Stellingen behorend bij proefschrift
"Coloration and binaural decoloration of sound due to reflections"

A. M. Salomens

1. De kleur van een geluid wordt omschreven als het kenmerk waarmee de luisteraar kan beoordelen dat twee geluiden, op dezelfde manier gepresenteerd en met dezelfde luidheid, verschillen.
   — Dit proefschrift, paragraaf 1.2

2. De sterkte van kleuring ten gevolge van reflecties hangt het sterkst samen met de onderlinge afstand van de reflecties in de tijd, meer dan met het aantal reflecties.
   — Dit proefschrift, paragraaf 2.6

3. Binaurale ontkleurings is het sterkst als linker- en rechter- signaal (spectraal) complementair zijn, maar kleur is dan beslist nog niet afwezig.
   — Dit proefschrift, paragraaf 6.7

   — Dit proefschrift, paragraaf 6.7

5. De eenzaamheid van slechthorenden wordt onderschat door horenden. Als slechthorende hoor je vaak niet bij.


7. Zolang de vegetariër in Frankrijk als een afwijking wordt beschouwd, heeft de Franse Haute Cuisine een kans gemist.

8. Het belang van wetenschappelijk onderzoek wordt veelal niet gevonden in de motivering vooraf maar door rechtvaardiging achteraf.

9. Selecteren aan de poort van de universiteit zou niet (alleen) op cijfers, maar veeleer op motivatie moeten gebeuren.

10. Veel maatschappelijke verworvenheden worden als rechten beschouwd, terwijl deze zouden moeten worden beschouwd als voorrechten die opgegeven moeten worden zodra de maatschappelijke situatie daarom vraagt.

COLORATION AND BINAURAL DECOLORATION
OF SOUND
DUE TO REFLECTIONS

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.F. Wakker,
in het openbaar te verdedigen ten overstaan van een commissie,
door het College van Dekanen aangewezen,
op woensdag 20 december 1995 te 10.30 uur

door

Aurelia Maria SALOMONS

natuurkundig ingenieur

geboren te Alphen aan den Rijn
Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. F.A. Bilsen

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter
Prof. dr. ir. F.A. Bilsen, TU Delft, promotor
Prof. dr. ir. A. van den Bos, TU Delft
Prof. dr. ir. T. Houtgast, VU Amsterdam
Prof. dr. ir. A.J.M. Houtsma, TU Eindhoven
Prof. dr. G.J.F. Smets, TU Delft
Dr. ir. J. Raatgever, TU Delft
Dr. ir. D. de Vries, TU Delft

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Salomons, Aurelia Maria

Coloration and binaural decoloration of sound due to reflections / Aurelia Maria Salomons. - [S.l. : s.n.] ;
Delft : Eburon P&L [distr.]
Proefschrift Technische Universiteit Delft. - Met lit.
opg. - Met samenvatting in het Nederlands.
ISBN 90-5651-024-X
Trefw.: klankkleur / akoestiek.

NWO

Dit onderzoek werd gefinancierd door het NWO, stichting Psychon.
voor Bas en onze ouders
Doch wy hebben hier niets ter nedergezet 't gene wy den Leser voor onfeylbaare waarheden willen opdringen, maar sullen ons geerne laten overtuugen van die gene welke met betere redenen, en welgevondeerde experientie, onse fautes konnen aanwijzen.

Menno van Coehoorn (1685)
Contents

List of Notations ........................................................................................................... v

Chapter 1  Introduction to the Perception of Coloration

1.1 Introduction ........................................................................................................... 1
1.2 The definition of coloration .............................................................................. 1
1.3 Reflections and coloration .............................................................................. 2
1.4 Binaural decoloration ...................................................................................... 4
1.5 Concert halls and coloration .......................................................................... 6
1.6 Measurement methods for the strength of coloration ................................. 6
1.7 Aim of the present work ................................................................................ 7

Chapter 2  Coloration of Artificial Signals

2.1 Introduction ........................................................................................................ 9
2.2 White noise with a single repetition: (an)harmonic cosine noise ............... 11
2.3 White noise with multiple repetitions ............................................................ 14
2.4 Results of coloration threshold measurements for a single repetition ........ 16
  2.4.1 Cosine noise thresholds .............................................................................. 17
  2.4.2 Thresholds of harmonic cosine noise compared to literature values .... 19
  2.4.3 Influence of interaural correlation ............................................................. 19
2.5 The relation between pitch and coloration for harmonic and anharmonic cosine noise ................................................................. 20
  2.5.1 The difference in coloration thresholds for uncorrelated harmonic and anharmonic cosine noise ....................................................... 21
  2.5.2 Coloration and pitch strength thresholds compared ............................ 23
Chapter 3  Binaural Decoloration Experiments

3.1 Introduction ........................................................................................................35
3.2 Experiments with harmonic cosine noise with different delay times at left and right ear ............................................................ 37
3.3 Experiments with complementary cosine noise ........................................39
   3.3.1 Experiments with equal delay times at left and right ear .......... 39
   3.3.2 Experiments with different delay times at left and right ear .......... 43
3.4 Coloration thresholds for a dichotic signal with multiple repetitions .......................................................... 44
3.5 Conversion of gain in amplification for dichotic harmonic cosine noise .................................................................................. 47
3.6 Summary and conclusions on binaural decoloration ........................................ 50

Chapter 4  Coloration of Concert Hall Signals

4.1 Introduction ........................................................................................................53
4.2 Paired-comparison of concert hall signals. ............................................... 55
4.3 Coloration thresholds of hall signals ........................................................................ 56
4.4 Binaural decoloration of hall signals ............................................................... 58
4.5 Dependence of coloration perception on location in a concert hall ................. 61
4.6 Conclusions on coloration of concert halls ....................................................... 64

Chapter 5  Criteria for Coloration

5.1 Introduction ........................................................................................................67
5.2 The $A_0$ and $B_0$ coloration criteria .................................................................... 68
Contents

5.3 Frequency analysis by the ear in relation to the perception of coloration ........................................... 70
5.4 The spectral criterion .................................................................................................................. 73
5.5 The autocorrelation criterion .................................................................................................. 75
5.6 Coloration of harmonic cosine noise with long delay times .............................................. 78
5.7 Threshold prediction for signals with multiple reflections .................................................. 78
5.8 Conclusions on criteria for coloration ...................................................................................... 81

Chapter 6  Modelling of Binaural Decoloration

6.1 Introduction .............................................................................................................................. 85
6.2 Physiological background of perception ................................................................................ 86
6.3 The physiology of neural networks for binaural interaction .............................................. 87
6.4 The Central Spectrum model ................................................................................................. 88
6.5 Binaural decoloration modelled with the Central Spectrum model ...................................... 92
  6.5.1 Model for interaurally correlated signals ........................................................................ 92
  6.5.2 Model for interaurally uncorrelated signals .................................................................. 95
6.6 Criteria of coloration applied to concert hall recordings .................................................... 98
6.7 The coloration of uncorrelated complementary cosine noise considered ................................. 100
  6.7.1 Zurek's binaural decoloration model ........................................................................... 100
  6.7.2 'Best ear'-model ............................................................................................................ 102
  6.7.3 Pitch sensation model ................................................................................................. 107
6.8 Conclusions on binaural hearing models .............................................................................. 109

Chapter 7  General Conclusions

7.1 Introduction .............................................................................................................................. 111
7.2 Coloration of artificial signals ............................................................................................... 111
7.3 Binaural decoloration experiments ....................................................................................... 112
7.4 Coloration of concert hall signals ........................................................................................ 113
7.5 Criteria for coloration for dichotic signals .......................................................................... 114
Contents

7.6 Modelling of binaural decoloration ........................................ 115

Appendix A Delay Times of Multiple Repetitions .......................... 117

Appendix B Psycho-physical Methods to Measure Coloration

B.1 Introduction ........................................................................... 119
B.2 Threshold of coloration .......................................................... 119
  B.2.1 The Two Alternative Forced Choice method ....................... 119
  B.2.2 The up/down-strategy ....................................................... 120
  B.2.3 The trajectory method ..................................................... 121
  B.2.4 The adjustment method .................................................. 123
B.3 Scaling of coloration by comparison ....................................... 124
  B.3.1 Paired-comparison ......................................................... 124
  B.3.2 Multidimensional scaling ............................................... 125
B.4 Experimental Environment ................................................... 127

Appendix C Manikin Recordings and Coloration ............................ 129

Appendix D Harmonic Cosine Noise Convolved
  with Auditory Filters .................................................................. 133

Appendix E The Calculation of the ‘Internal’ Autocorrelation ........... 135

References .................................................................................. 139

Summary ...................................................................................... 143

Samenvatting ............................................................................... 145

Acknowledgements ...................................................................... 147

Biography .................................................................................... 148
# List of Notations

