available to, each scene?

Perhaps the amount of sound could be measured in information. If N numbers specify a sound and each number can have one of n distinguishable values, and if the numbers are uncorrelated, then the number S of possible different sounds is n^N, and the information content of a sound is $\log_2 S$ or $N \log_2 n$. In fact the numbers in a sound PR are correlated, and calculating the information content is more complicated (see Figure 8.6). Nevertheless, the point to note is that the information is broadly proportional to N and to the log of n.

In the two cases we are comparing, we are either mapping all positive scenes to S/2 sounds without PIA, or to S sounds with PIA. In the first case the information available per scene is (roughly) $\log_2 S/2$ or $\log_2 S - 1$, and in the second $\log_2 S$, a difference of only one bit, namely the bit that would tell us whether the picture was positive or negative. Hardly worth making a fuss about.

What is the corresponding factor in the case of the Fourier magnitudes? For each true-to-life scene, how many nonsense scenes are there with the same Fourier

magnitudes? Working simply as before, suppose we have say N/2 independent phases and N/2 independent magnitudes per scene, and that each phase can have n distinct values. There are then $n^{N/2}$ different scenes corresponding to every set of magnitudes, 1 scene true to life and $n^{N/2} - 1$ others. In the first case we are mapping all true-to-life scenes to $S/n^{N/2}$ sounds, and in the second to S sounds. The information available per scene is roughly $\log_2 S - (N \log_2 n)/2$ in the first case and $\log_2 S$ in the second, a gain of $(N \log_2 n)/2$ bits per scene for PIA. This seems worth pursuing.

Note that the discussion has been kept very simple and that no distinction between scenes and patches has been made.

10.3.3 Scheme 3

Scheme 3, the polar piano transform, may be tested first as it stands, with the exception that the way colour is mapped should be corrected (see Section 10.1.5).

After that, however, there are no restrictions on what may be varied in order to try to improve it. The tests (see below) should be of such a nature that they show convincingly

first

(a) what variable to adjust next

and after some work

(b) what value for that variable is best.

It is not possible or desirable now to guess or plan what might be the course of events under (a) above.

10.3.4 Scheme 4

Scheme 4, the free-field patch transform, will require to be completed before it can be tested. This means

1 Deciding on a sensible method of choosing, in the final optophone, the next patch to sound. This is presently done by a weighted interest function (Section 9.2.8 and Figure 9.21).

2 Calculating the statistics of patches chosen in this way.

3 Deriving a KL transform for the patches, as has been done several times for other things in this thesis.

4 Changing the sign (and sometimes within limits

the order) of the patch and sound basis functions being matched until some sensible results are obtained, namely distinctive features matching distinctive sounds.

This item 4 will be long and tiresome and highly subjective, and for me would have been very interesting. The reason is that the quantitative sensibleness of whatever pairing list is tried is guaranteed by the method (barring mistakes), provided the order of the basis functions is not disturbed other than by swapping functions of equal or nearly equal eigenvalue.

The human input will supply any qualitative sensibleness achieved. To my knowledge this task has never before been attempted, and there is no knowing whether it will work. If it doesn't, then you will have to proceed with a mapping based on randomly signed basis functions. We know already that these do not produce namable sounds from namable shapes. What we don't know is whether that matters.

A third option, of choosing a patch-to-spectrum mapping entirely subjectively and bypassing the KL method altogether, is not recommended. First, the options become even more numerous, and second, the resulting mapping will not even be quantitatively sensible.

Having chosen the method of selecting the patches to sound and thence the patch-to-spectrum mapping, there remains to derive the method of presenting the sound so that the user can tell where in the xy plane the sound is supposed to be coming from. My starting point would have been Hirakana & Yamasaki (1983), who derived direction-dependent impulse responses to be applied to any sound signal to make it appear to come from anywhere on a sphere surrounding the listener's head (apart from the neck), with checks as appropriate against Wightman & Kistler (1989), Makous & Middlebrooks (1990), Wenzel et al (1993), and whatever else might turn up.

There now comes another example of the importance of the researcher's general approach and attitude to life. The question is: what is the appropriate scaling between the angle subtended at the camera between two points in the scene and the angles used to generate the sounds of the patches centred on those two points? The question arises because the camera might range from say -30° to +30° elevation (0° being dead ahead), while the work of Hirakana extends fron -30° to +210° (and sounds actually sound different throughout the whole circle). Similar remarks apply to azimuth.

Or even, going back a step, does the researcher ask himself the question at all? If not then the scaling is automatically 1, and there arises approximately a ten-

fold loss in localisation accuracy (the location of ten times fewer patches can be distinguished since they are all crammed into the forward field of vision of the camera).

Even if the researcher thinks of the question, he may decide that a distortion would be confusing to the user, and still use a scaling of 1. This is the kind of patronising preemptive censorship that makes my blood boil. To my mind, the distortion would disappear within a few days' use at most, and there can be no excuse for such a reduction in performance.

Happily, in this case no such suggestion has been made or is likely to be made by anyone, so I can use language appropriate to the sin without causing personal offence.

10.4 Recommendations - Tests

To begin with, concentrate on those features of the mapping that should be immediately accessible. For instance, concerning the colour mapping of Scheme 3, it should be much easier to tell the colour of the centre of the scene than of any off-centre object, because only the centre of the scene ever occupies the whole spectrum and prevents other colours being sounded simultaneously. Any peripheral colour perception would only come very much

later and be a bonus.

Initally, of course, one can expect no sensation of colour at all. One would just notice that, in a colourful environment, there was one period around the middle of each sound which was in a definite musical key. A musical key (chord) is a common enough sensation. The question will then be how to learn to associate these keys to colours. One idea would be a colour chart hung on the wall, which the user could check against at will.

The same idea might be used for objects. One could place some objects on a table, and similar objects elsewhere in the room, say on shelves on the walls. The idea would be to handle an object on the table, to know what it was and what it sounded like, and then explore the rest of the room, at a distance, for something sounding similar.

Such an exercise should prove very instructive. Initially, the objects might be placed on the shelves with the same orientation and lighting as on the table, thus guaranteeing a similar sound. Later, the orientation of the objects on the shelves might be kept secret, and one would have to remember all the sounds the object on the table made as it was manipulated in order to find it on the shelf.

The point is to start from something guaranteed to work,

and move on from there. This presupposes that the schemes being tested have aspects that are guaranteed to work - something to remember in designing them.

See Fish (1976) for some initial ideas.

10.5 Recommendations - hardware

10.5.1 Two classes of hardware

There are two classes of hardware, both optophones, that it is essential to distinguish, namely the prototypes designed to test the mappings on the one hand, and the first marketable product on the other.

10.5.2 Prototypes for Schemes 3 and 4

The time has come (April 1994) to build a prototype optophone. The reason is not that a mapping has been perfected but that an optophone is necessary in order to perfect any mapping.

Be careful not to let difficulties of implementation distort or destroy the characteristics of the mapping it is desired to test. The research so far has been guided

by a deliberate policy of ignoring hardware and real-time algorithmic questions. If at all possible, these should not be allowed to gain the upper hand now. For one thing, only one of the two schemes will ever appear in marketable form.

Beware of "testing something else because it might be simpler", of answering questions that aren't being asked. For instance, I expect that a prototype capable of performing the mappings of Schemes 3 or 4 will be very complicated, with the computational hardware probably desk-bound and mains-powered. This may be necessary in order to have the flexibility to try out major or minor variations, a flexibility not required in the final product.

It is to be expected that the prototypes for Schemes 3 and 4 will have many components in common. Whether one talks of one or two prototypes will just be a matter of choice.

10.5.3 First marketable optophone

The final product will be relatively inflexible, with only those parameters still adjustable as have proved necessary in the early tests.

These may be of two kinds. First, parameters that need adjusting as the user's competence develops. Presentation speed comes to mind as one of these.

Second, parameters that need adjusting according to the task at hand. It may turn out, for instance, that the optimum value of some parameter is different for reading.

In either case a choice will arise as to whether to leave the variable variable and provide an extra knob to adjust it, or to fix it at some intermediate value and have a simpler and cheaper and worse optophone. One of the objects of the tests will be to settle these issues in a convincing way.

It should be expected that the first marketable optophone will be a completely different animal from the prototypes, at least as concerns packaging. The optophones should be on the tough side: they will be called on to operate in all weathers and not always be handled gently. Experienced people will be required here, to design the casing, power pack, knobs and so on. There are so many excellent electronic consumer items around now that there can be no excuse for amateurism. I even have the strong impression that there are firms that specialise in packaging people's prototypes in this way.

Plan to have the money in hand to have 50 of these first marketable optophones made. Sell them at a profit.

Solicit and listen to complaints and suggestions. Note in particular how these vary with the period of use.

REFERENCES

HARDWARE - BLIND AIDS - GENERAL

RW Mann (1965) "The evaluation and simulation of mobility aids for the blind", American Foundation for the Blind Research Bulletin, Oct, 93-98.

TD Sterling, EA Bering, SV Pollack & HG Vaughan (eds) (1971) "Visual prosthesis: the interdisciplinary dialogue", Academic Press, 382 pp.

JA Brabyn (1980) ""Artificial vision": a review of sensory aids for the blind", IEEE 1980 Frontiers of Engineering in Health Care, IEEE, 98-104.

JA Brabyn (1982) "New developments in mobility and orientation aids for the blind", IEEE Trans Biomed Eng, vol BME-29, n° 4, 285-289.

WJ Perkins (ed) (1983) "High technology aids for the disabled", Butterworths, 216 pp.

S Tachi, RW Mann & D Rowell (1983) "Quantitative comparison of alternative sensory displays for mobility aids for the blind", IEEE Trans Biomed Eng, vol BME-30, n° 9, 571-577.

JM Gill (1984) "Aids for the visually handicapped", Microprocessors & Microsystems, vol 8, n° 10, 517-519.

L Kay (1984) "Electronic aids for blind persons: an interdisciplinary subject", IEE Proc, vol 131, part A, n° 7, 559-576.

E Peli & T Peli (1984) "Image enhancement for the visually impaired", Optical Engineering, vol 23, n° 1, 47-51.

DH Warren & ER Strelow (eds) (1984) "Electronic spatial sensing for the blind", NATO Conference, Lake Arrowhead, 10-13 Sep 1984, Martinus Nijhoff, 521 pp.

RM Fish (1985) "Electronic aids for the blind", Radio Electronics, vol 56, 57-59.

JM Gill (1986) "International survey of aids for the visually disabled", Brunel University Research Unit for the Blind, 206 pp. Available from RNIB, London.

JM Gill (1986) "International register of research on visual disability", Brunel University Research Unit for the Blind, 72 pp. Available from RNIB, London.

G Moore (1986) "Literacy and computing for the blind", Electronics & Power, Jul, 513-516.

CM Scheff (1986) "Experimental model for the study of changes in the organisation of human sensory information processing through the design and testing of noninvasive prosthetic devices for sensory impaired people", SIGCAF Newsletter, vol 36, 3-10.

PL Emiliani (1989) "Concerted research programme on 'technology and blindness'", Computers for handicapped persons (Austria), R Oldenbourg (Vienna), 344-350.

N Ohnishi, Y Kawai & N Sugie (1989) "A support system for the blind to recognise a diagram", Images of the 21st century, Annual International IEEE Engineering in Medicine and Biology Society Conference, Seattle, 9-12 Nov 1989, IEEE, 1510-1511.

R Bucken (1990) "Aids for the handicapped", Funkschau (Germany), n° 10, 4 May, 39-40. In German.

HARDWARE - BLIND AIDS - ECHO TO AUDITORY

MD Altschuler, JL Potsdamer, G Frieder & MJ Manthey (1980) "A medium-range vision aid for the blind", Conference, Boston, 8-10 Oct 1980, IEEE, 1000-1002.

VV Lebedev & VL Scheiman (1980) "An assessment of the possibilities of building an echolocator for the blind", Telecom Radio Eng Part 2, vol 35, n° 3, Mar, 97-100.

BA Goldstein & WR Wiener (1981) "Acoustic analysis of the Sonic Guide", J Acoust Soc Am, vol 70, n° 2, 313-320.

Z Kucerowsky, MA Alford, E Brannen, TWW Stewart, SJ Lupker & WJ McClelland (1981) "Ultrasonic mobility aid for the blind", International Toronto, 5-7 Oct 1981, IEEE, 102-103.

AD Heyes (1983) "Human navigation by sound", Physics in Technology, vol 14, Mar, 68-75.

AG Dodds, DD Clark-Carter & CI Howarth (1984) "The Sonic Pathfinder: an evaluation", J Visual Impairment & Blindness, May, 203-206.

J Gosch (1984) "Acoustic aid helps blind to navigate", Electronics, vol 57, 22 Mar, 80.

VA Eliseev, VV Lebedev, VA Usik & EA Tishenko (1985) "A visual prosthesis for the blind", USSR patent n° SU 1168243 A. In Russian.

AD Heyes (1986) "Whatever happened to the Sonic Pathfinder?", Electronics & Wireless World, vol 93, Apr, 38.

BN Schenkman (1986) "Identification of ground materials with the aid of tapping sounds and vibrations of long canes for the blind", Ergonomics, vol 29, n° 8, 985-998.

Y Yonezawa (1986) "Binaural sensitivity to direction cue in an acoustic spatial sensor", Acustica, vol 61, n° 2, Aug, 140-144.

M Brissaud & G Grange (1988) "Systemes ultrasonores d'aide a la localisation pour non-voyants", Acustica, vol 66, 53-55.

T Ikufube, T Sasaki & C Peng (1991) "A blind mobility aid modeled after echolocation of bats", IEEE Trans Biomed Eng, vol BME-38, n° 5, May, 461-465.

HARDWARE - BLIND AIDS - NEUROSURGICAL IMPLANTS

AF Shackil (1980) "An electronic human eye", IEEE Spectrum, vol 17, n° 9, Sep, 88-91.

E Hitchcock (1982) "Development of a visual prosthesis", Applied Neurophysiol, vol 45, 25-31.

PEK Donaldson (1983) "Engineering visual prostheses", IEEE Eng in Med and Biol Mag, vol 2, n° 2, Jun, 14-18.

HARDWARE - BLIND AIDS - OPTICAL TO AUDITORY

RM Fish (1976) "An audio display for the blind", IEEE Trans Biomed Eng, vol BME-23, n° 2, Mar, 144-154.

SA Dallas Jr (1980) "Sound pattern generator", World Intellectual Property Organisation patent application n° WO 82/00395.

E Kurcz (1981) "Heliotrope: an optoelectronic aid for the blind", Polish Tech Rev, n° 4, 29-30.