<table>
<thead>
<tr>
<th>symbol</th>
<th>quantity</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2AFC</td>
<td>two alternative forced choice</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>amplification: the power of the coloured signal with respect to the total signal in graphs expressed as $20\log(a)$ (a dB)</td>
<td></td>
</tr>
<tr>
<td>$A_0$</td>
<td>spectral criterion for coloration</td>
<td>dB</td>
</tr>
<tr>
<td>ACS</td>
<td>Acoustic Control System</td>
<td></td>
</tr>
<tr>
<td>$B_0$</td>
<td>autocorrelation criterion for coloration</td>
<td></td>
</tr>
<tr>
<td>$c_c(t)$</td>
<td>output peripheral filter</td>
<td></td>
</tr>
<tr>
<td>C(f)</td>
<td>auditory filter</td>
<td></td>
</tr>
<tr>
<td>CAP</td>
<td>Central Activity Pattern</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>Central Spectrum</td>
<td></td>
</tr>
<tr>
<td>$D_0$</td>
<td>CAP criterion - Deviation between a coloured CAP and a flat CAP</td>
<td></td>
</tr>
<tr>
<td>$D_0^\prime$</td>
<td>CAP criterion - Deviation between a coloured CAP and a white CAP</td>
<td></td>
</tr>
<tr>
<td>$\mathcal{F}^{-1}$</td>
<td>inverse Fourier transform</td>
<td></td>
</tr>
<tr>
<td>$\mathcal{F}$</td>
<td>Fourier transform</td>
<td></td>
</tr>
<tr>
<td>$f$</td>
<td>frequency</td>
<td>kHz</td>
</tr>
<tr>
<td>$f_c$</td>
<td>centre frequency</td>
<td>kHz</td>
</tr>
<tr>
<td>$g$</td>
<td>gain of repetition in graphs expressed as $20\log(g)$ (g dB)</td>
<td></td>
</tr>
<tr>
<td>$h(t)$</td>
<td>impulse response</td>
<td></td>
</tr>
<tr>
<td>$j$</td>
<td>imaginary unit: $j^2 = -1$</td>
<td></td>
</tr>
<tr>
<td>$P_{I_c}(t,\tau_i)$</td>
<td>output signal binaural interaction network</td>
<td></td>
</tr>
<tr>
<td>$P(I_c,\tau_i)$</td>
<td>power function for the interaction of left and right ear signals, the Central Activity Pattern</td>
<td></td>
</tr>
</tbody>
</table>
List of Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{\text{min}}(f_c)$</td>
<td>power function for the interaction of left and right ear signals according to the 'best ear'-model</td>
</tr>
<tr>
<td>$P_u(f_c)$</td>
<td>power function for the interaction of left and right ear signals for uncorrelated signals</td>
</tr>
<tr>
<td>$P_z(f_c)$</td>
<td>power function for the interaction of left and right ear signals according to Zurek</td>
</tr>
<tr>
<td>$s(t)$</td>
<td>time signal</td>
</tr>
<tr>
<td>$t$</td>
<td>time</td>
</tr>
<tr>
<td>$T_i$</td>
<td>delay time of repetition $i$</td>
</tr>
<tr>
<td>$W(f_c)$</td>
<td>equivalent rectangular bandwidth</td>
</tr>
<tr>
<td>$w_1(f)$</td>
<td>weighing function for frequency dominance</td>
</tr>
<tr>
<td>$w_2(t_i)$</td>
<td>weighing function for internal delay</td>
</tr>
<tr>
<td>$</td>
<td>H(f)</td>
</tr>
<tr>
<td>$</td>
<td>S(f)</td>
</tr>
<tr>
<td>$</td>
<td>S_c(f_c)</td>
</tr>
<tr>
<td>$\delta(t)$</td>
<td>unit impulse</td>
</tr>
<tr>
<td>$\theta$</td>
<td>autocorrelation time</td>
</tr>
<tr>
<td>$\rho'(\theta)$</td>
<td>revised autocorrelation weighing function (Bilsen, 1968)</td>
</tr>
<tr>
<td>$\rho(\theta)$</td>
<td>autocorrelation weighing function (Bilsen, 1968)</td>
</tr>
<tr>
<td>$\tau_i$</td>
<td>internal delay</td>
</tr>
<tr>
<td>$\varphi(\theta)$</td>
<td>autocorrelation</td>
</tr>
<tr>
<td>$\varphi_c(\theta)$</td>
<td>internal autocorrelation of spectrum</td>
</tr>
<tr>
<td>$\varphi_a(\theta)$</td>
<td>approximation of $\varphi_c(\theta)$</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction to the Perception of Coloration

1.1 Introduction

The subject of the research described in this thesis is the perception of coloration of sound signals. The aim of the research is to model coloration perception and binaural decoloration. The research has been carried out with psycho-physical experiments on human subjects, where the sound signals used in the experiments are either digitally processed synthetic signals or sound signals recorded in concert halls.

In this chapter a definition is given of the colour and coloration of a signal, in relation to the definitions of pitch and timbre as found in the American Standard of Acoustical Terminology (1960). Also, the definition of binaural decoloration is given (§ 1.2). In § 1.3 the mechanism of coloration due to reflections is explained, while in § 1.4 the binaural decoloration is considered. The coloration of sound due to reflections in concert halls is deliberated in § 1.5. This is followed by a brief discussion of the psycho-physical measuring methods used in this research to determine the strength of coloration. Finally, a short outline is given of the contents of the thesis.

1.2 The definition of coloration

The definition of the colour of a sound is given here in analogy with the definitions of pitch and timbre as found in the American Standard of Acoustical Terminology (1960). The definition for the pitch of a sound signal is: "Pitch is that attribute of cochlear sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends primarily upon the frequency of the sound stimulus, but it also depends upon the sound pressure and wave form of the stimulus. (Note: the pitch of a sound may be described by the frequency or frequency level of that simple tone having a specified sound pressure level which is judged by listeners to produce the same pitch.)". The definition for timbre is: "Timbre is that attribute of cochlear sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar." The American Standard of Acoustical Terminology does not contain a definition of 'colour' or 'coloration', therefore a definition is proposed here in analogy with the above expressions. With the definitions of pitch and timbre in mind a definition of the colour and coloration of a sound signal can be given as: "The colour of a sound signal is that
attribute of cochlear sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness are dissimilar"; it thus comprises the timbre, rhythm sensation as well as the pitch of the signal. Note that the colour of a signal can be judged best when comparing it to white noise.

The colour of a signal is a quality of the signal, which may be changed by the surroundings in which the sound is produced or by the apparatus that (re)produces the sound. This change in colour is called coloration: "The coloration of a signal is the audible distortion which alters the (natural) colour of the sound." Examples of everyday coloration is e.g. the hollow quality of sound in a bathroom, or a change in the equaliser setting of the audio amplifier.

Coloration can be perceived best when the coloured signal is compared to the original signal. This is not always available, and therefore coloration is often judged by comparing a signal to an 'internal reference'. For a listener used to natural sound signals, such as music performed in a well-rated concert hall, the music produced by low-quality loudspeakers or in a low-quality concert hall is perceived as coloured; a listener gets accustomed to the coloration of his surroundings, a similar effect applies for timbre (compare Letowski, 1992). To obtain an unbiased judgement of the coloration that has been added by the surroundings, irrespective of any internal reference, white noise is generated in the hall and recorded. The colour of this recorded signal in comparison to white noise can then be considered to represent the coloration added by the surroundings (compare Kuttruff, 1979).

The strength of the coloration depends on whether the listening situation is binaural or monaural (Zurek, 1979). Upon closing one ear the coloration that may be present becomes more distinct, while insofar one has not noticed any coloration, it may be perceived when one ear is closed. This phenomenon is called 'binaural decorrelation', which is therefore defined as: "Binaural decorrelation is a decrease in coloration when listening binaurally compared to monaural listening". The present thesis describes the binaural decorrelation with the aid of the Central Spectrum Model (Raatgever and Bilsen, 1986)

1.3 Reflections and coloration

Coloration is often predominantly the result of reflections. On arriving at the listener the direct signal interferes with the delayed reflections and the resulting signal is coloured with respect to the direct signal: the power of some frequencies of the direct signal is attenuated on interference with the reflections, while for other frequencies an amplification results. The coloration depends on the difference in arrival time between the reflections and the direct signal. The effect of the coloration caused by a single reflection is very prominent when considering white noise to which a repetition is added with a delay time comparable to an early reflection, e.g. between 0.5 and 50
ms. In the white noise signal no pitch can be perceived, but the coloured signal has a very distinct pitch or rattle, which changes with the delay time. The power spectrum of the coloured signal displays a cosine shape, with the distance between the peaks of the cosine being inversely proportional to the delay time of the repetition. This shape can be understood from the impulse response of the system (Fig. 1.1):

\[ h(t) = \frac{(\delta(t) + g\delta(t - T))}{\sqrt{1 + g^2}} \]

1.1

in which \( T \) stands for the delay time and \( g \) for the gain of the repetition. The denominator \( \sqrt{1 + g^2} \) is used to keep the level of the output signal independent of the gain. The Fourier Transform of this impulse response is given by:

\[ H(f) = \frac{(1 + ge^{-j2\pi fT})}{\sqrt{1 + g^2}} \]

1.2

and hence the frequency response (which for white noise input is equal to the power spectrum of the output signal) possesses a cosine dependence on \( f \) and \( T \):

\[ |H(f)|^2 = 1 + \frac{2g}{1 + g^2} \cos(2\pi fT) \]

1.3

In reference to the shape of the power spectrum, white noise to which a single repetition is added is called “cosine noise”. The amplitude of the peaks of the cosine and the ‘modulation depth’, depends on the gain of the repetition.