MF Deering (1984) "Computer vision requirements in blind mobility aids", in DH Warren & ER Strelow (eds) Electronic spatial sensing for the blind, NATO Conference, Lake Arrowhead, 10-13 Sep 1984, Martinus Nijhoff, 521 pp, 65-82.

JT Tou & M Adjouadi (1984) "Computer vision for the blind", in DH Warren & ER Strelow (eds) Electronic spatial sensing for the blind, NATO Conference, Lake Arrowhead, 10-13 Sep 1984, Martinus Nijhoff, 521 pp, 83-124.

AR O'Hea (1987) "A general-purpose optical-to-auditory seeing aid for the blind: design requirements and computer simulations", MSc Dissertation, Brunel University Department of Computer Science, 105 pp.

F Furuno (1989) "Colour discriminating apparatus for the blind", Computers for handicapped persons (Austria), R Oldenbourg (Vienna), 134-143.

K Itoh & Y Yonezawa (1989) "Support systems for handwriting characters for the blind using feedback of sound imaging signals", Computers for handicapped persons (Austria), R Oldenbourg (Vienna), 325-330.

PBL Meijer (1989) "Image audio transformation system, particularly as a visual aid for the blind", European patent application n° 0 410 045 A1.

L Nielsen, MA Mahowald & C Mead (1989) "Seehear", in C Mead (ed) Analog VLSI and neural systems, Addison Wesley, 371 pp, 207-227.

PBL Meijer (1992) "An experimental system for auditory image representations", IEEE Trans Biomed Eng, vol BME-39, n° 2, Feb, 112-121.

HARDWARE - BLIND AIDS - OPTICAL TO TACTILE

N Bel (1980) "Integrated capacitive imaging display for the blind", in WA Kaiser & WE Proebster (eds) From Electronics to Microelectronics, Stuttgart, 24-28 Mar 1980, North Holland, 549-551.

T Pun & F de Coulon (1981) "Image processing for visual prosthesis", IEEE Computer Society Conference on Pattern Recognition and Image Processing, Dallas, 3-5 Aug 1981, 120-126.

E Corcoran (1985) "Whatever happened to tactile vision substitutions?", IEEE Spectrum, vol 22, n° 9, Sep, 20.

KA Kaczmarek, P Bach-y-Rita, WJ Tompkins & JG Webster (1985) "A tactile vision-substitution system for the blind: computer-controlled partial image sequencing", IEEE Trans Biomed Eng, vol BME-32, n° 8, Aug, 602-608.

KA Kaczmarek, JG Webster, P Bach-y-Rita & WJ Tompkins (1991) "Electrotactile and vibrotactile displays for sensory substitution systems", IEEE Trans Biomed Eng, vol BME-38, n° 1, Jan, 1-16.

HARDWARE - GENERAL

MH Jones (1977) "A practical introduction to electronic circuits", Cambridge University Press, 237 pp.

C Mead (1989) "Analog VLSI and neural systems", Addison Wesley, 371 pp.

HARDWARE - SIGNAL PROCESSING

GE Brebner & RT Ritchings (1988) "Image transform coding: a case study involving real time signal processing", IEE Proc, vol 135, part E, n° 1, Jan, 41-48.

EA Lee (1988) "Programmable DSP Architectures: part 1", IEEE ASSP Magazine, Oct, 4-14.

J Nolan (1989) "Working with DSP", Electronics & Wireless World, Jan, 52-55.

M Newell & J Rasure (1991) "A VLSI system for real-time linear operations and transforms", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 8, Aug, 1914-1917.

NR Shanbhag (1991) "An improved systolic architecture for 2-D digital filters", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 5, May, 1195-1202.

J Wilbur & FJ Taylor (1991) "Single-modulus RNS implementation of Wigner-Ville time-varying spectral estimations", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 4, Apr, 1020-1023.

AC Erickson & BS Fagin (1992) "Calculating the FHT in hardware", IEEE Trans Signal Processing, vol 40, n° 6, Jun, 1341-1353.

HARDWARE - SPEECH

RL Damper (1982) "Speech technology: implications for biomedical engineering", J Med Eng & Tech, vol 6, n° 4, Jul, 135-149.

MATHEMATICS - FRACTALS

N Dodd (1987) "Multispectral texture synthesis using fractal concepts", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 5, Sep, 703-707.

JM Keller, RM Crownover & RY Chen (1987) "Characteristics of natural scenes related to the fractal dimension", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 5, Sep, 621-627.

MF Barnsley (1988) "Fractals everywhere", Academic Press, 396 pp.

HO Peitgen & D Saupe (eds) (1988) "The science of fractal images", Springer Verlag, 312 pp.

G Zorpette (1988) "Fractals: not just another pretty picture", IEEE Spectum, Oct, 29-31.

DC Knill, D Field & D Kersten (1990) "Human discrimination of fractal images", J Opt Soc Am A, vol 7, n° 6, Jun, 1113-1123.

DS Mazel & MH Hayes (1992) "Using iterated function systems to model discrete sequences", IEEE Trans Signal Processing, vol 40, n° 7, Jul, 1724-1734.

MF Barnsley & LP Hurd (1993) "Fractal image compression", AK Peters, 244 pp.

G Vines & MH Hayes (1993) "Nonlinear address maps in a one-dimensional fractal model", IEEE Trans Signal Processing, vol 41, n° 4, Apr, 1721-1724.

MATHEMATICS - GENERAL

JJ Tuma (1979) "Engineering mathematics handbook (2nd edition)", McGraw Hill, 394 pp.

M Abramowitz & IA Stegun (eds) (1972) "Handbook of mathematical functions (9th printing)", Dover, 1046 pp.

MATHEMATICS - INFORMATION THEORY

G Raisbeck (1963) "Information theory: an introduction for scientists and engineers", Massachusetts Institute of Technology, 105 pp.

I Selin (1965) "Detection theory", Princeton University Press, 128 pp.

JF Young (1971) "Information theory", Butterworths, 168 pp.

AM Rosie (1973) "Information and communication theory", van Nostrand Reinhold, 221 pp.

J Campbell (1984) "Grammatical man: information, entropy, language and life", Pelican, 319 pp.

JH van Hateren (1992) "A theory of maximizing sensory information", Biol Cybern, vol 68, 23-29.

MATHEMATICS - NUMERICAL ANALYSIS

JJ Moré"The Levenberg-Marquart algorithm: implementation and theory".

MJD Powell (1977) "Numerical methods for fitting functions of two variables", in DAH Jacobs (ed) The state of the art in numerical analysis, Conference, York, 12-15 Apr 1976, Academic Press, 978 pp, 563-604.

RL Burden & JD Faires (1985) "Numerical analysis", Prindle Weber & Schmidt, 676 pp.

RE Bellman & RS Roth (1986) "Methods in approximation: techniques for mathematical modelling", D Reidel, 224 pp.

MATHEMATICS - MATRIX THEORY

A Graham (1979) "Matrix theory and applications for engineers and scientists", Ellis Horwood, 295 pp.

GH Golub & CF van Loan (1983) "Matrix computations", North Oxford Academic, 476 pp.

RA Horn & CA Johnson (1985) "Matrix analysis", Cambridge University Press, 561 pp.

H Anton (1987) "Elementary linear algebra (5th edition)", John Wiley & Sons, 529 pp.

MATHEMATICS - PRINCIPAL COMPONENT ANALYSIS

S Watanabe (1965) "Karhunen-Loève expansion and factor analysis: theoretical remarks and applications", 4th Conference on Information Theory, 1965, 635-660.

K Fukunaga & WLG Koontz (1970) "Application of the Karhunen-Loève expansion to feature selection and ordering", IEEE Trans Comput, vol C-19, n° 4, Apr, 311-318.

DN Lawley & AE Maxwell (1971) "Factor analysis and statistical method", Butterworths, 153 pp.

J Johnston (1972) "Econometric methods", McGraw Hill, 437 pp.

J Kittler & PC Young (1973) "A new approach to feature selection based on the Karhunen-Loève expansion", Pattern Recog, vol 5, 335-352.

HC Andrews & CL Patterson III (1976) "Singular value decomposition (SVD) image coding", IEEE Trans Commun, vol COM-1976, Apr, 425-432.

S Daultrey (1976) "Principal components analysis", University of East Anglia Geo Abstracts, 51 pp.

AK Jain (1976) "A fast Karhunen-Loève transform for a class of random processes", IEEE Trans Commun, vol COM-1976, Sep, 1023-1029.

AK Jain (1976) "Some new techniques in image processing", ONR Symposium on Current Problems in Image Science, Naval Postgraduate School Monterey California, 10-12 Nov 1976, 201-223.

P Sanyal & DH Foley (1976) "Feature selection by a modified Fukunaga-Koontz transform and its graphic interpretation", Proc MSAC 76, Symposium on Automatic Computation and Control, Milwaukee, 1976, 445-452.

AK Jain (1977) "Partial differential equations and finite-difference methods in image processing: Part 1 Image representation", J Optim Theory & Appl, vol 23, n° 1, Sep, 65-91.

K Ozeki (1979) "A coordinate-free theory of eigenvalue analysis related to the method of principal components and the Karhunen-Loève expansion", Inform & Control, vol 42, 38-59.

KVM Fernando & H Nicholson (1980) "Discrete double-sided Karhunen-Loève expansion", IEE Proc, vol 127, part D, n° 4, Jul, 155-160.

K Fukunaga & JM Mantock (1981) "Nonparametric feature extraction", Conference on Alternative Futures, Oct 1981, IEEE, 352-356.

J Karhunen & E Oja (1982) "New methods of stochastic approximation of truncated Karhunen-Loève expansions", 6th International Conference on Pattern Recognition, Munich, Oct 1982, IEEE, 550-553.

JP Keating, JE Michalek & JT Riley (1983) "A note on the optimality of the Karhunen-Loève expansion", Pattern Recogn Letters, vol 14, May, 203-204.

GW Queen, JK Bryan & JN Gowdy (1983) "Improved techniques for image data compression", 15th South-Eastern Symposium on System Theory, Huntsville, Mar 1983, IEEE, 18-22.

MH Savoji & RE Burge (1985) "On different methods based on the Karhunen-Loève expansion and used in image analysis", Comput Vis Graph & Image Process, vol 29, 259-269.

PM Farrelle & AK Jain (1986) "Recursive block coding: a new approach to transform coding", IEEE Trans Commun, vol COM-34, n° 2, Feb, 161-179.

JB Burl (1989) "Estimating the basis functions of the Karhunen-Loève transform", IEEE Trans Acoust Speech Signal Process, vol ASSP-37, n° 1, Jan, 99-105.

M Kirby & L Sirovich (1990) "Application of the Karhunen-Loève procedure the characterization of human faces", IEEE Trans Pattern Anal Machine Intel, vol PAMI-12, n° 1, Jan, 103-108.

M Turk & A Pentland (1991) "Eigenfaces for recognition", J Cognitive Neuroscience, vol 3, no 1, 71-86.

A Pentland, B Moghaddam, T Starner, O Oliyide & M Turk (1994) "View-based and modular eigenspaces for face recognition", MIT Media Laboratory Perceptual Computing Section Technical Report n° 245, MIT, 11 pp.

MATHEMATICS - SIGNAL PROCESSING - CEPSTRUM ANALYSIS

DG Childers, DP Skinner & RC Kemerait (1977) "The cepstrum: a guide to processing", Proc IEEE, vol 65, n° 10, Oct, 1428-1443.

DE Dudgeon (1977) "The computation of two-dimentional cepstra", IEEE Trans Acoust Speech Signal Process, vol ASSP-25, n° 6, Dec, 476-484.

T Kobayashi & S Imai (1984) "Spectral analysis using generalized cepstrum", IEEE Trans Signal Processing, vol ASSP-32, n° 6, Dec, 1235-1238.

MC Steckner & DJ Drost (1989) "Fast cepstrum analysis using the Hartley transform", IEEE Trans Acoust Speech Signal Process, vol ASSP-37, n° 8, Aug, 1300-1302.

RT Sokolov & JC Rogers (1993) "Time-domain cepstral transformations", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1161-1169.

I Yamada, K Sakaniwa & S Tsujii (1993) "A new multidimensional isomorphic operator and its properties", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1486. Article submitted for publication.

MATHEMATICS - SIGNAL PROCESSING - DATA COMPRESSION

TS Huang & OJ Tretiak (eds) (1972) "Picture bandwidth compression", Gordon & Breach, 734 pp.

AK Jain (1981) "Image data compression: a review", Proc IEEE, vol 69, n° 3, Mar, 349-389.

V Cappellini (ed) (1985) "Data compression and error control techniques with applications", Academic Press.

DE Pearson & JA Robinson (1985) "Visual communication at very low data rates", Proc IEEE, vol 73, n° 4, Apr, 795-812.

M Goldberg, PR Boucher & S Shlien (1986) "Image compression using adaptive vector quantization", IEEE Trans Commun, vol COM-34, n° 2, Feb, 180-187.

DE Pearson (1986) "Transmitting sign language for the deaf", Nat Electron Rev, 65-68.

JA Saghri & AG Tescher (1986) "Adaptive transform coding based on chain coding concepts", IEEE Trans Commun, vol COM-34, n° 2, Feb, 112-117.

G Zorpette (1988) "Fractals: not just another pretty picture", IEEE Spectum, Oct, 29-31.

G Karlsson & M Vetterli (1990) "Theory of two-dimensional multirate filter banks", IEEE Trans Acoust Speech Signal Process, vol ASSP-38, n° 6, Jun, 925-937.

MATHEMATICS - SIGNAL PROCESSING - GENERAL

HC Andrews (1970) "Computer techniques in image processing", Academic Press, 187 pp.

DV Widder (1971) "An introduction to transform theory", Academic Press, 253 pp.

WK Pratt (1978) "Digital image processing", John Wiley, 750 pp.

NT Kottegoda (1980) "Stochastic water resources technology", Macmillan, 384 pp.

RE Crochiere & LR Rabiner (1981) "Interpolation and decimation of digital signals - A tutorial review ", Proc IEEE, vol 69, n° 3, Mar, 300-331.

NB Jones (ed) (1982) "Digital signal processing", Peter Peregrinus, 490 pp.

DE Dudgeon & RM Mersereau (1984) "Multidimensional digital signal processing", Prentice Hall, 400 pp.

P Denyer & D Renshaw (1985) "VLSI signal processing: a bit-serial approach", Addison Wesley.

J Strawn (ed) (1985) "Digital audio signal processing: an anthology", William Kaufmann.

J Strawn (ed) (1985) "Digital audio engineering: an anthology", William Kaufmann, 144 pp.