![Fig. 1.1 Symbolic signal circuit (a) and power spectrum (b) of cosine noise, for \( T=1.0 \text{ ms and } g=1.0 \) (top), and \( T=3.0 \text{ ms and } g=0.3 \) (bottom)](image)

The character of the colour sensation varies with the delay time. For small delay times, smaller than 1.0 ms, the colour is only a high hiss, which becomes lower with increasing delay time. If the delay time approaches 1 ms, the hiss turns more and more into a noisy pitch sensation. For delay times between 1 to 20 ms the pitch is very distinct, and one would be able to play a melody by varying the delay time. This type of pitch is called repetition pitch (Bilsen, 1968; Yost, 1978). When the delay time
is increased beyond 20 ms, the colour perception becomes a rattle sensation, which is sometimes referred to as infra pitch (Warren et al., 1980). With increasing delay time the rattle rhythm becomes slower, turning into a thumping sound, suitably called the 'motor boating effect' (Warren et al., 1980). For delay times of this size, the spectral content of the signal no longer determines the coloration and hence no pitch is perceived. The peaks of the cosine-shaped spectrum are too close together to be resolved by the ear, but one can perceive the time delay of the repetition, in the form of a rhythm sensation. Listening tests have revealed that experienced listeners can perceive the rattle and thumping effects for white noise and its repetition up to a delay time of 80 ms. For speech or music, a repetition with larger delay time (tenth of seconds) is perceived as an echo.

When more than one reflection is present, the reflections can mutually attenuate or intensify. If the reflections are very regularly spaced in time, the peaks in the spectrum of the coloured signal sharpen up and the timbre of the coloured signal becomes more salient. On the other hand, if the reflections are less regularly spaced in time, the pitch sensation wears off and becomes more noise like; although the peaks of the coloured spectrum flatten, the coloration can still be perceived provided that the delay time differences are small, about one ms (see this thesis, Chapter 2).

In addition to the delay time, the strength of the coloration also depends on the gain of the reflection. If the reflection is attenuated with respect to the direct sound, the coloration decreases.

1.4 Binaural decoloration

When listening to a sound in a room with substantial coloration, a noticeable difference occurs when one ear is closed: the coloration becomes worse. Insofar one has not noticed it, upon listening with only one ear the coloration may become clearly present. Generally, the sound becomes more unnatural when listening with only one ear (monaural listening condition) than when listening with two ears (binaural). This decrease of coloration through binaural interaction will be called binaural decoloration. It is most apparent when the direct sound is white noise. The (de)coloration is less when listening to music, for the continuous spectral changes in music reduce the effect to some extent.

In a natural listening situation, the individual ears receive different signals because the ears are placed differently with respect to the source, and the shielding effect of the head generally reduces the sound intensity. In a laboratory situation the left and right signal can be either the same (diotic) or different (dichotic). The definition of 'binaural decoloration' also applies to the dichotic versus the diotic listening situation. When a signal is presented diotically the strength of the coloration increases with respect to the dichotic situation. In this thesis the effect of the 'binaural decoloration' is
tested by comparing the strength of the coloration in diotic and dichotic listening situations.

![Diagram showing different delay times of reflections at left and right ear](image)

*Fig. 1.2 Different delay times of reflections at left and right ear*

A dichotic listening condition that mimics practical situations, is white noise to which repetitions are added that have different delay times for the left and right ear. Consider, for example, the situation that each ear is presented with cosine noise from correlated noise sources, with the repetition delay time at the left ear being 0.5 ms smaller than that at the right ear (compare Zurek, 1979). This delay time difference is comparable to the difference in arrival time of a reflection due to a reflecting surface placed near the left ear (Fig. 1.2).

Besides decoloration, other characteristic properties of dichotic signals (when presented with a headphone) are broadness of the image and lateralisation. The broadness of the signal relates to the interaural correlation. With fully correlated left and right ear signals, the sound perception is a very small image exactly in the middle of the head. When the correlation decreases, the image gets 'broader'. When the correlation is zero, the image fills the head between left and right ear (Chernyak and Dubrovsky, 1968; Potter, 1993). Lateralisation is the shift of the image from the middle of the head to one side. This 'shift' can be created by applying a level or phase difference between the signals of left and right ear. If the intensity of the left ear signal increases (Interaural Intensity Differences, IID), the image will shift to the left ear; when the left signal arrives earlier than the right signal, the image shifts to the left side (Interaural Time Differences, ITD). The effects which interaural correlation, time and intensity differences have on the perceived images has been described by Blauert (1983), Blauert and Lindeman (1986), Keulen, Bilsen and Raatgever (1991) and Potter (1993), among others. One must be aware that these specific aspects of dichotic signals can give unwanted cues when measuring binaural decoloration (see Chapter 3).
1.5 Concert halls and coloration

Reflections play an important role in building acoustics. Many criteria which are used to rate a concert hall, such as the reverberation time, the spaciousness, the clarity and speech intelligibility are based on the relative energy content of the reflections. Reflections are essential to the appreciation of the music: when music is played in an anechoic room, where the walls are covered with absorbing material, the music is 'dry' and boring. On the other hand, when the reflections are too strong and too regularly spaced in time, for instance as a consequence of a strongly reflecting surface, the coloration caused by the reflections will corrupt the music. This form of coloration should be avoided at all costs in a concert hall. The aspect of coloration due to reflections also has great significance for modern acoustic control systems where artificial reflections are added electronically in order to provide a concert hall with electronically adjustable acoustics. The additional coloration can be unacceptable if the delay times of the artificial reflections are not chosen carefully. In this thesis an investigation is made of coloration as a function of the delay times of multiple reflections in order to access this effect (See also Ando and Gottlob, 1979).

1.6 Measurement methods for the strength of coloration

There are several methods to measure the strength of the coloration of a signal, based on listening tests with human subjects. The methods can be divided into two groups. One group of methods aims at finding the threshold of coloration, which is defined as the minimum amount of coloration which is audible when the coloured signal is compared with the non-coloured signal. In all experiments presented here the non-coloured signal is white noise, while the coloured signals are either artificially coloured signals (white noise with added repetitions), or signals recorded in concert halls. The coloured signal is compared to a white noise reference and the subject has to detect a difference. Then, the strength of the coloration is systematically reduced until the level of coloration is reached where just no difference is detected, this is the threshold of coloration. The coloration threshold is expected to be low when the coloration at maximum (i.e. the gain of the repetition is 0 dB) is strong. This method is suitable to provide a quantitative measure of the coloration.

The other group of methods aims to scale the coloration of various signals by mutual comparison of the coloured signals. The signals are compared in pairs and the subject indicates which of the two in his opinion has the strongest coloration. The signals can be scaled from weak to strong coloration. This method is very convenient to compare the quality of coloration of various signals.

The results of the threshold measurements and the paired-comparison give insight in how coloration is valued. The aim of the research is to find criteria with which the subjective assessment of the coloration can be predicted if the spectral content of the
signal is known. These criteria are of great practical use in the design of concert halls, in order to estimate the coloration of the concert hall. The coloration criteria discussed in this thesis are based on an elaboration of the criteria proposed by Atal et al (1962).

1.7 Aim of the present work

As stated in paragraph 1.5, more insight in the role of coloration is required when acoustic control systems are used to design 'virtual' concert halls, and this forms the motivation for the present research. Emphasis will be put on the prediction of the coloration which may occur due to such systems (i.e. the coloration as a result of adding reflections). Since most listeners use both ears when judging a concert hall, realistic coloration models should incorporate the effect of binaural decoloration. In summary, the present research project was initiated with the following aims:

- a theoretical study of the perception of coloration due to repetitions
- provide insight in the coloration found in concert halls
- validation of coloration criteria using artificially coloured signals and concert hall signals
- research on binaural decoloration and development of a model for binaural decoloration based on the Central Spectrum model.

A challenge when measuring coloration is formed by the aspects which comprise the concept of coloration, namely the timbre and the pitch sensation and the rattle sensation. Pitch and timbre-aspects relate to 'spectral coloration', which is the main topic of this study. Rattle appears for large reflection delay times and is referred to as 'time dependent' coloration or infrapitch (Warren, 1980). The question remains as to which aspect a subject uses in the experiments for the assessment of colour. Therefore, the threshold results will be compared to pitch threshold (§ 2.5.2) and to profile analysis (§ 2.8) while in addition multidimensional scaling is applied (§ 4.5) in order to shed some light on this problem.

In Chapter 2 the results are given of experiments with artificially coloured signals, which possess the same spectrum for left and right ear. Chapter 3 describes the experiments with artificial signals concerning binaural decoloration. In Chapter 4 the results of experiments relating to the coloration of concert halls are presented, using signals recorded in nine concert halls in the Netherlands. In Chapter 5 the criteria of coloration are described and tested in relation to the results found in Chapter 2. In Chapter 6 binaural hearing models are discussed and tested on the binaural decoloration experiments of Chapter 3 and 4. Finally, Chapter 7 summarises the major conclusions, and discusses suggestions for further research.
Chapter 2
Coloration of Artificial Signals

2.1 Introduction

Coloration can be perceived if, for instance due to reflections, some frequencies are attenuated and others are intensified. In concert halls this effect can occur through interference between the direct signal coming from the source and the reflections of this signal from the walls, ceilings, etc. The reflected signals are delayed with respect to the direct signal, and often weakened by absorption when reflected. This interaction can be simulated artificially by adding repetitions to a signal, and in electronic acoustic systems this approach is used for generating 'virtual halls'. To avoid unnatural or strong coloration, it is necessary to understand the effect of reflections on coloration. The following parameters of the reflections are of interest: the number of reflections, the delay time, gain and phase with respect to the direct signal, and in dichotic listening situations the interaural correlation and delay time difference between the reflections of left- and right ear signal.

For the case that one reflection is present, so called 'harmonic cosine noise', the strength of the coloration as function of its delay time has been investigated for fully correlated signals by Atal, Schroeder and Kuttruff (1962), Bilsen and Ritsma (1970), Zurek (1979) and for fully uncorrelated signals by Zurek (1979). In § 2.4 these measurements are repeated. Although not mentioned as such in Zurek (1979), there is a difference between the coloration threshold of uncorrelated and correlated harmonic cosine noise, as will be shown in § 2.4.3.

The signal for which the phase of the reflection is in counter-phase with respect to the direct signal, ‘anharmonic cosine noise’, has been considered previously in pitch experiments (Bilsen, 1968; Yost et al, 1978; Raatgever and Bilsen, 1992). Anharmonic cosine noise appears to cause a weaker and more ambiguous pitch sensation than harmonic cosine noise. Section 2.5 will go further into this matter, for both anharmonic cosine noise and comb filter noise (i.e. with a feedback reflection). Based on existing diotic coloration models, the coloration threshold for harmonic and anharmonic cosine noise are expected to be identical, since these models give a relation between the strength of the coloration and the maximum modulation depth of the signal, which is the same for harmonic and anharmonic cosine noise. If, however, a difference in the coloration threshold between harmonic and anharmonic cosine
noise exists, as is found in the experiments, it is expected that there is a relation between the pitch sensation and the coloration of the signals.

In § 2.6 the effect on the coloration threshold is investigated when multiple reflections colour the sound. The number of reflections is varied and the strength of the reflections altered. It is expected that the coloration can be either weakened or intensified by the mutual interaction of the reflections, depending on their delay time.

In § 2.7 a paired comparison is presented, in which the coloured signals are compared directly. The motivation for these experiments is the fact that in halls the coloration is perceived above threshold. With the comparison a coloration scale value is calculated, and correlation is sought between this scale value and the coloration threshold. A negative correlation is expected: if the threshold is low, the (unmasked) coloration is strong and hence a high value on the coloration scale is expected.

In a coloration experiment, the subject compares a coloured signal and a white signal. The strategy with which a subject discerns between both stimuli, may be based on the perceptual information about the stimuli, such as pitch or timbre. Another strategy can be to discern between the difference in shape of both spectra directly. In order to find out whether the subject uses the spectral shape in the coloration experiments, a comparison is made between the coloration thresholds and the thresholds found in experiments in which the subject is forced to use the difference between two spectra only. Such types of experiments are 'profile analysis', on which extensive research has been conducted by Green (Green, 1986; Bernstein and Green, 1987; Green, 1988). In § 2.8 a comparison will be made between the results of profile analysis and coloration thresholds.

In this chapter the research is restricted to signals having the same spectrum for left- and right ear signal. Difference in delay times for left and right reflections can cause binaural decoimation, and this forms the topic of research in Chapter 3.

In the experiments the repetitions are delayed with the aid of a signal processor. For white noise† all frequencies between certain boundaries (say 10 Hz and 10 kHz) are present with the same intensity level. When the interaction of the direct signal and the artificial reflections produces notable coloration, a difference is perceived

† "The purely random process \( \{\xi(t)\} \) has a uniform non-normalized spectrum over the frequency range \( (\sim 0, \infty) \), and is correspondingly referred to as 'continuous white noise'. (...) In physical applications one can use what is termed 'band limited white noise'. This is a process whose spectral density function is constant over a large, but finite, range, and effectively zero outside this range." (Priestley, 1989, page 235)
between the white noise signal and the artificially coloured signal. This effect is used
to measure the coloration thresholds with either an adjustment method (§ 2.4) which
is a fast method to measure many thresholds in a short period, a trajectory method
(§ 2.5) with which small differences between thresholds can be detected, or the
up/down method as described by Levitt (1971), in § 2.6, which is an accurate method
for complex coloured signals. In addition, as mentioned above, in § 2.7 the coloured
signals are compared in pairs far above threshold.

2.2 White noise with a single repetition: (an)harmonic cosine noise

When to white noise one delayed version is added, the resulting signal is called
‘harmonic cosine noise’. This signal has a very distinct coloration, namely a tonal
sensation, referred to as repetition pitch (Bilsen, 1968). The pitch of this tonal sensa-
tion depends on the delay time of the repetition. The spectrum of the signal is cosine
shaped, with peaks at frequencies that are an integer multiple of the reciprocal of the
delay time T. The pitch resembles the pitch of a pure tone with a frequency equal to
the frequency of the first peak of the harmonic cosine noise, f_0 = 1/T and is extracted
from the spectrum around the frequency 4f_0, which is the dominant region (Bilsen,
1977). The tonal sensation is clearest for delay times between 2 ms and 20 ms. For
smaller delay times the coloration becomes weaker and more noise-like (compare
the effect of a simple low-pass filter). For larger delay times the sensation gradually
shifts from a low frequency ‘tone’ to a rhythm sensation, which can be perceived up
to delay times of 100 ms. To give a description of the perception for delay times of
about 40 ms, the term ‘motor-boating’ is sometimes used. (Warren et al, 1980). For
more on the rhythm sensation the reader is referred to Handel (1992) and Kaembach
(1993).

An example of a harmonic cosine noise spectrum is given in Fig. 2.1(top) for a delay
time of 5.0 ms. If the repetition is shifted in phase with respect to the input signal, the
spectrum will be anharmonic. In this thesis anharmonic cosine noise is used to refer
to a phase shift with 180° for all frequency peaks of the signal. This phase shift is
realised by subtracting the repetition from the input signal. The spectrum is shifted by
1/2T with respect to harmonic cosine noise, see Fig. 2.1(bottom).

The sensation of anharmonic cosine noise is also more or less tonal, but the pitch
sensation is not so strong as in the case of harmonic cosine noise, while the tonal
sensation is ambiguous: if the pitch of anharmonic cosine noise is matched to a pure
sine tone, the matching consists two or even three possible pitch perceptions (Bilsen,
Fig. 2.1  Top: harmonic cosine noise, bottom: anharmonic cosine noise, $T=5.0 \text{ ms}$

In the experiments a coloured noise test signal is compared to white noise in a four interval two-alternative-forced-choice (2AFC) method (see Appendix B.2.1). During such experiments the strength of the coloration of the test signal is varied. For cosine noise the coloration strength can be modified by means of adjusting the gain of the repetition, see Fig. 2.2a. Here $n(t)$ is the original noise signal, and $s_g(t)$ the resulting cosine noise test signal. Another method of varying the strength of the coloration is adding uncorrelated white noise, in order to ‘mask’ the colour, see Fig. 2.2b. In this case $n_1(t)$ and $n_2(t)$ represent two uncorrelated noise signals of equal intensity, $s_a(t)$ the resulting test signal.

Fig. 2.2  Alternative strategies for varying the strength of the coloration of the test signals: The gain of the repetition is changed (a) or uncorrelated white noise is added to the coloured signal (b)

The ratio of the power of the coloured signal with respect to the total signal (coloured signal plus masker) is adjusted by means of the amplification ‘$a$’. This amplification is gradually decreased until no difference in coloration can be detected between the in coloration weakened signal and white noise. ‘Masking’ the coloration is a more general method than adjusting the gain of the repetitions, and can also be used for
non-artificial signals such as concert hall recordings. In order to show that the schemes of Fig. 2.2a and Fig. 2.2b can be considered equivalent, the corresponding expressions for the power spectrum are compared. For a given value of the amplification 'a' an equivalent gain 'g' is calculated. For the case that the gain of the repetition is altered, as in Fig. 2.2a, the power spectrum of the output signal is given by (see Eq. 1.3)†

\[
\left| S_g(f) \right|^2 = \frac{2g}{1+g^2} \cos(2\pi f T)
\]

2.1

For the situation of Fig. 2.2b the output time signal belonging to this input is:

\[
s_a(t) = \sqrt{1-a^2} n_1(t) + a \frac{1}{\sqrt{2}} (n_2(t) + n_2(t-T))
\]

2.2

And so the output power spectrum is:

\[
\left| S_a(f) \right|^2 = 1 + a^2 \cos(2\pi f T)
\]

2.3

The amplification 'a' can be written in terms of gain 'g' by equating Eq. 2.1 to Eq. 2.3:

\[
a = \sqrt{2g} \frac{1}{1+g^2}
\]

2.4

and which is depicted in Fig. 2.3.