FA Wilson (1985) "Audio (Elements of Electronics Book 6)", Bernard Babani, 308 pp.

RC Gonzalez & P Wintz (1987) "Digital image processing", Addison Wesley, 503 pp.

RA Roberts & CT Mullis (1987) "Digital signal processing", Addison Wesley, 578 pp.

JV Candy (1988) "Signal processing - the modern approach", McGraw Hill, 386 pp.

LB Jackson (1991) "Signals systems and transforms", Addison Wesley.

MATHEMATICS - SIGNAL PROCESSING - HALFTONING

T Scheermesser, M Broja & O Bryngdahl (1993) "Adaptation of spectral contraint to electronically halftoned pictures", J Opt Soc Am A, vol 10, n° 3, Mar, 412-417.

MATHEMATICS - SIGNAL PROCESSING - HIDDEN MARKOV MODELS

LR Rabiner & BH Juang (1986) "An introduction to hidden Markov models", IEEE ASSP Magazine, Jan, 4-16.

MATHEMATICS - SIGNAL PROCESSING - INTERFRAME IMAGE CODING

JA Roese, WK Pratt & GS Robinson (1977) "Interframe cosine transform image coding", IEEE Trans Commun, vol COM-25, n° 11, Nov, 1329-1339.

MATHEMATICS - SIGNAL PROCESSING - IMAGE RESTORATION

WK Pratt (1972) "Generalised Wiener filtering techniques", IEEE Trans Comput, vol C-21, n° 7, 636-641.

BR Hunt (1973) "The application of constrained least squares estimation to image restoration by digital computer", IEEE Trans Comput, vol C-22, n° 9, 805-812.

BR Hunt (1975) "Digital image processing", Proc IEEE, vol 63, n° 4, 693-708.

AK Jain (1977) "A fast Karhunen-Loève transform for digital restoration of images degraded by white and coloured noise", IEEE Trans Comput, vol C-26, n° 6, 560-571.

WK Pratt & F Davarian (1977) "Fast computational techniques for pseudoinverse and Wiener image restoration", IEEE Trans Comput, vol C-26, n° 6, 571-580.

AK Jain & JR Jain (1978) "Partial differential equations and finite difference methods in image processing: Part II Image restoration", IEEE Trans Automat Control, vol AC-23, n° 5, Oct, 817-834.

AM Tekalp & G Pavlovic (1991) "Image restoration with multiplicative noise: incorporating sensor nonlinearity", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 9, Sep, 2132-2136.

MATHEMATICS - SIGNAL PROCESSING - INVARIANCE

RA Altes (1978) "The Fourier-Mellin transform and mammalian hearing", J Acoust Soc Am, vol 63, n° 1, 174-183.

JJ Koenderink & AJ van Doorn (1982) "Invariant features of contrast detection: an explanation in terms of self-similar detector arrays", J Opt Soc Am, vol 72, n° 1, Jan, 83-87.

DH Foster & JI Kahn (1985) "Internal representations and operations in the visual comparison of transformed patterns: effects of pattern point-inversion, positional symmetry and separation", Biol Cybern, vol 51, 305-312.

RA Messner & HH Szu (1985) "An image processing architecture for real-time generation of scale and rotation invariant patterns", Comput Graphics Image Process, vol 31, 50-66.

C Braccini & G Gambardella (1986) "Form-invariant linear filtering: theory and applications", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 6, 1612-1628.

JI Kahn & DH Foster (1986) "Horizontal-vertical structure in the visual comparison of rigidly transformed patterns", J Exper Psychol: Human Percept & Perform, vol 12, n° 4, 422-433.

E de Castro & C Morandi (1987) "Registration of translated and rotated images using finite Fourier transforms", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 5, Sep, 700-703.

H Wechsler (1987) "Invariance in pattern recognition", in PW Hawkes (ed) Advances in Electronics and Electron Physics, vol 69, Academic Press, 261-322.

JE Cutting (1988) "Affine distortions of pictorial space: some predictions for Goldstein (1987) that La Gournerie (1859) might have made", J Exper Psychol: Human Percept & Perform, vol 14, n° 2, 305-311.

J Segman, J Rubinstein & YY Zeevi (1992) "The canonical coordinates method for pattern deformation: theoretical and computational considerations", IEEE Trans Pattern Anal Machine Intel, vol 14, n° 12, Dec, 1171-1183.

EP Simoncelli, WT Freeman, EH Adelson & DJ Heeger (1992) "Shiftable multiscale transforms", IEEE Trans Information Theory, vol 38, n° 2, Mar, 587-607.

MATHEMATICS - SIGNAL PROCESSING - LINEAR PREDICTION

J Makhoul (1975) "Linear prediction: a tutorial review", Proc IEEE, vol 63, n° 4, 561-580.

HW Strube (1980) "Linear prediction on a warped frequency scale", J Acoust Soc Am, vol 68, n° 4, 1071-1076.

MR Schroeder & BS Atal (1985) "Code-excited linear prediction (CELP): high-quality speech at very low bit rates", Proc IEEE Int Conf Acoust Speech Signal Proc, 937-940.

E Kr}ger & HW Strube (1988) "Linear prediction on a warped frequency scale", IEEE Trans Acoust Speech Signal Process, vol ASSP-36, n° 9, 1529-1531.

MATHEMATICS - SIGNAL PROCESSING - MAXIMUM ENTROPY

A Papoulis (1981) "Maximum entropy and spectral estimation: a review", IEEE Trans Acoust Speech Signal Process, vol ASSP-29, n° 6, Dec, 1176-1186.

MATHEMATICS - SIGNAL PROCESSING - NEURAL NETWORKS

JJ Hopfield (1988) "Artificial neural networks", IEEE Circuit Device Mag, Sep, 3-10.

J Xing & GL Gerstein (1993) "A neural network model for texture discrimination", Biol Cybern, vol 69, 97-108.

MATHEMATICS - SIGNAL PROCESSING - PATTERN RECOGNITION

HC Andrews (1972) "Introduction to mathematical techniques in pattern recognition", Wiley Interscience, 242 pp.

S Watanabe (ed) (1972) "Frontiers of pattern recognition", Academic Press, 602 pp.

JR Ullmann (1973) "Pattern recognition techniques", Butterworths, 412 pp.

S Ullman & W Richards (eds) (1984) "Image understanding", Ablex, 268 pp.

MATHEMATICS - SIGNAL PROCESSING - PHASE AND MAGNITUDE

MH Hayes, JS Lim & AV Oppenheim (1980) "Signal reconstruction from phase or magnitude", IEEE Trans Acoust Speech Signal Process, vol ASSP-28, n° 6, 672-680.

AV Oppenheim & JS Lim (1981) "The importance of phase in signals", Proc IEEE, vol 69, n° 5, 529-541.

MH Hayes (1982) "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform", IEEE Trans Acoust Speech Signal Process, vol ASSP-30, n° 2, 140-154.

SH Nawab, TF Quatieri & JS Lim (1983) "Signal reconstruction from short-time Fourier transform magnitude", IEEE Trans Acoust Speech Signal Process, vol ASSP-31, n° 4, 986-998.

B Yegnanarayana, DK Saikia & TR Krishnan (1984) "Significance of group delay functions in signal reconstruction from spectral phase or magnitude", IEEE Trans Acoust Speech Signal Process, vol ASSP-32, n° 3, 610-622.

S Shitz & YY Zeevi (1985) "On the duality of time and frequency domain signal reconstruction from partial information", IEEE Trans Acoust Speech Signal Process, vol ASSP-33, n° 6, 1486-1498.

A van den Bos (1987) "A new method for synthesis of low-peak-factor signals", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, n° 1, Jan, 120-122.

B Yegnanarayana & A Raghunathan (1987) "Representation of images through group-delay functions", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, n° 2, 237-240.

J Le Roux, P Sole, A Murat Tekalp & A Tanju Erdem (1993) "Tekalp-Erdem estimator gives the LS estimate for Fourier and log-Fourier modulus", IEEE Trans Signal Processing, vol 41, n° 4, Apr, 1705-1707.

J Weng (1993) "Windowed Fourier phase: completeness and signal reconstruction", IEEE Trans Signal Processing, vol 41, n° 2, Feb, 657-666.

MATHEMATICS - SIGNAL PROCESSING - SPACE-FREQUENCY ANALYSIS (see also: PSYCHOPHYSICS - SIGHT - 2D VISION - General)

G Wackersreuther (1986) "On two-dimensional polyphase filter banks", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 1, Feb, 192-199.

AC Bovik, N Gopal, T Emmoth & A Restrepo (1992) "Localized measurement of emergent image frequencies by Gabor wavelets", IEEE Trans Information Theory, vol 38, n° 2, Mar, 691-712.

K Grochenig & WR Madych (1992) "Multiresolution analysis, Haar bases, and self-similar tilings of R^n", IEEE Trans Information Theory, vol 38, n° 2, Mar, 556-568.

SL Hahn (1992) "Multidimensional complex signals with single-orthant spectra", Proc IEEE, vol 80, n° 8, Aug, 1287-1300.

J Kovačevic & M Vetterli (1992) "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for R^n", IEEE Trans Information Theory, vol 38, n° 2, Mar, 533-555.

GL Sicuranza (1992) "Quadratic filters for signal processing", Proc IEEE, vol 80, n° 8, Aug, 1263-1285.

R Wilson, AD Calway & ERS Pearson (1992) "A generalized wavelet transform for Fourier analysis: the multiresolution Fourier transform and its application to image and audio signal analysis", IEEE Trans Information Theory, vol 38, n° 2, Mar, 674-690.

JI Yellott Jr & GJ Iverson (1992) "Uniqueness properties of higher-order autocorrelation functions", J Opt Soc Am A, vol 9, n° 3, Mar, 388-404.

AN Akansu, RA Haddad & H Caglar (1993) "The binomial QMF-wavelet transform for multiresolution signal decomposition", IEEE Trans Signal Processing, vol 41, n° 1, Jan, 13-19.

YY Zeevi & E Shlomot (1993) "Nonuniform sampling and antialiasing in image representation", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1223-1236.

MATHEMATICS - SIGNAL PROCESSING - TIME-FREQUENCY ANALYSIS

WD Mark (1970) "Spectral analysis of the convolution and filtering of nonstationary stochastic processes", J Sound Vib, vol 11, n° 1, 19-63.

RW Schafer & LR Rabiner (1973) "Design and simulation of a speech analysis-synthesis system based on short-time Fourier analysis", IEEE Trans Audio Elecroacoust, vol AU-21, n° 3, 165-174.

JA Moorer (1976) "The synthesis of complex audio spectra by means of discrete summation formulas", J Audio Eng Soc, vol 24, n° 9, 717-727.

MR Portnoff (1976) "Implementation of the digital phase vocoder using the fast Fourier transform", IEEE Trans Acoust Speech Signal Process, vol ASSP-24, n° 3, 243-248.

JB Allen (1977) "Short-term spectral analysis, synthesis, and modification by discrete Fourier transform", IEEE Trans Acoust Speech Signal Process, vol ASSP-25, n° 3, 235-238.

JA Moorer (1978) "The use of the phase vocoder in computer music applications", J Audio Eng Soc, vol 26, 42-45.

JA Moorer (1979) "The digital coding of high-quality musical sound", J Audio Eng Soc, vol 27, n° 9, 657-666.

RE Crochiere (1980) "A weighted overlap-add method of short-time Fourier analysis/synthesis", IEEE Trans Acoust Speech Signal Process, vol ASSP-28, n° 1, 99-102.

JN Holmes (1980) "The JSRU channel vocoder", IEE Proc, vol 127, part F, n° 1, Feb, 53-60.

MR Portnoff (1980) "Time-frequency representation of digital signals and systems based on short-time Fourier analysis", IEEE Trans Acoust Speech Signal Process, vol ASSP-28, n° 1, 55-69.

SM Kay & SL Marple Jr (1981) "Spectrum analysis - a modern perspective", Proc IEEE, vol 69, n° 11, Nov, 1380-1419.

MR Portnoff (1981) "Time-scale modification of speech based on short-time Fourier analysis", IEEE Trans Acoust Speech Signal Process, vol ASSP-29, n° 3, 374-390.

MR Portnoff (1981) "Short-time Fourier analysis of sampled speech", IEEE Trans Acoust Speech Signal Process, vol ASSP-29, n° 3, 364-373.

JA Stuller (1982) "Generalised running discrete transforms", IEEE Trans Acoust Speech Signal Process, vol ASSP-30, n° 1, 60-68.

TACM Claasen & WFG Mecklenbr{uker (1983) "The aliasing problem in discrete-time Wigner distributions", IEEE Trans Acoust Speech Signal Process, vol ASSP-31, n° 5, 1067-1072.

JM Kates (1983) "An auditory spectal analysis model using the chirp z-transform", IEEE Trans Acoust Speech Signal Process, vol ASSP-31, n° 1, 148-156.

TL Petersen & SF Boll (1983) "Critical band synthesis-analysis", IEEE Trans Acoust Speech Signal Process, vol ASSP-31, n° 3, 656-663.

E Paulus (1984) "A fast convolution procedure for discrete short-time spectral analysis with frequency-dependent resolution", IEEE Trans Acoust Speech Signal Process, vol ASSP-32, n° 5, 1100-1104.

T Dabelsteen & SB Pedersen (1985) "A method for computerised modification of certain natural animal sounds for communication study purposes", Biol Cybern, vol 52, 399-404.

DJ Hermes (1985) "Separation of time and frequency", Biol Cybern, vol 52, 109-115.

D Izraelevitz (1985) "Some results on the time-frequency sampling of the short-time Fourier transform magnitude", IEEE Trans Acoust Speech Signal Process, vol ASSP-33, n° 6, 1611-1613.

AJEM Janssen & TACM Claasen (1985) "On positivity of time-frequency distributions", IEEE Trans Acoust Speech Signal Process, vol ASSP-33, n° 4, 1029-1032.

W Martin & P Flandrin (1985) "Wigner-Ville spectral analysis of nonstationary processes", IEEE Trans Acoust Speech Signal Process, vol ASSP-33, n° 6, 1461-1470.

VJ Mathews & DH Youn (1985) "Analysis of the short-time unbiased spectrum estimation algorithm", IEEE Trans Acoust Speech Signal Process, vol ASSP-33, n° 1, 136-142.

BEA Saleh & NS Subotic (1985) "Time-variant filtering of signals in the mixed time-frequency domain", IEEE Trans Acoust Speech Signal Process, vol ASSP-33, n° 6, 1479-1485.

DH Youn & JG Kim (1985) "Short-time Fourier transform using a bank of low-pass filters", IEEE Trans Acoust Speech Signal Process, vol ASSP-33, n° 1, 182-185.

GF Boudreaux-Bartels & TW Parks (1986) "Time-varying filtering and signal estimation using Wigner distribution systhesis techniques", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 3, 442-451.

A Dembo & D Malah (1986) "WMMSE design of digital filter banks with specified composite response", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 6, Dec, 1529-1541.

RJ McAulay & TF Quatieri (1986) "Speech analysis/synthesis based on a sinusoidal representation", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 4, 744-754.

F Peyrin & R Prost (1986) "A unified definition for the discrete-time, discrete-frequency, and discrete-time/frequency Wigner distributions", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 4, 858-867.

JP Princen & A Bernard Bradley (1986) "Analysis/synthesis filter bank design based on time domain aliasing cancellation", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 5, Oct, 1153-1161.

TF Quatieri & RJ McAulay (1986) "Speech transformations based on a sinusoidal representation", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 6, 1449-1464.

G Wackersreuther (1986) "Some new aspects of filters for filter banks", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 5, Oct, 1182-1200.

B Boashash & PJ Black (1987) "An efficient real-time implementation of the Wigner-Ville distribution", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, n° 11, 1611-1618.

L Cohen (1987) "Wigner distribution for finite-duration or band-limited signals and limiting cases", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, n° 6, 796-806.

L Cohen (1987) "On a fundamental property of the Wigner distribution", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, nº 4, 559-561.

E de Boer & WA Dreschler (1987) "Auditory psychophysics: spectrotemporal representation of signals", Ann Rev Psychol, vol 38, 181-202.

A Dembo & D Malah (1987) "The design of optimal uniform filter banks with specified composite response", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, nº 6, Jun, 807-817.

AJEM Janssen (1987) "A note on "Positive time-frequency distributions"", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, nº 5, 701-705.

MJT Smith & TP Barnwell (1987) "A new filter bank theory for time-frequency representation", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, nº 3, Mar, 314-327.

PP Vaidyanathan (1987) "Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques", IEEE ASSP Magazine, Jul, 4-20.

M Vetterli (1987) "A theory of multirate filter banks", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, nº 3, Mar, 356-372.

KB Yu & SL Cheng (1987) "Signal synthesis from pseudo-Wigner distribution and applications", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, nº 9, 868-873.

B Boashash (1988) "Note on the use of the Wigner distribution for time-frequency signal analysis", IEEE Trans Acoust Speech Signal Process, vol ASSP-36, nº 9, 1518-1521.

TP Bronez (1988) "Spectral estimation of irregularly sampled multidimensional processes by generalised prolate spheroidal sequences", IEEE Trans Acoust Speech Signal Process, vol ASSP-36, nº 12, 1862-1873.

A Dembo & D Malah (1988) "Signal systhesis from modified discrete short-time transform", IEEE Trans Acoust Speech Signal Process, vol ASSP-36, nº 2, 168-181.

F Hlawatsch (1988) "A note on "Wigner distribution for finite-duration or band-limited signals and limiting cases"", IEEE Trans Acoust Speech Signal Process, vol ASSP-36, nº 6, 927-929.

MS Murphy & FJ Owens (1988) "Optimum design of running Fourier transform filter banks", IEE Digest, vol 1988, nº 11, IEE Electronics Division Professional Group E10 Colloquium on Speech Processing, London, 19-19 Jan 1988, IEE.

TQ Nguyen & pp Vaidyanathan (1988) "Maximally decimated perfect-recontruction FIR filter banks with pairwise mirror-image analysis (and synthesis) frequency responses", IEEE Trans Acoust Speech Signal Process, vol 36, n° 5, May, 693-706.

FJ Owens & MS Murphy (1988) "A short-time Fourier transform", Signal Processing, vol 14, 3-10.

SC Pei & TY Wang (1988) "The Wigner distribution of linear time-variant systems", IEEE Trans Acoust Speech Signal Process, vol ASSP-36, n° 10, 1681-1684.

M Vetterli (1988) "Running FIR and IIR filtering using multirate filter banks", IEEE Trans Acoust Speech Signal Process, vol 36, n° 5, May, 730-738.

LB White & B Boashash (1988) "On estimating the instantaneous frequency of a Gaussian random signal by use of the Wigner-Ville distribution", IEEE Trans Acoust Speech Signal Process, vol ASSP-36, n° 3, 417-420.

M Sun, CC Li, LN Sekhar & RJ Sclabassi (1989) "Efficient computation of the discrete pseudo-Wigner distribution", IEEE Trans Acoust Speech Signal Process, vol ASSP-37, n° 11, Nov, 1735-1742.

RA Altes (1990) "Wide-band, proportional-bandwidth Wigner-Ville analysis", IEEE Trans Acoust Speech Signal Process, vol ASSP-38, n° 6, Jun, 1005-1012.

RD Hippenstiel & PM de Oliveira (1990) "Time-varying spectral estimation using the instantaneous power spectrum", IEEE Trans Acoust Speech Signal Process, vol ASSP-38, n° 10, Oct, 1752-1759.

DL Jones & TW Parks (1990) "A high resolution data-adaptive time-frequency representation", IEEE Trans Acoust Speech Signal Process, vol ASSP-38, n° 12, Dec, 2127-2135.

Y Zhao, LE Atlas & RJ Marks (1990) "The use of cone-shaped kernels for generalised time-frequency representations of nonstationary signals", IEEE Trans Acoust Speech Signal Process, vol ASSP-38, n° 7, Jul, 1084-1091.

MR Dellomo & GM Jacyna (1991) "Wigner transforms, Gabor coefficients, and Weyl-Heisenberg wavelets", J Acoust Soc Am, vol 89, n° 5, May, 2355-2361.

C Griffin, P Rao & F Taylor (1991) "Roundoff error analysis of the discrete Wigner distribution using fixed-point arithmetic", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 9, Sep, 2096-2098.

B Harms (1991) "Computing time-frequency distributions", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 3, Mar, 727-729.

F Hlawatsch (1991) "Duality and classification of bilinear time-frequency signal representations", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 7, Jul, 1564-1574.

F Hlawatsch & W Krattenthaler (1991) "Phase matching algorithms for Wigner-distribution signal synthesis", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 3, Mar, 612-619.

W Krattenthaler & F Hlawatsch (1991) "Improved signal synthesis from pseudo-Wigner distribution", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 2, Feb, 506-509.

P Rao & FJ Taylor (1991) "Detection and localisation of narrow-band transient signals using the Wigner distribution", J Acoust Soc Am, vol 90, n° 3, Sep, 1423-1434.

O Rioul & M Vetterli (1991) "Wavelets and signal processing", IEEE SP magazine, Oct, 14-38.

B Boashash (1992) "Estimating and interpreting the instantaneous frequency of a signal - Part 1 Fundamentals - Part 2 Algorithms and applications", Proc IEEE, vol 80, n° 4, Apr, 519-568.

I Daubechies, S Mallat & AS Willsky (1992) "Introduction to the special issue on wavelet transforms and multiresolution signal analysis", IEEE Trans Information Theory, vol 38, n° 2, Mar, 529-531.

N Delprat, B Escudie, P Guillemain, R Kronland-Martinet, P Tchamitchian & B Torresani (1992) "Asymptotic wavelet and Gabor analysis: extraction of instantaneous frequencies", IEEE Trans Information Theory, vol 38, n° 2, Mar, 644-664.

L Haken (1992) "Computational methods for real-time Fourier synthesis", IEEE Trans Signal Processing, vol 40, n° 9, Sep, 2327-2329.

F Hlawatsch & W Krattenthaler (1992) "Bilinear signal synthesis", IEEE Trans Signal Processing, vol 40, n° 2, Feb, 352-363.

F Hlawatsch & GF Boudreaux-Bartels (1992) "Linear and quadratic time-frequency signal representations", IEEE ASSP Magazine, Apr, 21-67.

S Kadambe & G Faye Boudreaux-Bartels (1992) "A comparison of the existence of "cross terms" in the Wigner distribution and the squared magnitude of the wavelet transform and the short time Fourier transform", IEEE Trans Signal Processing, vol 40, n° 10, Oct, 2498-2517.

A Kumar, DR Fuhrmann, M Frazier & BD Jawerth (1992) "A new transform for time-frequency analysis", IEEE Trans Signal Processing, vol 40, n° 7, Jul, 1697-1707.

DS Mazel & MH Hayes (1992) "Using iterated function systems to model discrete sequences", IEEE Trans Signal Processing, vol 40, n° 7, Jul, 1724-1734.

K Nayebi, TP Barnwell & MJT Smith (1992) "Time-domain filter bank analysis: a new design theory", IEEE Trans Signal Processing, vol 40, n° 6, Jun, 1412-1429.

S Oh & RJ Marks (1992) "Some properties of the generalized time frequency representation with cone-shaped kernel", IEEE Trans Signal Processing, vol 40, n° 7, Jul, 1735-1745.

S Pei & I Yang (1992) "Computing pseudo-Wigner distribution by the fast Hartley transform", IEEE Trans Signal Processing, vol 40, n° 9, Sep, 2346-2349.

O Rioul & P Duhamel (1992) "Fast algorithms for discrete and continuous wavelet transforms", IEEE Trans Information Theory, vol 38, n° 2, Mar, 569-586.

MJ Shensa (1992) "The discrete wavelet transform: wedding the à trous and Mallat algorithms", IEEE Trans Signal Processing, vol 40, n° 10, Oct, 2464-2482.

M Vetterli & C Herley (1992) "Wavelets and filter banks: theory and design", IEEE Trans Signal Processing, vol 40, n° 9, Sep, 2207-2232.

J Wexler & S Raz (1992) "Wigner-space synthesis of discrete-time periodic signals", IEEE Trans Signal Processing, vol 40, n° 8, Aug, 1997-2006.

R Wilson, AD Calway & ERS Pearson (1992) "A generalized wavelet transform for Fourier analysis: the multiresolution Fourier transform and its application to image and audio signal analysis", IEEE Trans Information Theory, vol 38, n° 2, Mar, 674-690.

MG Amin (1993) "Introducing the spectral diversity", IEEE Trans Signal Processing, vol 41, n° 1, Jan, 185-193.

RG Baraniuk & DL Jones (1993) "A signal-dependent time-frequency representation: optimal kernal design", IEEE Trans Signal Processing, vol 41, n° 4, Apr, 1589-1601.

S Barash & Y Ritov (1993) "Logarithmic pruning of FFT frequencies", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1398-1400.

MU Bikdash & K Yu (1993) "Analysis and filtering using the optimally smoothed Wigner distribution", IEEE Trans Signal Processing, vol 41, n° 4, Apr, 1603-1617.

B Boashash & P O'Shea (1993) "Use of the cross Wigner-Ville distribution for estimation of instantaneous frequency", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1439-1445.

W Krattenhaler & F Hlawatsch (1993) "Time-frequency design and processing of signals via smoothed Wigner distributions", IEEE Trans Signal Processing, vol 41, n° 1, Jan, 278-287.

PJ Loughlin, JW Pitton & LE Atlas (1993) "Bilinear time-frequency representations: new insights and properties", IEEE Trans Signal Processing, vol 41, n° 2, Feb, 750-767.

AK Soman & pp Vaidyanathan (1993) "On orthonormal wavelets and paraunitary filter banks", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1170-1183.

G Vines & MH Hayes (1993) "Nonlinear address maps in a one-dimensional fractal model", IEEE Trans Signal Processing, vol 41, n° 4, Apr, 1721-1724.

MATHEMATICS - SIGNAL PROCESSING - TRANSFORMS

RVL Hartley (1942) "A more symmetrical Fourier analysis applied to transmission problems", Proc IRE, Mar, 144-150.

N Ahmed, T Natarajan & KR Rao (1974) "Discrete cosine transform", IEEE Trans Comput, vol C-1974, Jan, 90-93.

M Hamidi & J Pearl (1976) "Comparison of the cosine and Fourier transforms of Markov-1 signals", IEEE Trans Acoust Speech Signal Process, vol ASSP-1976, Oct, 428-429.

AK Jain (1979) "A sinusoidal family of unitary transforms", IEEE Trans Pattern Anal Machine Intel, vol PAMI-1, n° 4, Oct, 356-365.

RN Bracewell (1983) "Discrete Hartley transform", J Opt Soc Am, vol 73, n° 12, Dec, 1832-1835.

RN Bracewell (1984) "The fast Hartley transform", Proc IEEE, vol 72, n° 8, Aug, 1010-1018.

BG Lee (1984) "A new algorithm to compute the discrete cosine transform", IEEE Trans Acoust Speech Signal Process, vol ASSP-32, n° 6, Dec, 1243-1245.

RN Bracewell (1986) "The Fourier transform and its applications - 2nd edition", McGraw Hill International Editions, 474 pp.

RN Bracewell (1986) "The Hartley transform", Oxford University Press, 160 pp.

KG Beauchamp (1987) "Tranforms for engineers", Oxford Science Publications Clarendon Press.

DA Jaffe (1987) "Spectrum analysis tutorial - Part 1: The discrete Fourier transform", Computer music journal, vol 11, n° 2, Sum, 9-24.

DA Jaffe (1987) "Spectrum analysis tutorial - Part 2: Properties and applications of the discrete Fourier transform", Computer music journal, vol 11, n° 3, Fal, 17-35.

AO Steinhardt (1988) "Householder transforms in signal processing ", IEEE ASSP Magazine, Jul, 4-12.

AC Erickson & BS Fagin (1992) "Calculating the FHT in hardware", IEEE Trans Signal Processing, vol 40, n° 6, Jun, 1341-1353.

JC Ehrhardt (1993) "Hexagonal fast Fourier transform with rectangular output", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1469-1472.

HV Sorensen & CS Burrus (1993) "Efficient computation of the DFT with only a subset of input or output points", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1184-1199.

MATHEMATICS - SIGNAL PROCESSING - WINDOWS

FJ Harris (1978) "On the use of windows for harmonic analysis with the discrete Fourier transform", Proc IEEE, vol 66, n° 1, 51-83.

AH Nuttall (1981) "Some windows with very good sidelobe behaviour", IEEE Trans Acoust Speech Signal Process, vol ASSP-29, n° 1, 84-91.

CA Greenhall (1990) "Orthogonal sets of data windows constructed from trigonometric polynomials", IEEE Trans Acoust Speech Signal Process, vol ASSP-38, n° 5, May, 870-872.

JW Adams (1991) "A new optimal window", IEEE Trans Acoust Speech Signal Process, vol ASSP-39, n° 8, Aug, 1753-1769.

MATHEMATICS - STATISTICS

CG Paradine & BHP Rivett (1964) "Statistical methods for technologists", English Universities Press, 288 pp.