Fig. 2.3 The gain g as a function of the amplification a if Fig. 2.2a is compared to Fig. 2.2b

† Strictly spoken, the deterministic function S(f) does not exists for white noise n(t). In this case, \( \left| S(f) \right|^2 \) is defined as the Fourier transform of the expectation: \( \left| S(f) \right|^2 = \mathcal{F} \left[ E\{ n(t) n(t+\theta) \} \right] = 1 \)
In most of the literature concerning cosine noise (e.g. Atal et al, 1962; Bilsen and Ritsma, 1970; Yost and Hill, 1978; Zurek, 1979) the definition of the threshold of coloration is expressed in the gain of the repetition in dB. When adding uncorrelated white noise, a more appropriate measure for the threshold is the level of the coloured signal with respect to the overall level, in dB. Since the overall level of the signal is independent of the amount of white noise and equal to 1.0, this threshold definition leads to the amplification ‘a’, given in dB. For a binaural listening situation the equivalence of the two schemes discussed above is only valid under diotic conditions, i.e. when the same harmonic cosine noise is presented to both ears. When harmonic cosine noise signals with different delay times are applied, or when multiple reflections are considered, such an equivalence is generally not present; section 3.5 will further elaborate in this matter.

2.3 White noise with multiple repetitions

In real-life situations, coloration often results from more than only one repetition. Therefore, experiments have been performed with white noise to which multiple repetitions are added (Fig. 2.4). This signal scheme is realised with the aid of a Digital Signal Processor (Ariel, DSP16), which (with standard sampling rate) could cope with up to 16 repetitions.

![Diagram](image)

**Fig. 2.4** White noise \( n_2(t) \) added to its repetitions, with delay time \( T_i \) and gain \( g_i \) 'masked' by uncorrelated white noise \( n_1(t) \)

The number of parameters that can be changed increases significantly as not only the delay times, but also the delay time differences are important. In addition, the level-ratio of the repetitions has influence on the final spectrum. In the threshold detection experiments, the coloration is weakened by adding uncorrelated white...
noise. The spectrum belonging to the scheme of Fig. 2.4 is given in Eq. 2.5, in which the gain of the direct sound is denoted with \( g_0 \) and its delay time \( T_0 = 0 \) ms:

\[
|S(f)|^2 = 1 + a^2 \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} \frac{2g_j}{\sum_{k=0}^{N} g_k^2} \cos \left( 2\pi f \left( T_j - T_i \right) \right)
\]

In case the intensity of each repetition is equal to that of the direct sound, the following applies:

\[
g_i = 1, \quad \sum_{i=0}^{N} g_i^2 = N + 1
\]

The ratio of the total intensity of the repetitions and that of the direct sound equals the number of repetitions:

\[
P_{\text{reflections}} / P_{\text{direct}} = \sum_{i=1}^{N} g_i^2 / g_0^2 = N
\]

For another set of measurements the gain values are chosen such that the sum of the gain values for the repetitions is equal to the gain of the direct sound:

\[
g_0 = 1, \quad g_{i \geq 1} = \frac{1}{N}, \quad \sum_{i=0}^{N} g_i^2 = 1 + \frac{1}{N}
\]

With these gain values the coloration level decreases by increasing number of repetitions, as the total power of the repetitions decreases. The intensity ratio of repetitions and direct sound is 1/N.

A 'set of delay times' refers to a certain ratio of delay times with 1, 2, 4, 8, 12 and 16 repetitions and with all \( g_i \) having the same value. The delay times of the 'sets' in the experiments are chosen such that the repetitions are either regularly or irregularly spaced in time, which give regular or irregular spectra, respectively. With a regular spectrum the delay times have integer ratios. The smallest delay time (difference) occurs just as many times as there are repetitions. In the case of irregular spectra the delay times are chosen from prime numbers, so the delay time ratios are non-integer values. In the experiments, the delay times of irregular spectra are chosen such, that any delay time difference occurs only once for six or a smaller number of repetitions, and twice if there are eight or more repetitions.

For irregular spectra two sets of delay times are studied, one set with small delay times, with differences of about 1 ms and another set with larger delay times, with differences of about 9 ms on the average. For all sets the coloration threshold is measured, and for two sets the coloration is scaled by paired-comparison.

If by applying feedback a reflection with a certain delay is added infinite times to the direct signal, the result is comb filter noise, which has a regular spectrum with very sharp peaks. The power spectrum of comb filter noise to which masking white noise
is added, is given by:

\[ |S(f)|^2 = \left(1 - a^2\right) + a^2 \frac{1}{1 + g^2 - 2g \cos(2\pi f T)} \]

(2.8)

In Fig. 2.5 the signal scheme for producing comb filter noise is given, together with the corresponding power spectrum.

![Signal scheme and power spectrum](image)

Fig. 2.5 Signal scheme (a) and power spectrum of comb filter noise, with \(a=1\), \(g=0.8\) and \(T=2.0\) ms (b)

### 2.4 Results of coloration threshold measurements for a single repetition

This section summarises the results of coloration threshold measurements for white noise signals to which one repetition is added. The stimuli are presented to the subject by means of a headphone. The measurement procedure used is the adjustment method. In this experimental set-up two stimuli are presented, a cosine noise signal and a white noise signal, in random order. The subject can adjust the gain of the repetition of the cosine noise until (just) no difference is detected between both stimuli. The interaural correlation of the white noise signal matches that of the coloured signal, i.e. an interaural (un)correlated coloured test signal is compared to an (un)correlated white noise signal. The duration of a stimulus is 1000 ms and the time between the stimuli is 750 ms. The experiment starts at maximum coloration of the cosine noise (top-down method, see Tab. B.4). The adjustment track is ended after 15 reversals, and the threshold is calculated from the mean of the level points after the fourth reversal. The step size in the gain was taken 2 dB, and the 95% confidence interval of the thresholds is calculated to be 2 dB (halve interval). (see § B.2.4 for more detailed experiment description). The subject is asked to pay attention to the coloration only, when adjusting the gain of the repetition.

In § 2.4.1 the results of correlated and uncorrelated harmonic cosine noise are given (see Fig. 2.6a), i.e. harmonic cosine noise with either correlated or uncorrelated input sources for the left and right ear. Also the thresholds for uncorrelated anharmonic
cosine noise are presented (compare Fig. 2.6b).

Fig. 2.6 Harmonic cosine noise generated from a single source (a) and anharmonic cosine noise generated from two uncorrelated sources (b) with different delay times for left and right ear signal \((T_L, T_R)\)

In § 2.4.2 the results of uncorrelated and correlated harmonic cosine noise are compared to the results found in the literature (of Atal et al, 1962; Bilsen, 1968; Zurek, 1979) and § 2.4.3 gives a survey of the differences between the results for uncorrelated harmonic cosine noise and correlated harmonic cosine noise. In § 2.4.4 the differences between the thresholds of harmonic and anharmonic cosine noise are studied and compared to thresholds of pitch perception.

2.4.1 Cosine noise thresholds

In these experiments, the noise input for the signals presented to left and right ear is correlated. The reference is correlated white noise. The delay time of the cosine noise is the same for both ears, so the only cue the subject can use to find the discrimination threshold is the coloration of the cosine noise. The thresholds have been determined for cosine noise signals with delay times between 2.0 and 20.0 ms, the range where the spectral coloration is present (see § 5.4). Since the research is restricted to spectral coloration, the delay times of the repetitions in most of the experiments are restricted to this interval. For each signal the experiment was repeated three times. Five subjects participated in the experiments.

In Fig. 2.7 the threshold gain is given as a function of the delay time, each dot represents the result of an individual subject (see Tab. B.7 for information on the subjects), while the solid line connects the mean results, averaged over all subjects.
Fig. 2.7  Coloration thresholds of correlated harmonic cosine noise

In the following experiments the thresholds of uncorrelated harmonic and anharmonic cosine thresholds are measured for various delay times. The input sources for the left and right ear signal are uncorrelated white noise sources, and the cosine noise is compared to uncorrelated white noise. Again, equal delay times are used for the left and right ear signal. The experimental set-up and assessment procedure is the same as for the correlated harmonic cosine noise measurements. The thresholds are determined for delay times between 0.5 and 60 ms, the range in which the coloration is perceptible, both spectral coloration and rattle. In these experiments five experienced subjects participated. In Fig. 2.8a and 2.8b the results are presented.

Fig. 2.8  Coloration threshold of uncorrelated harmonic (a) and anharmonic (b) cosine noise

A first observation from comparing Fig. 2.8a and Fig. (b) reveals that the harmonic cosine noise thresholds are somewhat lower (about 3 dB) than the anharmonic cosine noise thresholds. Section 2.5 will go further into this matter.
2.4.2 Thresholds of harmonic cosine noise compared to literature values

Thresholds of correlated harmonic cosine noise were also measured by Zurek (1979) and earlier by Atal, Schoeder and Kuttruff (1962) and Bilsen and Ritsma (1970), and more recently by Lassche (1990). In Fig. 2.9a the mean results (averaged over the subjects) are depicted for each of the investigations measured. The threshold of Atal et al and Bilsen and Ritsma is found at 50% correct responses in detecting difference between white and coloured noise, Zurek used the two-interval 2AFC up/down method (Levitt, 1971), thus 71% correct. Lassche used the trajectory method described in § B.2.3, which gives a threshold at 75% correct responses.