TW Anderson (1984) "An introduction to multivariate statistical analysis", John Wiley, 675 pp.

A Papoulis (1984) "Probability, random variables and stochastic processes", McGraw Hill International Editions, 576 pp.

PSYCHOPHYSICS - BLINDNESS - GENERAL

M von Senden (1960) "Space and sight: the perception of space and shape in the congenitally blind before and after operation", Methuen, 348 pp.

K Trouern-Trend & EA Bering Jr (1969) "Blindness in the United States", 2nd University of Chicago Conference on Visual prosthesis, Chicago, 2-4 Jun 1969, 371-378.

MA Heller & JM Kennedy (1990) "Perspective taking, pictures, and the blind", Percept & Psychophys, vol 48, n° 5, 459-466.

PSYCHOPHYSICS - CROSSMODAL STUDIES

RH Henneman & ER Long (1954) "A comparison of the visual and auditory senses as channels for data presentation (technical report n° 54-363)", Wright Air Development Centre, 38 pp.

GH Mowbray & JW Gebhard (1958) "Man's senses as information channels", John Hopkins University Applied Physics Laboratory, 64 pp.

SS Stevens (1958) "Some similarities between hearing and seeing", Laryngoscope, vol 68, International Conference on Audiology, 14 May 1957, 508-527.

R Jakobson (1964) "On visual and auditory signs", Phonetica, vol 11, 216-220.

B Julesz (1971) "Critical bands in vision and audition", 7th International Congress on Acoustics, Budapest, Akademiai Kiado (Budapest), 445-448.

B Julesz (1971) "Foundations of cyclopean perception", University of Chicago Press, 406 pp, 50-53 & 66-74.

EE David & PB Denes (eds) (1972) "Human communication: a unified view", McGraw Hill.

B Julesz & I Hirsh (1972) "Visual and auditory perception: an essay of comparison", in EE David & PB Denes (eds) Human communication: a unified view, McGraw Hill.

LE Marks (1978) "The unity of the senses", Academic Press, 289 pp.

W Slawson (1985) "Sound color", University of California Press, 266 pp.

LE Marks, R Szczesiul & P Ohlott (1986) "On the cross-modal perception of intensity", J Exper Psychol: Human Percept & Perform, vol 12, n° 4, 517-534.

LE Marks (1987) "On crossmodal similarity: auditory-visual interactions in speeded discrimination", J Exper Psychol: Human Percept & Perform, vol 13, n° 3, 384-394.

JW Grau & DG Kemler Nelson (1988) "The distinction between integral and separable dimensions: evidence for the integrality of pitch and loudness", J Exper Psychol: General, vol 117, n° 4, 347-370.

S Handel (1988) "Space is to time as vision is to audition: seductive but misleading", J Exper Psychol: Human Percept & Perform, vol 14, n° 2, 315-317.

LE Marks (1989) "On crossmodal similarity: the perceptual structure of pitch, loudness and brightness", J Exper Psychol: Human Percept & Perform, vol 15, n° 3, 586-602.

RD Melara (1989) "Dimensional interaction between color and pitch", J Exper Psychol: Human Percept & Perform, vol 15, n° 1, 69-79.

RD Melara (1989) "Similarity relations among synesthetic stimuli and their attributes", J Exper Psychol: Human Percept & Perform, vol 15, n° 2, 212-231.

RD Melara & LE Marks (1990) "Interaction among auditory dimensions: timbre, pitch, and loudness", Percept & Psychophys, vol 48, n° 2, 169-178.

LM Ward (1990) "Mixed-method mixed-modality psychophysical scaling", Percept & Psychophys, vol 48, n° 6, 571-582.

BM Bennett, DD Hoffman & C Prakash (1991) "Unity of perception", Cognition, vol 38, 295-334.

I Liu, Y Zhu & J Wu (1992) "The long-term modality effect: in search of differences in processing logographs and alphabetic words", Cognition, vol 43, 31-66.

X Seron, M Pesenti, M Noel, G Deloche & J Cornet (1992) "Images of numbers or "when 98 is upper left and 6 sky blue"", Cognition, vol 44, 159-196.

SA Shamma (1992) "Hearing as seeing - Space and time in auditory processing", in FH Eeckman (ed) Analysis and modelling of neural systems, Kluwer Academic Publishers, 253-274.

DR Perrott, B Costantino & J Cisneros (1993) "Auditory and visual localization performance in a sequential discrimination task", J Acoust Soc Am, vol 93, n° 1, Apr, 2134-2138.

RE Cytowic (1994) "The man who tasted shapes", Little Brown. Reported in the Independent on Sunday, 13 Feb 1994.

PSYCHOPHYSICS - GENERAL

CW Savage (1970) "The measurement of sensation: a critique of perceptual psychophysics", University of California Press, 578 pp.

DH Krantz, RC Atkinson, RD Luce & P Suppes (eds) (1974) "Contemporary developments in mathematical psychology vol II: measurement, psychophysics and neural information processing", WH Freeman, 468 pp.

JH van Hateren (1992) "A theory of maximizing sensory information", Biol Cybern, vol 68, 23-29.

PSYCHOPHYSICS - HEARING - ATTENTION

MR Leek, ME Brown & MF Dorman (1991) "Informational masking and auditory attention", Percept & Psychophys, vol 50, n° 3, 205-214.

PSYCHOPHYSICS - HEARING - AUDITION-BASED SOUND REPRESENTATION

H Hermansky, BA Hanson & H Wakita (1985) "Perceptually based linear predictive analysis of speech", Proc ICASSP 85, 509-512.

H Hermansky, K Tsuga, S Makino & H Wakita (1986) "Perceptually based processing in automatic speech recognition", Proc ICASSP 86, Tokyo, 1971-1974.

O Ghitza (1987) "Auditory nerve representation criteria for speech analysis/synthesis", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, n° 6, 736-740.

W Heinbach (1988) "Aurally adequate signal representation: the part-tone-time-pattern", Acustica, vol 67, 113-121.

H Hermansky (1990) "Perceptual linear predictive (PLP) analysis of speech", J Acoust Soc Am, vol 87, n° 4, Apr, 1738-1753.

SE Scrugs & GH Wakefield (1992) "Time-frequency representations of auditory signatures: Dynamic signal models", J Acoust Soc Am, vol 92, n° 4, Oct, 124th Acoustical Society of America Meeting.

X Yang, K Wang & SA Shamma (1992) "Auditory representation of acoustic signals", IEEE Trans Information Theory, vol 38, n° 2, Mar, 824-839.

O Ghitza (1993) "Adequacy of auditory models to predict human internal representation of speech sounds", J Acoust Soc Am, vol 93, n° 4, Apr, 2160-2171.

MF Spiegel, MC Picardi & DM Green (1981) "Signal and masker uncertainty in intensity discrimination", J Acoust Soc Am, vol 70, n° 4, 1015-1019.

MF Spiegel & DM Green (1982) "Signal and masker uncertainty with noise maskers of varying duration, bandwidth, and centre frequency", J Acoust Soc Am, vol 71, n° 5, 1204-1210.

DM Green & G Kidd Jr (1983) "Further studies of auditory profile analysis", J Acoust Soc Am, vol 73, n° 4, 1260-1265.

DM Green, G Kidd Jr & MC Picardi (1983) "Successive versus simultaneous comparison in auditory intensity discrimination", J Acoust Soc Am, vol 73, n° 2, 639-643.

DM Green (1983) "Profile analysis: a different view of auditory intensity discrimination", Am Psychol, Feb, 133-142.

DM Green, CR Mason & G Kidd Jr (1984) "Profile analysis: critical bands and duration", J Acoust Soc Am, vol 75, n° 4, 1163-1167.

LR Bernstein & DM Green (1987) "Detection of simple and complex changes of spectral shape", J Acoust Soc Am, vol 82, n° 5, 1587-1592.

WA Yost & CS Watson (eds) (1987) "Auditory processing of complex sounds", Lawrence Erlbaum, 328 pp.

PF Assmann & Q Summerfield (1990) "Modeling the perception of concurrent vowels: vowels with different fundamental frequencies", J Acoust Soc Am, vol 88, n° 2, Aug, 680-697.

WAC van den Brink & T Houtgast (1990) "Spectro-temporal integration in signal detection", J Acoust Soc Am, vol 88, n° 4, Oct, 1703.

CS Watson, DC Foyle & GR Kidd (1990) "Limits of auditory pattern discrimination for patterns with various durations and numbers of components", J Acoust Soc Am, vol 88, n° 6, Dec, 2631-2638.

G Kidd Jr, CR Mason, RM Uchanski, MA Brantley & P Shah (1991) "Evaluation of simple models of auditory profile analysis using random reference spectra", J Acoust Soc Am, vol 90, n° 3, Sep, 1340-1354.

H Dai & DM Green (1993) "Discrimination of spectral shape as a function of stimulus duration", J Acoust Soc Am, vol 93, n° 2, Feb, 957-965.

G Kidd Jr (1993) "Individual differences in the improvement in spectral shape discrimination due to increasing number of nonsignal tones", J Acoust Soc Am, vol 93, nº 2, Feb, 992-996.

J Zera & DM Green (1993) "Detecting temporal asynchronous standards", J Acoust Soc Am, vol 93, nº 3, Mar, 1571-1052.

J Zera, ZA Onsan, QT Nguyen & DM Green (1993) "Auditory profile analysis of harmonic signals", J Acoust Soc Am, vol 93, nº 6, Jun, 3431-3441.

PSYCHOPHYSICS - HEARING - AUDITORY STREAMING

Y Tougas & AS Bregman (1990) "Auditory streaming and the continuity illusion", Percept & Psychophys, vol 47, 121-126.

GR Kidd & CS Watson (1992) "The "proportion-of-the-total-duration rule" for the discrimination of auditory patterns", J Acoust Soc Am, vol 92, nº 6, Dec, 3109-3118.

PSYCHOPHYSICS - HEARING - BINAURAL EFFECTS

NV Franssen (1964) "Stereophony", Philips Technical Library.

MF Yama (1982) "Differences between psychophysical "suppression effects" under diotic and dichotic listening conditions", J Acoust Soc Am, vol 72, nº 5, 1380-1383.

A Kohlrausch (1988) "Auditory filter shape derived from binaural masking experiments", J Acoust Soc Am, vol 84, nº 2, 573-583.

WA Yost (1988) "The masking-level difference and overall masker level: restating the internal noise hypothesis", J Acoust Soc Am, vol 83, nº 4, 1517-1521.

WA Yost & RH Dye Jr (1988) "Discrimination of interaural differences of level as a function of frequency", J Acoust Soc Am, vol 83, nº 5, 1846-1851.

PSYCHOPHYSICS - HEARING - COCHLEAR MODELS

MA Viergever (?) "Cochlear mechanics: a review".

MH Holmes & LA Rubenfeld (eds) (1981) "Mathematical modeling of the hearing process", Troy, 21-25 Jul 1980, Springer Verlag, 104 pp.

F Grandori (1984) "Inverse solutions for auditory evoked potential fields", in XJR Avula, RE Kalman, AI Liapis, EY Rodin (eds) Mathematical modelling in science and technology, Zurich, 15-17 Aug 1983, Pergamon, 1006 pp, 768-773.

MH Holmes & JD Cole (1984) "Cochlear mechanics: analysis for a pure tone", J Acoust Soc Am, vol 76, n° 3, Sep, 767-778.

JB Allen (1985) "Cochlear modeling", IEEE ASSP Magazine, Jan, 3-29.

RS Chadwick (1985) "Three-dimensional effects on low-frequency cochlear mechanics", Mech Res Comm, vol 12, n° 4, 181-186.

RJ LeVeque, CS Peskin & PD Lax (1985) "Solution of a two-dimensional cochlea model using transform techniques", SIAM J App Math, vol 45, n° 3, Jun, 450-464.

ST Neely (1985) "Mathematical modeling of cochlear mechanics", J Acoust Soc Am, vol 78, n° 1, Jul, 345-352.

HW Strube (1985) "A computationally efficient basilar-membrane model", Acustica, vol 58, 207-214.

E Zwicker (1986) "A hardware cochlear nonlinear preprocessing model with active feedback", J Acoust Soc Am, vol 80, n° 1, Jul, 146-153.

RJ Diependaal, H Duifhuis, HW Hoogstraten & MA Viergever (1987) "Numerical methods for solving one-dimensional cochlear models in the time domain", J Acoust Soc Am, vol 82, n° 5, Nov, 1655-1666.

RW Guelke, JP Ramakers & AE Bunn (1987) "Modelling the cochlea", Electron and Wireless World, vol 93, n° 1611, Jan, 19-21.

RF Lyon & C Mead (1988) "An analog electronic cochlea", IEEE Trans Acoust Speech Signal Process, vol 36, n° 7, Jul, 1119-1134.

E de Boer & C Kruidenier (1990) "On ringing limits of the auditory periphery", Biol Cybern, vol 63, 433-442.

F Mammano (1990) "Modeling auditory system nonlinearities through Volterra series", Biol Cybern, vol 63, 307-313.

PSYCHOPHYSICS - HEARING - DIFFERENCE LIMENS - General

E Schorer (1989) "Vergleich eben erkennbaren Unterschiede und Variationen der Frequenz und Amplitude von Schallen", Acustica, vol 68, 183-199.

E Schorer (1989) "Ein Funktionsschema eben wahrnehmbarer Frequenz- und Amplituden{nderungen", Acustica, vol 68, 268-287.

PSYCHOPHYSICS - HEARING - DIFFERENCE LIMENS - Intensity

JH Johnson, CW Turner, JL Zwislocki & RH Margolis (1993) "Just noticeable differences for intensity and their relation to loudness", J Acoust Soc Am, vol 93, n° 2, Feb, 983-991.

PSYCHOPHYSICS - HEARING - DIFFERENCE LIMENS - Frequency

D Gabor (1946) "Theory of communication", J Inst Electr Eng, vol 93, 429-457.

BCJ Moore (1973) "Frequency difference limens for short-duration tones", J Acoust Soc Am, vol 54, n° 3, 610-619.

BCJ Moore (1973) "Frequency difference limens for narrow bands of noise", J Acoust Soc Am, vol 54, n° 4, 888-896.

BCJ Moore (1974) "Relation between the critical bandwidth and the frequency difference limen", J Acoust Soc Am, vol 55, n° 2, 359.

P Baldi & W Heiligenberg (1988) "How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers", Biol Cybern, vol 59, 313-318.

JP Gagné & PM Zurek (1988) "Resonance-frequency discrimination", J Acoust Soc Am, vol 83, n° 6, Jun, 2293-2299.

BCJ Moore & BR Glasberg (1989) "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation", J Acoust Soc Am, vol 86, n° 5, 1722-1732.

S Deutsch (1990) "On the determination of input sound frequencies by the auditory central processor", IEEE Trans Biomed Eng, vol BME-37, n° 6, Jun, 556-564.

PSYCHOPHYSICS - HEARING - DIFFERENCE LIMENS - Periodicity

I Pollack (1990) "Detection and discrimination thresholds for auditory periodicity", Percept & Psychophys, vol 47, 105-111.

PSYCHOPHYSICS - HEARING - DIFFERENCE LIMENS - Timing

A Michelsen (ed) (1985) "Time resolution in auditory systems", 11th Danavox Gamle Avernaes, 28-31 Aug 1984, Springer-Verlag, 242 pp.

IJ Hirsh, CB Monahan, KW Grant & PG Singh (1990) "Studies in auditory timing: 1 Simple patterns", Percept & Psychophys, vol 47, n° 3, 215-226.

CB Monahan & IJ Hirsh (1990) "Studies in auditory timing: 2 Rhythm patterns", Percept & Psychophys, vol 47, n° 3, 226-242.

BCJ Moore, RW Peters & BR Glasberg (1993) "Detection of temporal gaps in sinusoids: effects of frequency and level", J Acoust Soc Am, vol 93, n° 3, Mar, 1563-1570.

J Zera & DM Green (1993) "Detecting temporal asynchronous standards", J Acoust Soc Am, vol 93, n° 3, Mar, 1571-1052.

PSYCHOPHYSICS - HEARING - ECHOLOCATION

JA Simmons (1989) "A view of the world through the bat's ear: the formation of acoustic images in echolocation", Cognition, vol 33, 155-199.

PSYCHOPHYSICS - HEARING - GENERAL

JL Goldstein, T Baer & NSY Kiang"A theoretical treatment of latency, group delay, and tuning characteristics for auditory-nerve responses to clicks and t... 133-141.

I Pollack & L Ficks (1954) "Information of elementary multidimensional auditory displays", J Acoust Soc Am, vol 26, n° 2, 155-158.

GA Miller (1956) "The magical number seven, plus or minus two: some limits on our capacity for processing information", Psychol Rev, vol 63, n° 2, 81-97.

G von Békésy (1960) "Experiments in hearing", McGraw Hill, 745 pp.

LA Jeffress (1968) "Mathematical models of auditory detection", J Acoust Soc Am, vol 44, n° 1, 187-203.

H Levitt (1971) "Transformed up-down methods in psychoacoustics", J Acoust Soc Am, vol 49, n° 2, 467-477.

SE Gerber (1974) "Introductory hearing science: physical and psychological concepts", WB Saunders, 299 pp.

JR Miller & EC Carterette (1975) "Perceptual space for musical structures", J Acoust Soc Am, vol 58, n° 3, 711-720.

S Singh (ed) (1975) "Measurement procedures in speech, hearing and language", University Park Press, 470 pp.

R Plomp (1976) "Aspects of tone sensation: a psychophysical study", Academic Press, 167 pp.

EF Evans & JP Wilson (eds) (1977) "Psychophysics and physiology of hearing", Keele, 12-16 Apr 1977, Academic Press, 525 pp.

MA Gerzon (1978) "Mathematics and sound perception", J Audio Eng Soc, vol 26, 46-50.

R Taylor (1979) "Noise", Pelican, 274 pp.

PM Haughton (1980) "Physical principles of audiology", Adam Hilger, 183 pp.

JM Festen & R Plomp (1981) "Relations between auditory functions in normal hearing", J Acoust Soc Am, vol 70, n° 2, 356-369.

JJ Barucha & K Stoeckig (1986) "Reaction time and musical expectancy: priming of chords", J Exper Psychol: Human Percept & Perform, vol 12, n° 4, 403-410.

E de Boer & WA Dreschler (1987) "Auditory psychophysics: spectrotemporal representation of signals", Ann Rev Psychol, vol 38, 181-202.

WA Yost & CS Watson (eds) (1987) "Auditory processing of complex sounds", Lawrence Erlbaum, 328 pp.

H Duifhuis, JW Horst & HP Wit (eds) (1988) "Basic issues in hearing", 8th International Symposium on Hearing, Paterswolde, 5-9 Apr 1988, Academic Press, 470 pp.

JP Gagné & PM Zurek (1988) "Resonance-frequency discrimination", J Acoust Soc Am, vol 83, n° 6, 2293-2299.

BCJ Moore (1989) "An introduction to the psychology of hearing (3rd edition)", Academic Press, 350 pp.

E de Boer & C Kruidenier (1990) "On ringing limits of the auditory periphery", Biol Cybern, vol 63, 433-442.

CA Fowler (1990) "Sound-producing sources as objects of perception: rate normalization and nonspeech perception", J Acoust Soc Am, vol 88, n° 3, Sep, 1236-2904.

RL Diehl, MA Walsh & KR Kluender (1991) "On the interpretability of speech/nonspeech comparisons: a reply to Fowler", J Acoust Soc Am, vol 89, n° 6, Jun, 2905-2909.

CA Fowler (1991) "Auditory perception is not special: we see the world, we feel the world, we hear the world", J Acoust Soc Am, vol 89, n° 6, Jun, 2910-2915.

WM Hartmann (1993) "Auditory demonstrations on compact disk for large N", J Opt Soc Am , vol 93, n° 1, Jan, 1-16.

PSYCHOPHYSICS - HEARING - GLIDES AND CHIRPS

GJ Dooley & BCJ Moore (1988) "Detection of linear frequency glides as a function of frequency and duration", J Acoust Soc Am, vol 84, n° 6, 2045-2057.

A Kohlrausch (1988) "Masking patterns of harmonic complex tone maskers and the role of the inner ear transfer function", in H Duifhuis, JW Horst & HP Wit (eds) Basic issues in hearing, 8th International Symposium on Hearing, Paterswolde, 5-9 Apr 1988, Academic Press, 470 pp, 339-350.

PSYCHOPHYSICS - HEARING - INFORMATION

H Jacobson (1950) "The informational capacity of the human ear", Science, vol 112, 4 Aug, 143-144.

PSYCHOPHYSICS - HEARING - MASKING - General

BCJ Moore & BR Glasberg (1981) "Auditory filter shapes derived in simultaneous and forward masking", J Acoust Soc Am, vol 70, n° 4, 1003-1014.

BR Glasberg, BCJ Moore & I Nimmo-Smith (1984) "Comparison of auditory filter shapes derived with three different maskers", J Acoust Soc Am, vol 75, n° 2, 536-544.

LE Humes & W Jesteadt (1989) "Models of the additivity of masking", J Acoust Soc Am, vol 85, n° 3, Mar, 1285-1294.

PSYCHOPHYSICS - HEARING - MASKING - Simultaneous

RD Patterson (1976) "Auditory filter shapes derived with noise stimuli", J Acoust Soc Am, vol 59, n° 3, 640-654.

RD Patterson, I Nimmo-Smith, DL Weber & R Milroy (1982) "The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold", J Acoust Soc Am, vol 72, n° 6, 1788-1803.

S Fidell, R Horonjeff, S Teffeteller & DM Green (1983) "Effective masking bandwidths at low frequencies", J Acoust Soc Am, vol 73, n° 2, 628-638.

RA Lutfi (1983) "Additivity of simultaneous masking", J Acoust Soc Am, vol 73, n° 1, 262-267.

RA Lutfi (1983) "Simultaneous masking and unmasking with bandlimited noise", J Acoust Soc Am, vol 73, n° 3, 899-905.

BCJ Moore & BR Glasberg (1983) "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns", J Acoust Soc Am, vol 74, n° 3, 750-753.

BR Glasberg, BCJ Moore, RD Patterson & I Nimmo-Smith (1984) "Dynamic range and asymmetry of the auditory filter", J Acoust Soc Am, vol 76, n° 2, 419-427.

RA Lutfi & RD Patterson (1984) "On the growth of masking asymmetry with stimulus intensity", J Acoust Soc Am, vol 76, n° 3, 739-745.

SP Bacon & NF Viemeister (1985) "Simultaneous masking by gated and continuous sinusoidal maskers", J Acoust Soc Am, vol 78, n° 4, 1220-1230.

BCJ Moore (1985) "Additivity of simultaneous masking, revisited", J Acoust Soc Am, vol 78, n° 2, 488-494.

BCJ Moore & BR Glasberg (1987) "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns", Hearing Research, vol 28, 209-225.

BCJ Moore, RW Peters & BR Glasberg (1990) "Auditory filter shapes at low center frequencies", J Acoust Soc Am, vol 88, n° 1, Jul, 132-140.

MJ Shailer, BCJ Moore, BR Glasberg, N Watson & S Harris (1990) "Auditory filter shapes at 8 and 10 kHz", J Acoust Soc Am, vol 88, n° 1, Jul, 141-148.

S Rosen & D Stock (1992) "Auditory filter bandwidths as a function of level at low frequencies (125 Hz - 1 kHz)", J Acoust Soc Am, vol 92, n° 2, Aug, 773-781.

PSYCHOPHYSICS - HEARING - MASKING - Suppression

M Terry & BCJ Moore (1977) ""Suppression" effects in forward masking", J Acoust Soc Am, vol 62, n° 3, 781-784.

DL Weber (1978) "Suppression and critical bands in band-limiting experiments", J Acoust Soc Am, vol 64, n° 1, 141-150.

DL Weber & DM Green (1979) "Suppression effects in backward and forward masking", J Acoust Soc Am, vol 65, n° 5, 1258-1267.

BCJ Moore & BR Glasberg (1982) "Interpreting the role of suppression in psychophysical tuning curves", J Acoust Soc Am, vol 72, n° 5, 1374-1379.

MF Yama (1982) "Differences between psychophysical "suppression effects" under diotic and dichotic listening conditions", J Acoust Soc Am, vol 72, n° 5, 1380-1383.

DL Weber (1983) "Do off-frequency simultaneous maskers suppress the signal?", J Acoust Soc Am, vol 73, n° 3, 887-893.

PSYCHOPHYSICS - HEARING - MASKING - Temporal

H Duifhuis (1973) "Consequences of peripheral frequency selectivity for nonsimultaneous masking", J Acoust Soc Am, vol 54, n° 6, 1471-1488.

CE Robinson & I Pollack (1973) "Interaction between forward and backward masking: a measure of the integrating period of the auditory system", J Acoust Soc Am, vol 53, n° 5, 1313-1316.

H Fastl (1976) "Temporal masking effects: I Broad band noise masker", Acustica, vol 35, n° 5, 287-302.

H Fastl (1977) "Temporal masking effects: II Critical band noise masker", Acustica, vol 36, n° 5, 317-331.

T Houtgast (1977) "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker", J Acoust Soc Am, vol 62, n° 2, 409-415.

MJ Penner (1977) "Detection of temporal gaps in noise as a measure of the decay of auditory sensation", J Acoust Soc Am, vol 61, n° 2, 552-557.

MJ Penner (1978) "A power law resulting in a class of short-term integrators that produce time-intensity trades for noise bursts", J Acoust Soc Am, vol 63, n° 1, 195-201.

H Fastl (1979) "Temporal masking effects: III Pure tone masker", Acustica, vol 43, 282-294.

MJ Penner (1979) "Forward masking with equal-energy maskers", J Acoust Soc Am, vol 66, n° 6, 1719-1724.

NF Viemeister (1979) "Temporal modulation transfer functions based upon modulation thresholds", J Acoust Soc Am, vol 66, n° 5, 1364-1380.

GP Widin & NF Viemeister (1979) "Short-term spectral effects in pure-tone forward masking", J Acoust Soc Am, vol 66, n° 2, 396-399.

GP Widin & NF Viemeister (1979) "Intensive and temporal effects in pure-tone forward masking", J Acoust Soc Am, vol 66, n° 2, 388-395.

H Fastl & M Bechly (1981) "Post masking with two maskers: effects of bandwidth", J Acoust Soc Am, vol 69, n° 9, 1753-1757.

BCJ Moore (1981) "Interactions of masker bandwidth with signal duration and delay in forward masking", J Acoust Soc Am, vol 70, n° 1, 62-68.

J Verschuure (1981) "Pulsation patterns and nonlinearity of auditory tuning: I Psychophysical results; II Analysis of psychophysical results", Acustica, vol 49, 288-306.

DL Weber & BCJ Moore (1981) "Forward masking by sinusoidal and noise maskers", J Acoust Soc Am, vol 69, n° 5, 1402-1409.

W Jesteadt, SP Bacon & JR Lehman (1982) "Forward masking as a function of frequency, masker level, and signal delay", J Acoust Soc Am, vol 71, n° 4, 950-962.

G Kidd Jr & LL Feth (1982) "Effects of masker duration in pure-tone forward masking", J Acoust Soc Am, vol 72, n° 5, 1384-1386.

PJ Fitzgibbons (1983) "Temporal gap detection in noise as a function of frequency, bandwidth, and level", J Acoust Soc Am, vol 74, n° 1, 67-72.

M Florentine & S Buus (1983) "Temporal acuity as a function of level and frequency", Proc 11th Int Cong Acoust, 103-106.

BCJ Moore & BR Glasberg (1983) "Growth of forward masking for sinusoidal and noise maskers as a function of signal delay; implications for suppression in noise", J Acoust Soc Am, vol 73, n° 4, 1249-1259.

MJ Shailer & BCJ Moore (1983) "Gap detection as a function of frequency, bandwidth, and level", J Acoust Soc Am, vol 74, n° 2, 467-473.

RA Lutfi (1984) "Predicting frequency selectivity in forward masking from simultaneous masking", J Acoust Soc Am, vol 76, n° 4, 1045-1050.

E Zwicker (1984) "Dependence of post-masking on masker duration and its relation to temporal effects in loudness", J Acoust Soc Am, vol 75, n° 1, 219-223.

BCJ Moore (1985) "Comments on "Predicting frequency selectivity in forward masking from simultaneous masking"", J Acoust Soc Am, vol 78, n° 1, 253-260.

MJ Shailer & BCJ Moore (1985) "Detection of temporal gaps in bandlimited noise: effects of variations in bandwidth and signal-to-masker ratio", J Acoust Soc Am, vol 77, n° 2, 635-639.

SP Bacon & BCJ Moore (1986) "Temporal effects in masking and their influence on psychophysical tuning curves", J Acoust Soc Am, vol 80, n° 6, 1638-1645.

BCJ Moore & BJ O'Loughlin (1986) "The use of nonsimultaneous masking to measure frequency selectivity and suppression", Frequency Selectivity in Hearing, Academic Press, 179-250.