Despite the different experimental set-ups, most of the thresholds agree very well with the present results (difference within ±1.5 dB); only the results from Lassche are on average 4.5 dB lower than the other results, but the variation of the thresholds with the delay time shows the same trend. In Fig. 2.9b the mean results of uncorrelated harmonic cosine noise are given. The results are compared to the results found by Zurek. Again, good agreement is found (average difference: 1.4±2.3 dB).

![Graphs showing coloration thresholds](image)

**Fig. 2.9** Coloration thresholds of correlated (a) and uncorrelated (b) harmonic cosine noise

2.4.3 Influence of interaural correlation

In this section the difference between the thresholds of correlated and uncorrelated harmonic cosine noise is studied in more detail. The differences between the averaged thresholds may not appear to be large in the light of the large scatter among the individual subjects. However, if the difference between correlated and uncorrelated harmonic cosine thresholds is calculated for each individual (Fig. 2.10a) this reveals a quite consistent difference. The average difference is 3.9±1.9 dB and is for all subjects with 95% certainty significant between T=8 and T=20 ms.
In Fig. 2.10b the results averaged over the subjects are compared to the results as measured by Zurek, the error bars in this picture indicate the standard deviation among the results of the subjects. For both cases it is found that the thresholds for uncorrelated harmonic cosine noise are consistently lower than those of correlated harmonic cosine noise.

\[
\begin{align*}
\text{dB} & \quad 10.0 \\
\Delta g & \quad 5.0 \\
& \quad 0.0 \\
& \quad -5.0 \\
& \quad -10.0
\end{align*}
\]
\[
\begin{align*}
T & \quad 10.0 \\
& \quad 20.0 \\
& \quad 30.0
\end{align*}
\]

(a)

\[
\begin{align*}
\text{dB} & \quad 10.0 \\
\Delta g & \quad 5.0 \\
& \quad 0.0 \\
& \quad -5.0 \\
& \quad -10.0
\end{align*}
\]
\[
\begin{align*}
T & \quad 10.0 \\
& \quad 20.0 \\
& \quad 30.0
\end{align*}
\]

(b)

Fig. 2.10 Thresholds from uncorrelated harmonic cosine noise subtracted form those of correlated harmonic cosine noise (a), compared to the results of Zurek (b)

2.5 The relation between pitch and coloration for harmonic and anharmonic cosine noise

This section goes further into the difference between harmonic and anharmonic cosine noise thresholds and the possible relation between the pitch and coloration perception. The coloration thresholds for uncorrelated harmonic cosine noise are lower than that for uncorrelated anharmonic cosine noise (compare Fig. 2.8a and Fig. 2.8b, and Fig. 2.11a). Given the spectra of the signals this difference is remarkable. Since the modulation depth is equal, the coloration strength is expected to be the same for harmonic and anharmonic cosine noise. However, the position of the peaks differ, and the strength of the 'tonal' perception of anharmonic cosine noise is, in the delay time region from 2-10 ms, not as strong as it is for harmonic cosine noise (Yost, 1978; Bremer, 1992; Raatgever and Bilsen, 1992). The question arises whether the pitch strength has played a role in the coloration experiments. In order to investigate this, experiments have been carried out in which the difference in (correlated) harmonic and anharmonic signal thresholds were determined for both coloration and pitch strength (§ 2.5.2).
2.5.1 The difference in coloration thresholds for uncorrelated harmonic and anharmonic cosine noise

To study the difference between the coloration thresholds of uncorrelated harmonic and anharmonic cosine noise in more detail, the differences for the individual subjects are given in Fig. 2.11a. That the threshold for pitch perception need not be equal to the threshold of coloration was shown by Bilsen (1968) for correlated harmonic cosine noise (see Fig. 2.11b).

![Diagram](image)

*Fig. 2.11 Thresholds of uncorrelated anharmonic cosine noise subtracted from those of uncorrelated harmonic cosine noise thresholds (a) and comparison of pitch thresholds and coloration thresholds for correlated harmonic cosine noise (Bilsen, 1968) (b)*

With aid of a 2AFC- method the coloration threshold of harmonic cosine noise was found by comparing to white noise. The thresholds of perceptibility of pitch were determined by comparing harmonic cosine noise with delay time T with harmonic cosine noise with delay time 1.06 T, a pitch shift of 6%, in random order. The subject had to respond whether the pitch shifted ‘up’ or ‘down’. The thresholds were obtained by decreasing the gain of the repetitions until (just) no pitch shift could be perceived between the stimuli. In Fig. 2.11b the coloration and pitch strength thresholds are displayed. For delay times between 2 and 10 ms, the coloration threshold are about 4 dB higher than the pitch thresholds.
Fig. 2.12 Correlated harmonic thresholds subtracted from correlated anharmonic thresholds for coloration (a) and pitch strength of broad band (b) and small band (c) cosine noise (1) and comb filter noise (2) for individual subjects (from Swets 1994)
2.5.2 Coloration and pitch strength thresholds compared

The aim of this section is to verify whether the difference in coloration thresholds for harmonic and anharmonic cosine noise can be attributed to the result of the difference in pitch strength. For that purpose the coloration thresholds and the thresholds for pitch strength were determined for interaurally correlated harmonic and anharmonic cosine noise as well as for comb filter noise (Swets, 1994). To this end the 2AFC trajectory method as described in § B.2.3 was applied. The harmonic and anharmonic signals were presented in one experiment session, so the subject had no knowledge of whether a stimulus pair would be harmonic or anharmonic. In this way, the difference in performance between harmonic and anharmonic signals could be measured very accurately. All experiments were done with one group of subjects, so that potential subject differences were of less concern; four (reasonably) experienced subjects participated in the tests.

Since anharmonic cosine noise has an ambiguous pitch sensation, special care must be taken to find the threshold of pitch for this signal. If the pitch of anharmonic cosine noise is matched to a sinusoidal signal, the pitches found to match are 0.89/T and 1.14/T (Yost et al, 1978; Bilsen, 1968; Raatgever and Bilsen, 1992) while in addition a lower octave can be perceived, namely 0.5/T (Raatgever and Bilsen, 1992; Bremer, 1992). If in a pitch detection experiment anharmonic cosine noise with delay time T is compared to anharmonic cosine noise with 1.06T, the pitch shift perceived by the subject is not necessarily a pitch shift of 6%. If the pitch of the first stimulus is perceived as 0.89/T and the second as 0.89/1.06T, the subject will indeed perceive the second pitch 6% lower than the first pitch and his (correct) answer will be ‘pitch goes down’. However, if the first pitch is perceived as 0.89/T and the second pitch as 1.14/1.06 T, the subject will perceive a pitch shift of 21% up, and will answer accordingly. Another possibility is that in the first stimulus the subject perceives the pitch which corresponds to one octave. If in the second stimulus the octave is not noticed, the answer again will be that the pitch shift between the stimuli goes up. This shows that both possible answers to the 2AFC are valid responses. To reduce this ambiguity effect a pitch shift of 24% is used in the experiments, thus \( T_2 = 1.32 T_1 \) which is sufficient to give a pitch shift down for anharmonic cosine noise for all combinations where no octave is perceived. The subjects are provided with good/false information in the 2AFC-experiments, in order to avoid octave shift answers as much as possible.

The thresholds for perceptibility of pitch and coloration were measured for broadband cosine and comb filter noise (filtered between 20 Hz and 8 kHz) as well as for band filtered cosine and comb filter noise. The aim of the latter measurements is to force the subject to synthetic listening to the pitch generated by only three peaks of the cosine noise (see Fig. 2.13). An advantage is that when the first three peaks are removed from the signal, almost no octave shifts will be perceived (Raatgever and Bilsen, 1992; Bremer, 1992).
Fig. 2.13 Signals which differ in pitch and timbre (a) and signals which differ in timbre only (b)

A disadvantage of this method is that not only the pitch alters between stimulus 1 and 2, but the subject is aided with a timbre shift in the same direction. If the pitch of the signals shifts down, the timbre also shifts down. Since the point of interest for this experiment is the pitch and not the timbre, the timbre is randomised by measuring with different sets of peaks, e.g. the first stimulus with peak 4, 5 and 6, the second stimulus with 5, 6, and 7. Although it is difficult to distinguish between pitch and timbre, all four subjects were capable of hearing the difference after some training (cf. Semal and Demany, 1991).

In Fig. 2.12 the differences in thresholds between (interaurally correlated) harmonic and anharmonic cosine noise are given as measured by Swets. The thresholds are measured by masking the signals (the cosine as well as the comb filter noise) with white noise. The difference between the thresholds is given as $\Delta a$. In Fig. 2.12a the difference in coloration thresholds are depicted, for cosine noise (a1) and comb filter noise (a2). Fig. 2.12b gives the difference between the thresholds for the pitch strength for broad band cosine noise (Fig. 2.12b1) and comb filter noise (b2), and in Fig. 2.12c the results are given for the small band signals. The pitch of small band cosine noise is quite weak for smaller and larger delay times, so in this case fewer thresholds as a function of the delay times were measured.