SP Bacon & W Jesteadt (1987) "Effects of pure-tone forward masker duration on psychophysical measures of frequency selectivity", J Acoust Soc Am, vol 82, n° 6, 1925-1932.

TG Forrest & DM Green (1987) "Detection of partially filled gaps in noise and the temporal modulation transfer function", J Acoust Soc Am, vol 82, n° 6, 1933-1943.

BCJ Moore, PWF Poon, SP Bacon & BR Glasberg (1987) "The temporal course of masking and the auditory filter shape", J Acoust Soc Am, vol 81, n° 6, 1873-1880.

M Florentine, H Fastl & S Buus (1988) "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking", J Acoust Soc Am, vol 84, n° 1, 195-203.

G Formby & K Muir (1988) "Modulation and gap detection for broadband and filtered noise signals", J Acoust Soc Am, vol 84, n° 2, 545-550.

RA Lutfi (1988) "Interpreting measures of frequency selectivity: is forward masking special?", J Acoust Soc Am, vol 83, n° 1, 163-177.

BCJ Moore, BR Glasberg, CJ Plack & AK Biswas (1988) "The shape of the ear's temporal window", J Acoust Soc Am, vol 83, n° 3, 1102-1116.

DM Green & TG Forrest (1989) "Temporal gaps in noise and sinusoids", J Acoust Soc Am, vol 86, n° 3, Sep, 961-970.

SP Bacon (1990) "Effect of masker level on overshoot", J Acoust Soc Am, vol 88, n° 2, 698-702.

CJ Plack & BCJ Moore (1990) "Temporal window shape as a function of frequency and level", J Acoust Soc Am, vol 87, n° 5, May, 2178-2187.

PSYCHOPHYSICS - HEARING - MUSIC

RN Shepard (1964) "Circularity in judgements of relative pitch", J Acoust Soc Am, vol 36, n° 12, Dec, 2346-2353.

O Karolyi (1965) "Introducing music", Penguin, 176 pp.

JR Miller & EC Carterette (1975) "Perceptual space for musical structures", J Acoust Soc Am, vol 58, n° 3, 711-720.

R Shuter-Dyson & C Gabriel (1981) "The psychology of musical ability".

JJ Bharucha & K Stoeckig (1986) "Reaction time and musical expectancy: priming of chords", J Exper Psychol: Human Percept & Perform, vol 12, n° 4, 403-410.

C Palmer & CL Krumhansl (1987) "Independent temporal and pitch structures in determination of musical phrases", J Exper Psychol: Human Percept & Perform, vol 13, n° 1, 116-126.

D Deutsch (1992) "Paradoxes of musical pitch", Scientific American, Aug, 70-75.

T Kelsey (1993) "Musical harmonies disrupted by miner's 'revolutionary theory'", Independent, 2 Jun, 7.

J Sloboda (1994) "Becoming a musician", Annual British Association for the Advancement of Science Meeting, Keele University, Aug 1994, not published. Reported in Independent.

PSYCHOPHYSICS - HEARING - LOCALISATION

Y Hiranaka & H Yamasaki (1983) "Envelope representations of pinna impulse responses relating to three-dimensional localization of sound sources", J Acoust Soc Am, vol 73, n° 1, 291-296.

FL Wightman & DJ Kistler (1989) "Headphone simulation of free-field listening: I Stimulus synthesis, II Psychophysical validation", J Acoust Soc Am, vol 85, n° 2, Feb, 858-878.

JC Makous & JC Middlebrooks (1990) "Two-dimensional sound localisation by human listeners", J Acoust Soc Am, vol 87, n° 5, May, 2188-2200.

DR Perrot & K Saberi (1990) "Minimum audible angle thresholds for sources varying in both elevation and azimuth", J Acoust Soc Am, vol 87, n° 4, Apr, 1728-1731.

K Saberi & DR Perrot (1990) "Lateralisation thresholds obtained under conditions in which the precedence effect is assumed to operate", J Acoust Soc Am, vol 87, n° 4, Apr, 1732-1737.

DR Perrott, B Costantino & J Ball (1993) "Discrimination of moving events which accelerate or decelerate over the listening interval", J Acoust Soc Am, vol 93, n° 2, Feb, 1053-1057.

DR Perrott, B Costantino & J Cisneros (1993) "Auditory and visual localization performance in a sequential discrimination task", J Acoust Soc Am, vol 93, n° 1, Apr, 2134-2138.

EM Wenzel, M Arrunda, DJ Kistler & FL Wightman (1993) "Localization using nonindividualized head-related transfer functions", J Acoust Soc Am, vol 94, n° 1, Jul, 111-123.

PSYCHOPHYSICS - HEARING - OTOACOUSTIC EMISSIONS

C Dallmayr (1987) "Stationary and dynamical properties of simultaneous evoked otoacoustic emissions(SEOAE)", Acustica, vol 63, n° 4, Jun, 243-255.

PSYCHOPHYSICS - HEARING - PHASE EFFECTS

RH Gilkey & DE Robinson (1986) "Models of auditory masking: a molecular psychophysical approach", J Acoust Soc Am, vol 79, n° 5, 1499-1510.

MR Schroeder (1986) "Auditory paradox based on fractal waveform", J Acoust Soc Am, vol 79, n° 1, 186-189.

MR Schroeder & HW Strube (1986) "Flat-spectrum speech", J Acoust Soc Am, vol 79, n° 5, 1580-1583.

BK Smith, UK Sieben, A Kohlrausch & MR Schroeder (1986) "Phase effects in masking related to dispersion in the inner ear", J Acoust Soc Am, vol 80, n° 6, 1631-1637.

BCJ Moore & BR Glasberg (1987) "Factors affecting thresholds for sinusoidal signals in narrow-band maskers with fluctuating envelopes", J Acoust Soc Am, vol 82, n° 1, 69-79.

RD Patterson (1987) "A pulse ribbon model of monaural phase perception", J Acoust Soc Am, vol 82, n° 5, Nov, 1560-1586.

GP Schooneveldt & BCJ Moore (1987) "Comodulation masking release (CMR): effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band", J Acoust Soc Am, vol 82, n° 6, 1944-1956.

A van den Bos (1987) "A new method for synthesis of low-peak-factor signals", IEEE Trans Acoust Speech Signal Process, vol ASSP-35, n° 1, 120-122.

WM Hartmann & J Pumplin (1988) "Noise power fluctuations and the masking of sine signals", J Acoust Soc Am, vol 83, n° 6, 2277-2289.

A Scherer (1988) "Erkl{rung der spektralen Verdeckung mit Hilfe von Mith¦rschwellen- und Suppressionsmustern", Acustica, vol 67, 1-18.

GP Schooneveldt & BCJ Moore (1988) "Failure to obtain comodulation masking release with frequency-modulated maskers", J Acoust Soc Am, vol 83, n° 6, 2290-2292.

C Kaernbach (1993) "Temporal and spectral basis of the features perceived in repeated noise", J Acoust Soc Am, vol 94, n° 1, Jul, 91-97.

PSYCHOPHYSICS - MULTIDIMENSIONAL SCALING

JB Kruskal (1964) "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis", Psychometrika, vol 29, nº 1, Mar, 1-27.

JB Kruskal (1964) "Nonmetric multidimensional scaling: a numerical method", Psychometrika, vol 29, nº 2, Jun, 115-129.

N Cliff (1966) "Orthogonal rotation to congruence", Pychometrika, vol 31, nº 1, Mar, 33-42.

EE Roskam (1970) "The method of triads for nonmetric multidimensional scaling", Ned Tijdschrift Psychol, vol 25, 404-417.

WA Wagenaar & P Padmos (1971) "Quantitative interpretation of stress in Kruskal's multidimensional scaling technique", Br J Math Statist Psychol, vol 24, 101-110.

RN Shepard, AK Romney & SB Nerlove (1972) "Multidimensional scaling: vol I theory, vol II applications", Seminar Press, 584 pp.

JC Lingoes & EE Roskam (1973) "A mathematical and empirical analysis of two multidimensional scaling algorithms", Psychometric Society, 93 pp.

RN Shepard (1974) "Representation of structure in similarity data: Problems and prospects", Psycometrika, vol 39, nº 4, Dec, 373-421.

JR Miller & EC Carterette (1975) "Perceptual space for musical structures", J Acoust Soc Am, vol 58, nº 3, 711-720.

JD Carroll & P Arabie (1980) "Multidimensional scaling", Ann Rev Psychol, vol 31, 607-649.

RN Shepard (1980) "Multidimensional scaling, tree-fitting, and clustering", Science, vol 210, 24 Oct, 390-398.

SS Schiffman, ML Reynolds & FW Young (1981) "Introduction to multidimensional scaling: theory, methods and applications", Academic Press, 413 pp.

FW Young (1984) "Scaling", Ann Rev Psychol, vol 35, 55-81.

PSYCHOPHYSICS - SPEECH - GENERAL

JN Holmes (1972) "Speech synthesis", Mills & Boon, 68 pp.

S Singh (ed) (1975) "Measurement procedures in speech, hearing and language", University Park Press, 470 pp.

DA Sanders (1977) "Auditory perception of speech: an introduction to principles and problems", Prentice Hall, 246 pp.

LR Rabiner & RW Schafer (1978) "Digital processing of speech signals", Prentice Hall, 512 pp.

DB Pisoni (1985) "Speech perception: some new directions in research and theory", J Acoust Soc Am, vol 78, n° 1, Jul, 381-388.

RJ McAulay & TF Quatieri (1986) "Speech analysis/synthesis based on a sinusoidal representation", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 4, 744-754.

TF Quatieri & RJ McAulay (1986) "Speech transformations based on a sinusoidal representation", IEEE Trans Acoust Speech Signal Process, vol ASSP-34, n° 6, 1449-1464.

MR Schroeder & HW Strube (1986) "Flat-spectrum speech", J Acoust Soc Am, vol 79, n° 5, 1580-1583.

JN Holmes (1988) "Speech synthesis and recognition", Van Nostrand Reinhold, 198 pp.

O Ghitza (1993) "Adequacy of auditory models to predict human internal representation of speech sounds", J Acoust Soc Am, vol 93, n° 4, Apr, 2160-2171.

PSYCHOPHYSICS - SIGHT - COLOUR

DH Kelly (1989) "Opponent-color receptive-field profiles determined from large-area psychophysical measurements", J Opt Soc Am A, vol 6, n° 11, Nov, 1784-1793.

TM Man & DL MacAdam (1989) "Three-dimensional scaling of the uniform color scales of the Optical Society of America", J Opt Soc Am A, vol 6, n° 1, Jan, 128-138.

DL MacAdam (1990) "Redetermination of colors for uniform scales", J Opt Soc Am A, vol 7, n° 1, Jan, 113-115.

MA Webster, KK de Valois & E Switkes (1990) "Orientation and spatial-frequency discrimination for luminance and chromatic gratings", J Opt Soc Am A, vol 7, n° 6, Jun, 1034-1049.

I Abramov, J Gordon & H Chan (1991) "Color appearance in the peripheral retina", J Opt Soc Am A, vol 8, n° 2, Feb, 404-414.

SL Guth (1991) "Model for color vision and light adaptation", J Opt Soc Am A, vol 8, n° 6, Jun, 976-993.

C Oleari (1991) "Uniform-scale chromaticity diagram with angular coordinates in zero-curvature space", J Opt Soc Am A, vol 8, n° 2, Feb, 415-421.

I Abramov, J Gordon & H Chan (1992) "Color appearance across the retina: effects of a white surround", J Opt Soc Am A, vol 9, nº 2, Feb, 195-202.

M Melgosa, E Hita, J Romero & L Jimenez del Barco (1992) "Some classical color differences calculated with new formulas", J Opt Soc Am A, vol 9, nº 8, Aug, 1247-1254.

C Oleari (1993) "Uniform color space for 10º visual field and OSA uniform color scales", J Opt Soc Am A, vol 10, nº 7, Jul, 1490-1498.

HJ Trussell (1993) "DSP solutions run the gamut for color systems", IEEE SP magazine, Apr, 8-23.

PSYCHOPHYSICS - SIGHT - EYE MOVEMENTS

D Noton & L Stark (1971) "Scanpaths in saccadic eye movements while viewing and recognizing patterns", Vision Res, vol 11, 929-42.

AT Bahill, MR Clark & L Stark (1975) "The main sequence, a tool for studying human eye movements", Math Biosci, vol 24, 191-204.

CM Zingale & E Kowler (1987) "Planning sequences of saccades", Vision Res, vol 27, nº 8, 1327-1341.

W Becker & R Jürgens (1990) "Human oblique saccades: quantitative analysis of the relation between horizontal and vertical components", Vision Res, vol 30, nº 6, 893-920.

SG Whittaker & RW Cummings (1990) "Foveating saccades", Vision Res, vol 30, nº 9, 1363-1366.

DS Greenhouse & TE Cohn (1991) "Saccadic suppression and stimulus uncertainty", J Opt Soc Am A, vol 8, nº 3, Mar, 587-595.

PSYCHOPHYSICS - SIGHT - GENERAL

PJ Barber & D Legge (1976) "Perception and information (Essential Psychology A4)", Methuen, 144 pp.

D Marr (1982) "Vision", WH Freeman, 397 pp.

MR Schroeder (1983) "The eikonal equation", Mathematical Intelligencer, vol 5, 36-37.

RL Gregory (1986) "Eye and brain: the psychology of seeing", Weidenfeld and Nicholson, 256 pp.

C Blakemore (ed) (1990) "Vision: coding and efficiency", Cambridge University Press, 448 pp.

RJ Watt (1991) "Understanding Vision", Academic Press, 301 pp.

PSYCHOPHYSICS - SIGHT - INFORMATION

H Jacobson (1951) "The informational capacity of the human eye", Science, vol 113, 16 Mar, 292-293.

PSYCHOPHYSICS - SIGHT - VISUAL FIELD

RT Brooke (1951) "The variation of critical fusion frequency with brightness at various retinal locations", J Opt Soc Am, vol 41, Dec, 1010-1016.

A Ransom-Hogg & L Spillmann (1980) "Perceptive field size in fovea and periphery of the light- and dark-adapted retina", Vision Res, vol 20, 221-228.

J Saarinen (1987) "Perception of positional relationships between line segments in eccentric vision", Perception, vol 16, nº 5, 583-591.

JS Pointer & RF Hess (1989) "The contrast sensitivity gradient across the human visual field: with emphasis on the low frequency range", Vision Res, vol 29, nº 9, 1133-1151.

JS Pointer & RF Hess (1990) "The contrast sensitivity gradient across the major oblique meridians of the human visual field", Vision Res, vol 30, nº 3, 497-501.