In Fig. 2.14 the results as found by Swets, averaged over all subjects, are compared to the difference in coloration thresholds described earlier in this section (a) and compared to the difference in pitch thresholds as found by Yost and Hill (1978) (b), both results are expressed in $\Delta a$. The latter thresholds are presumably measured for correlated cosine noise, although the article mentioned is not clear about this. The thresholds are found by detecting difference between the pitch of a stimulus and a pitch of the next stimulus shifted by 10% in a 2AFC up-down method. Since in the experiments of Yost and Hill the subject has to detect the difference between two
stimuli instead of a pitch shift up or down, the direction of the perceived shift is irrele-
vant, and a pitch shift of 10% is sufficient. A slight drawback of their experimental
set-up is that it is not certain whether the subject uses pitch as a cue when detecting
the difference.

Fig. 2.14 The coloration threshold differences (anharmonic minus harmonic) as
found by Swets compared to the results as described earlier in this section
(a) and the broad band pitch threshold differences compared to the results
from Yost (b)

The difference between the coloration thresholds as measured by Swets is in close
agreement with the difference as described earlier in this section (Fig. 2.14a,
average difference 0.57±0.83 dB), and the difference between the pitch thresholds is
comparable to that of Yost (Fig. 2.14b, average difference 1.8±2.0 dB). The
conclusion is that there is a small, but consistent, difference between the harmonic
and anharmonic coloration threshold and that this difference is independent from the
measurement method. If the harmonic / anharmonic difference of coloration
thresholds is compared to the difference in pitch strength thresholds, it appears that
this difference is of the same order: about 3.0±0.5 dB, especially between 2 and 20
ms. Also when the results of cosine noise are compared to those of comb filter noise
(Fig. 2.12), the difference is found to be of the same order as well. So, for both
coloration and pitch threshold, the difference between harmonic and anharmonic is
comparable (respectively 3 and 2 dB). This supports the idea, that when measuring
coloration thresholds of signals with a distinct pitch, the subject finds his threshold
aided by the pitch. This notion is further supported by the observation that the
difference found between the coloration and pitch strength thresholds do not differ
significantly from zero for most of the measurements (see Fig. 2.15 for the results
averaged over all subjects, average difference: about 0.5± 1 dB) These differences
are therefore smaller than in the results found by Bilsen (1968), see also Fig. 2.11b,
where the pitch strength thresholds are lower than the coloration thresholds. If the
absolute thresholds of Bilsen and Swets are compared, it appears that (for T in the
range of 2-10 ms) the pitch strength thresholds are comparable ('a' within ±0.6 dB), but the coloration thresholds for Swets are about a = 2.5 dB lower, and comparable to the results found by Lassche (see Fig. 2.9a). Since Lassche and Swets both used the trajectory method for determining the coloration thresholds, it is possible that this difference is due to the experimental method.

\[\text{Fig. 2.15 Coloration thresholds subtracted from pitch strength thresholds for correlated cosine (a) and comb filter noise (b) as measured by Swets, averaged over individual subjects}\]

### 2.6 Results of coloration threshold measurements for multiple repetitions

The signal of the experiments that are discussed in this section is white noise to which its multiple repetitions are added. In one experiment the influence of the gain of the repetitions is examined, in four other experiments the delay time of the repetitions is varied. The coloration of the signals is weakened by adding uncorrelated white noise. The test signal is compared to white noise in a 2AFC-procedure. The subject is provided with good/false-information. The threshold is found with an adaptive up/down procedure, in which two correct answers will give a level change of 2 dB down and with one incorrect answer the level will go up 2 dB (after Levitt, 1971). The error in the threshold is about 1 dB. In the graphs the different open symbols denote the results of the individual subjects, while the solid line connects the mean threshold averaged over the subjects. In the captions of the figures the delay times of the signal are given by \(T_1\), the first delay time after the direct signal, and \(\Delta T_m\), the mean delay time difference between successive repetitions and the standard deviation of this difference (the exact delay times are given in Appendix A). Along the horizontal axis of the figure the number of repetitions of the signal is given, while the vertical axis is the threshold coloration level, expressed in 'a' (see Eq. 2.5). For the results given in Fig. 2.16a and Fig. 2.16b the delay times are the same.
In Fig. 2.16a the repetition level is $1/N$ of the direct signal (Eq. 2.7), while in Fig. 2.16b each repetition is of equal strength as the direct signal (Eq. 2.6).

**Fig. 2.16** Coloration thresholds, irregularly spaced delay times: $T_1 = 0.94$ ms, $\Delta T_m = 0.99 \pm 0.36$ ms for $g_i = g_0/N$ (a) and $g_i = g_0$ (b)

**Fig. 2.17** Coloration thresholds for multiple repetitions, respectively spaced irregular and regular in time, $T_1 = 8.5$, $\Delta T_m = 8.9 \pm 3.3$ ms (a) and $T_1 = 8.5$, $\Delta T_m = 8.5 \pm 0.0$ ms (b), $g_i = g_0$

In Fig. 2.16a, the total energy in the repetitions is $1/N$ of that of the direct noise, and the decrease of the coloration with increasing $N$ is evident. In Fig. 2.16b the coloration threshold is nearly independent of the number of repetitions, although there is quite a variation in the character of the coloration of the signals. For more than one repetition, the sensation is not longer tonal, but more noise like.

In the next experiments the gain of each repetition is equal to that of the direct noise (Eq. 2.6). In Fig. 2.17a and Fig. 2.17b the delay time differences are about 9 ms. In Fig. 2.17a the results are given for repetitions irregularly spaced in time while in
Fig. 2.17b the results are displayed for regularly spaced repetitions.

Fig. 2.18 Spectra of signals with 8 repetitions spaced irregular in time: $T_f = 8.46$, $\Delta T_m = 8.9 \pm 3.3$ ms (a), and regular in time: $T_f = 8.5$, $\Delta T_m = 8.5 \pm 0.0$ ms (b)

As can be seen, there is a striking difference between the thresholds as a function of the number of repetitions. For the irregularly spaced repetitions, the threshold increases with increasing number of repetitions: the coloration becomes less. For the regularly spaced repetitions the threshold decreases: the coloration becomes stronger with increasing number of repetitions. It can be understood from Fig. 2.18a and Fig. 2.18b that the sharp, regular peaks in spectrum (b) result in a stronger coloration than the irregular peaks in spectrum (a).

2.7 Results of paired-comparison

When the signals are scaled with paired-comparison (see § B.3.1) the coloration of the signals is compared more directly, above threshold. In the tests the subject compares all signals per pair and has to indicate which signal of the two is the 'most coloured' signal. In first instance the subject will indicate the signal with the strongest coloration; if the colour of the two signals is comparably strong, the subject will use other cues, such as the presence of high-pitched colour or rattle. In this way a subjective scale of the 'quality' of the coloration is generated.

Two sets of signals were measured in this way, with the same irregularly spaced delay times. The delay times are the same sets as those of Fig. 2.16, where the coloration thresholds of these signals are given. In Fig. 2.19 the results are shown of the paired-comparison. In this graph N=0 denotes white noise.
Coloration of Artificial Signals

Fig. 2.19 Results of paired-comparison, irregularly spaced delay times: $T_f = 0.94$, $\Delta T_m = 0.99 \pm 0.36$ for $g_i = g_0/N$ (a) and $g_i = g_0$ (b)

For the experiment depicted in Fig. 2.19b one subject does not agree with the other subjects (circle symbol, subject MS). The difference is mainly due to a higher rating of the signals with $N=1$ and $N=2$. If both these signals are removed from the comparison, the subjects all agree. The subject who deviates from the others is the experiment leader, recognising the signals by their coloration which influences the subjective rating. For this subject, the results for $N=1$ and $N=2$ in Fig. 2.19b are comparable to those of Fig. 2.19a; the difference between the signals of Fig. 2.19a and Fig. 2.19b is small for small values of $N$ and this subject rated the signals accordingly. The subjects that where not aware of this connection rated the signals with $g_i = g_0/N$ and $g_i = g_0$ independently.

When the results of the paired-comparison are compared to the coloration thresholds, one must consider that a strongly coloured signal has a high scale value, near to one, and a low threshold. So a correlation between the threshold levels and the scale value should be a negative. For the experiments depicted in Fig. 2.16a and Fig. 2.19a the scale value and thresholds agree: the higher the threshold, the lower the scale value with as minimum white noise ($N=0$). For the experiments from Fig. 2.16b and Fig. 2.19b, only a weak correlation between the thresholds and the scale values is found, the differences between the scale values are larger than expected from the thresholds.