CW Tyler & RD Hamer (1990) "Analysis of visual modulation sensitivity: IV Validity of the Ferry-Porter law", J Opt Soc Am A, vol 7, nº 4, Apr, 743-758.

E Peli, J Yang & RB Goldstein (1991) "Image invariance with changes in size: the role of peripheral contrast thresholds", J Opt Soc Am A, vol 8, nº 11, Nov, 1762-1774.

RP Scobey & PLE van Kan (1991) "A horizontal stripe of displacement sensitivity in the human visual field", Vision Res, vol 31, nº 1, 99-109.

I Abramov, J Gordon & H Chan (1992) "Color appearance across the retina: effects of a white surround", J Opt Soc Am A, vol 9, nº 2, Feb, 195-202.

P Bijl, JJ Koenderink & AML Kappers (1992) "Deviations from strict M scaling", J Opt Soc Am A, vol 9, nº 8, Aug, 1233-1239.

R Navarro, P Artal & DR Williams (1993) "Modulation transfer of the human eye as a function of retinal eccentricity", J Opt Soc Am A, vol 10, nº 2, Feb, 201-212.

YY Zeevi & E Shlomot (1993) "Nonuniform sampling and antialiasing in image representation", IEEE Trans Signal Processing, vol 41, n° 3, Mar, 1223-1236.

PSYCHOPHYSICS - SIGHT - 2D VISION - Edges

F Bergholm (1987) "Edge focusing", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 6, Nov, 726-741.

YG Leclerc & SW Zucker (1987) "The local structure of image discontinuities in one dimension", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 3, May, 341-355.

S Mallat & W Liang Hwang (1992) "Singularity and processing with wavelets", IEEE Trans Information Theory, vol 38, n° 2, Mar, 617-643.

PSYCHOPHYSICS - SIGHT - 2D VISION - Faces

M Turk & A Pentland (1991) "Eigenfaces for recognition", J Cognitive Neuroscience, vol 3, no 1, 71-86.

NL Etcoff & JL Magee (1992) "Categorical perception of facial expressions", Cognition, vol 44, 227-240.

AJ O'Toole, H Abdi, KA Deffenbacher & D Valentin (1993) "Low-dimensional representation of faces in higher dimensions of the face space", J Opt Soc Am A, vol 10, n° 3, Mar, 405-411.

G Rhodes, S Brake & AP Atkinson (1993) "What's lost in inverted faces?", Cognition, vol 47, 25-57.

A Pentland, B Moghaddam, T Starner, O Oliyide & M Turk (1994) "View-based and modular eigenspaces for face recognition", MIT Media Laboratory Perceptual Computing Section Technical Report n° 245, MIT, 11 pp.

PSYCHOPHYSICS - SIGHT - 2D VISION - Feature analysis

LA Riggs, F Ratliff, JC Cornsweet & TN Cornsweet (1953) "The disappearance of steadily fixated test objects", J Opt Soc Am, vol 43, n° 6, 495-501.

JJ Koenderink & AJ van Doorn (1986) "Dynamic shape", Biol Cybern, vol 53, 383-396.

A Treisman (1986) "Features and objects in visual processing", Scientific American, vol 255, n° 5, 106-115.

SM Pizer, WR Oliver & SH Bloomberg (1987) "Hierarchical shape description via the multiresolution symmetric axis transform", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 4, Jul, 505-511.

I Alter & EL Schwartz (1988) "Psychophysical studies of shape with Fourier descriptor stimuli", Perception, vol 17, 191-202.

PJ Burt (1988) "Smart sensing within a pyramid vision machine", Proc IEEE, vol 76, n° 8, 1006-1015.

KA Stevens & A Brookes (1988) "The concave cusp as a determiner of figure-ground", Perception, vol 17, 35-42.

A Treisman & S Gormican (1988) "Feature analysis in early vision: evidence from search asymmetries", Psychol Rev, vol 95, n° 1, 15-48.

HR Wilson & WA Richards (1989) "Mechanisms of contour curvature discrimination", J Opt Soc Am A, vol 6, n° 1, Jan, 106-115.

PSYCHOPHYSICS - SIGHT - 2D VISION - General (see also: MATHEMATICS - SIGNAL PROCESSING - SPACE-FREQUENCY ANALYSIS)

F Attneave (1954) "Some informational aspects of visual perception", Psychol Rev, vol 61, n° 3, 183-193.

R Sekuler (1974) "Spatial vision", ?, 195-232.

JJ Koenderink & AJ van Doorn (1978) "Visual detection of spatial contrast: influence of location in the visual field, target extent, and illuminance level", Biol Cybern, vol 30, 157-167.

JJ Koenderink & AJ van Doorn (1979) "The structure of two-dimensional scalar fields with application to vision", Biol Cybern, vol 33, 151-158.

C Braccini, G Gambardella, G Sandini & V Tagliasco (1982) "A model of the early stages of the human visual system: functional and topological transformations performed in the peripheral visual field", Biol Cybern, vol 44, 47-58.

G Hartmann (1982) "Recursive fields of circular receptive fields", Biol Cybern, vol 43, 199-208.

JJ Kulikowski, S Marčelya & PO Bishop (1982) "Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex", Biol Cybern, vol 43, 187-198.

JJ Koenderink (1984) "The structure of images", Biol Cybern, vol 50, 363-370.

G Westheimer (1984) "Spatial vision", Ann Rev Psychol, vol 35, 201-226.

N Graham (1985) "Detection and identification of near-threshold visual patterns", J Opt Soc Am, vol 2, n° 9, 1468-1482.

D Sagi & B Julesz (1985) ""Where" and "what" in vision", Science, vol 228, 1217-1219.

GJ Burton, ND Haig & IR Moorhead (1986) "A self-similar stack model for human and machine vision", Biol Cybern, vol 53, 397-403.

GL Schulman, MA Sullivan, K Gish & WJ Sakoda (1986) "The role of spatial-frequency channels in the perception of local and global structure", Perception, vol 15, 259-273.

MJ Carlotto (1987) "Histogram analysis using a scale-space approach", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 1, Jan, 121-129.

AR Dill, MD Levine & PB Noble (1987) "Multiple resolution skeletons", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 4, Jul, 495-504.

DJ Field (1987) "Relations between the statistics of natural images and the response properties of cortical cells", J Opt Soc Am, vol 4, n° 12, 2379-2394.

CB Fisher & MP Fracasso (1987) "The Goldmeyer effect in adults and children: environmental, retinal, and phenomenal influences on judgements of visual symmetry", Perception, vol 16, 29-39.

RM Haralick, SR Sternberg & X Zhuang (1987) "Image analysis using mathematical morphology", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, n° 4, Jul, 532-550.

JJ Koenderink & AJ van Doorn (1987) "Representation of local geometry in the visual system", Biol Cybern, vol 55, 367-375.

BJA Kr¦se (1987) "Local structure analysers as determinants of preattentive pattern discrimination", Biol Cybern, vol 55, 289-298.

AB Watson (1987) "Efficiency of a model human image code", J Opt Soc Am, vol 4, n° 12, 2401-2417.

AB Watson (1987) "Estimation of local spatial scale", J Opt Soc Am, vol 4, n° 8, 1579-1582.

G Borgefors (1988) "Hierarchical chamfer matching: a parametric edge matching algorithm", IEEE Trans Pattern Anal Machine Intel, vol PAMI-10, n° 6, Nov, 849-865.

JG Daugman (1988) "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression", IEEE Trans Acoust Speech Signal Process, vol ASSP-36, n° 7, 1169-1179.

DM Kammen & AL Yuille (1988) "Spontaneous symmetry-breaking energy functions and the emergence of orientation selective cortical cells", Biol Cybern, vol 59, 23-31.

JJ Koenderink (1988) "Operational significance of receptive field assemblies", Biol Cybern, vol 58, 163-171.

JJ Koenderink & W Richards (1988) "Two-dimensional curvature operators", J Opt Soc Am, vol 5, n° 7, 1136-1141.

E Micheli-Tzanakou (1988) "Neural aspects of vision and related technological advances", Proc IEEE, vol 76, n° 9, 1130-1142.

M Porat & YY Zeevi (1988) "The generalised Gabor scheme of image representation in biological and machine vision", IEEE Trans Pattern Anal Machine Intel, vol PAMI-10, n° 4, Jul, 452-468.

CH Teh & RT Chin (1988) "On image analysis by the method of moments", IEEE Trans Pattern Anal Machine Intel, vol PAMI-10, n° 4, Jul, 496-513.

RW Conners & CT Ng (1989) "Developing a quantitative model of human preattentive vision".

JG Daugman (1989) "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields", IEEE Trans Biomed Eng, vol BME-36, n° 1, Jan, 107-114.

DP Lulich & KA Stevens (1989) "Differential contributions of circular and elongated spatial filters to the cafe wall illusion", Biol Cybern, vol 61, 427-435.

A Rosenfeld (1989) "Computer vision: a source of models for biological visual processes?", IEEE Trans Biomed Eng, vol BME-36, n° 1, Jan, 93-96.

AB Watson & AJ Ahumada Jr (1989) "A hexagonal orthogonal-oriented pyramid as a model of image representation in visual cortex", IEEE Trans Biomed Eng, vol BME-36, n° 1, Jan, 97-106.

HR Wilson & WA Richards (1989) "Mechanisms of contour perception", J Opt Soc Am A, vol 6, n° 1, Jan, 106-115.

JJ Koenderink & AJ van Doorn (1990) "Receptive field families", Biol Cybern, vol 63, 291-297.

JB Martens (1990) "The Hermite transform: Theory; Applications", IEEE Trans Acoust Speech Signal Process, vol ASSP-38, nº 9, Sep, 1595-1618.

DG Stork & HR Wilson (1990) "Do Gabor functions provide appropriate descriptions of visual cortical receptive fields?", J Opt Soc Am A, vol 7, nº 8, Aug, 1362-1373.

AL Stewart & R Pinkham (1991) "A space-variant differential operator for visual sensitivity", Biol Cybern, vol 64, 373-379.

J Yang (1992) "Do Gabor functions provide appropriate descriptions of visual cortical receptive fields?: Comment", J Opt Soc Am A, vol 9, nº 2, Feb, 334-340.

PSYCHOPHYSICS - SIGHT - 2D VISION - Motion and temporal effects

DJ Fleet, PE Hallett & AD Jepson (1985) "Spatiotemporal inseparability in early visual processing", Biol Cybern, vol 52, 153-164.

JK Kearney, WB Thompson & DL Boley (1987) "Optical flow estimation: an error analysis of gradient-based methods with local optimization", IEEE Trans Pattern Anal Machine Intel, vol PAMI-9, nº 2, Mar, 229-244.

SJ Anderson, DC Burr & MC Morrone (1991) "Two-dimendional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision", J Opt Soc Am A, vol 8, nº 8, Aug, 1340-1351.

AC den Brinker & JAJ Roufs (1992) "Evidence for a generalized Laguerre transform of temporal events by the visual system", Biol Cybern, vol 67, 395-402.

PSYCHOPHYSICS - SIGHT - 2D VISION - Orientation

DH Hubel & TN Wiesel (1962) "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", J Physiol, vol 160, 106-54.

DH Hubel & TN Wiesel (1968) "Receptive fields and functional architecture of monkey striate cortex", J Physiol, vol 195, 215-243.

KA Stevens (1978) "Computation of locally parallel structure", Biol Cybern, vol 29, 19-28.

V Braitenberg & C Braitenberg (1979) "Geometry of orientation columns in the visual cortex", Biol Cybern, vol 33, 179-186.

SJ Anderson & DC Burr (1991) "Spatial summation properties of directionally selective mechanisms in human vision", J Opt Soc Am A, vol 8, n° 8, Aug, 1330-1339.

PSYCHOPHYSICS - SIGHT - 2D VISION - Phase effects

MJ Morgan, J Ross & A Hayes (1991) "The relative importance of local phase and local amplitude in patchwise image reconstruction", Biol Cybern, vol 65, 113-119.

PSYCHOPHYSICS - SIGHT - 2D VISION - Texture

T Caelli, B Julesz & E Gilbert (1978) "On perceptual analysers underlying visual texture discrimination: Part II", Biol Cybern, vol 29, 201-214.

T Caelli & B Julesz (1978) "On perceptual analysers underlying visual texture discrimination: Part I", Biol Cybern, vol 28, 167-175.

B Julesz (1981) "Textons, the elements of texture perception, and their interactions", Nature, vol 290, 91-97.

JR Bergen & B Julesz (1983) "Parallel versus serial processing in rapid pattern discrimination", Nature, vol 303, 696-698.

B Julesz & JR Bergen (1983) "Textons, the fundamental elements in preattentive vision and perception of textures", The Bell System Tech J, vol 62, n° 6, 1619-1645.

B Julesz (1986) "Texton gradients: the texton theory revisited", Biol Cybern, vol 54, 245-251.

MR Turner (1986) "Texture discrimination by Gabor functions", Biol Cybern, vol 55, 71-82.

I Fogel & D Sagi (1989) "Gabor filters as texture discriminator", Biol Cybern, vol 61, 103-113.

M Porat & YY Zeevi (1989) "Localized texture processing in vision: analysis and synthesis in the Gaborian space", IEEE Trans Biomed Eng, vol BME-36, n° 1, Jan, 115-129.

A Gorea & TV Papathomas (1993) "Double opponency as a generalized concept in texture segregation illustrated with stimuli defined by color, luminance, and orientation", J Opt Soc Am A, vol 10, n° 7, Jul, 1450-1462.

J Xing & GL Gerstein (1993) "A neural network model for texture discrimination", Biol Cybern, vol 69, 97-108.

JI Yellott Jr (1993) "Implications of triple correlation uniqueness for texture statistics and the Julesz conjecture",

J Opt Soc Am A, vol 10, n 5, May, 777-793.

PSYCHOPHYSICS - SIGHT - 2D VISION - Vision-based picture representation

JL Mannos & DJ Sakrison (1974) "The effects of a visual fidelity criterion on the encoding of images", IEEE Trans Inform Theory, vol IT-20, n° 4, Jul, 525-536.

DJ Sakrison (1977) "On the role of the observer and a distortion measure in image transmission", IEEE Trans Commun, vol COM-25, n° 11, Nov, 1251-1267.

AB Watson (1990) "Perceptual-components architecture for digital video", J Opt Soc Am A, vol 7, n° 10, Oct, 1943-1954.

PSYCHOPHYSICS - SIGHT - 3D VISION - General

JT Todd & RA Akerstrom (1987) "Perception of three-dimensional form from patterns of optical texture", J Exper Psychol: Human Percept & Perform, vol 13, n° 2, 242-255.