In Fig. 2.20a and Fig. 2.20b the thresholds are plotted against the scale of coloration to show the correlation between the results. The symbols in the graphs denote the individual subjects. The relation between scale value and threshold in Fig 2.19a shows an offset. This is due to the scaling of white noise, which is scaled as zero, while signals with a low threshold, near to zero, are scaled as a value larger than zero. The correlation between the coloration scale and threshold is -0.88. In
Fig. 2.20b there is little variation in the threshold level, so one would expect a small variation in the scale value. In contrast to this the variation in the scale is comparable to the variation in Fig. 2.20a, since the subjects can use the whole range of the scale. Note that the variation in scale level is not caused by random scatter or subject variability, as Fig. 2.19b shows that most of the subjects agree quite consistently on the scaling of the signals. The correlation between coloration scale and threshold is only -0.10.

![Graph](a)

![Graph](b)

**Fig. 2.20** Thresholds depicted against the coloration scale value, for \( g_i = g_0 / N \) (a) and \( g_i = g_0 \) (b)

From the results of Fig. 2.17b one can conclude that when the coloration of the signals is comparably strong, i.e. a relative deviation between the thresholds within 15% as in Fig. 2.16b, the subject still can rate the difference in the quality of the coloration. If the relative deviation is large, such as in Fig. 2.16a which is 82%, the coloration thresholds and the coloration scale will correlate.

### 2.8 Comparison of coloration thresholds and profile analysis

In this section a comparison is made between coloration thresholds and results of 'profile analysis' as given in Green (1986), in which profiles are described which can (more or less) be compared to correlated harmonic cosine noise. Therefore, the 'profile analysis'-results are compared to the coloration threshold of harmonic cosine noise. The motivation for this is to investigate the relevance of the spectral shape for the coloration perception.

The experimental task in the profile analysis is to discern between two discrete broadband spectra, one standard spectrum and the standard spectrum to which a 'signal' is added. In most experiments the standard spectrum is composed of "a set of equal-amplitude sinusoidal components" (Green, 1986). The signal is then formed by adding one or more sinusoids in-phase to components of the standard spectrum.
The level of the signal is adapted until 'just no difference' is detected between standard and standard plus signal. To avoid that the difference is detected by using the absolute intensity as a cue, a roving intensity level is applied to the total sound.

In Fig. 2.21a the standard and in Fig. 2.21b standard plus the signal are depicted, as described in Green (1986). The standard is formed of 21 sines which are spaced evenly on a logarithmic scale, in the form of either a sine or a cosine ripple, between 200 Hz and 5 kHz. The signal gives a modulation of the standard. In Green’s experiments the number of cycles ‘k’ in the sinusoidal variation are altered; the example in Fig. 2.21b the signal exists of a sine-ripple with two cycles (k=2). In Fig. 2.22 the spectrum of harmonic cosine noise is depicted on a logarithmic scale, showing a similar modulation. For an equal amount of 'cycles' in both the profile and the harmonic cosine noise, the delay time of the cosine noise equals k divided by the frequency range (4.8 kHz).

![Graph](image1.png)

Fig. 2.21 Standard profile (a) and standard plus sine profile for the case k=2 (b) as used by Green (1986)

![Graph](image2.png)

Fig. 2.22 Harmonic cosine noise plotted on a logarithmic scale (T=0.42 ms)

There are some important differences between the profile and the cosine noise. Firstly, the discrete profile spectrum versus the continuous cosine noise spectrum, secondly, for the profile the cycles of sinusoidal variation occur on a logarithmic scale and for the cosine noise on a linear scale. Thirdly, the frequency range of the harmonic cosine noise was not restricted between 0.2 and 5 kHz, while the profile was in Green's experiments.
In the profile analysis experiments the level of the signal is adapted in a 2AFC-task, with two-down one-up strategy and yields a 0.707 probability of being correct. The threshold in Green (1986) is given as a signal-to-standard ratio. Since the threshold of harmonic cosine noise is given as the threshold gain of the repetition, the profile thresholds have to be recalculated to an equivalent gain level. In Fig. 2.23 the recalculated profile thresholds, for sine and cosine rippled profiles, are compared to correlated harmonic cosine noise thresholds, as function of the number of cycles \( k \). As the equivalent delay time of harmonic cosine noise threshold for the range of \( k \) values on consideration varies between 0.2 and 2 ms, the harmonic cosine noise thresholds as found by Atal et al. (1962) are used.

Comparison of the thresholds reveals that the cosine noise thresholds are higher than the profile thresholds. Also, the coloration threshold decreases with increasing number of cycles (i.e. increasing delay time). The profile thresholds display no systematic variation with the number of cycles, for neither the sine nor the cosine rippled profile. The sine rippled profile gives an average threshold of \(-20.9\pm0.8\) dB, for the cosine ripple the average profile threshold is \(-20.9\pm3.7\) dB. For \( k=2 \) the coloration threshold is 9 dB higher than the average profile threshold, for \( k=8 \) this is reduced to 4.5 dB.

![Fig. 2.23 Profile thresholds for sine-rippled profile (closed symbols) and cosine-rippled (open diamond) as found by Green (1986) compared to correlated harmonic cosine noise coloration thresholds found by Atal (1962) (solid line)](image)

A possible explanation for the varying coloration threshold is that the subject is aided by the pitch sensation. For harmonic cosine noise the pitch is most distinct when the delay time is about 2 ms, while for smaller delay time the pitch sensation wears off. As a consequence the threshold increases. If the subject uses the profile of the
spectrum only, the threshold does not vary. These results strengthen the surmise that the pitch sensation plays an important role for the coloration perception, more than the spectral shape only. Since the profile and the cosine noise do differ in some number of points, it is recommended to repeat the coloration and profile experiments in future with signals which bear more likeness, e.g. (continuous) modulated noise from which the tops are placed linearly on a logarithmic scale, and filtered in the same range as the profile. Also a roving intensity should be considered in the coloration experiments.

2.9 Summary and conclusions of artificial coloration experiments

For uncorrelated harmonic cosine noise the threshold of coloration depends on the delay time, with a minimum threshold occurring for a delay time of about 2 ms (Fig. 2.8a). The thresholds measured in the present experiments were in accordance with the thresholds found by Zurek (1979), Atal et al (1962) and Bilsen (1968) (Fig. 2.9).

An average difference of $3.9 \pm 1.9$ dB is detected for harmonic cosine noise with correlated and uncorrelated input signal for delay times between 2 and 20 ms: uncorrelated harmonic cosine noise thresholds are lower than correlated harmonic cosine noise (Fig. 2.10).

It is shown that for anharmonic cosine and comb filter noise both the coloration and the pitch strength threshold is about 3 dB higher than that for the corresponding harmonic signals (Fig. 2.12, see also Swets, 1994). The difference in the threshold for the pitch strength corresponds to the experimental results of Yost (Yost, 1978; Yost and Hill, 1979). The amount of difference in harmonic and anharmonic thresholds is comparable for pitch and coloration (maximum 0.5 dB difference, see Fig. 2.15). This result supports the notion that when measuring the thresholds for coloration, the subject is aided by the pitch of the signals. Also, it explains that the coloration of anharmonic cosine noise is less, because the pitch is less distinct. For more on the perception of pitch of complex signals, the reader is referred to Terhard (1974), Gerson and Goldstein (1978) and Meddis and Hewitt (1991).

When more than one repetition is added to white noise, the threshold depends on the number of the repetitions, the relative gain, and on the delay time of the repetitions. If the direct signal dominates over the repetitions, the coloration is less with increasing number of repetitions, for both the quantity (Fig. 2.16a) and the quality (Fig. 2.19a) of the coloration. If the direct signal and repetitions are equally strong, the coloration depends mainly on the delay time differences between the successive repetitions. For repetitions which are irregularly spaced in time with small delay times differences (about 1 ms), the threshold hardly increases as number of the repetitions (Fig. 2.16b), although there is much difference in the quality of the coloration. This
difference in quality is shown by the results of the paired comparison (Fig. 2.19b). For irregularly spaced repetitions with a larger delay time difference, about 9 ms, the threshold increases as a function of the number of repetitions (Fig. 2.17a). When the repetitions are regular, the thresholds decrease as a function of the number of repetitions and the coloration gets more intense (Fig. 2.17b). This can be understood from the spectrum of the signal. If the number of regular spaced repetitions increases, the regular spaced peaks in the spectrum become sharper, and thus the coloration becomes more salient.

In § 2.8 the results of profile analysis were compared with the coloration thresholds. The profile considered, was a sine-formed signal, compared to a flat standard, the coloration threshold were those of harmonic cosine noise. It appeared that the profile threshold do not vary with the number of cycles as harmonic cosine noise thresholds do (Fig. 2.23). It strengthen the surmise that for harmonic cosine noise the thresholds are found with the aid of pitch perception rather than the shape of the spectrum. Since the profile signals and the harmonic cosine noise differ on a number of points, it is recommended to repeat the experiments of profile analysis and coloration with more comparable signals.
Chapter 3

Binaural Decoloration Experiments

3.1 Introduction

In this chapter binaural decoloration is topic of research. If the signals at left and right ear differ, the coloration can decrease with respect to the monotic or diotic listening situation. The question rises, when binaural decoloration will occur and how strong the binaural decoloration is, depending on the difference between the left- and right signal.

To this end, the coloration threshold of harmonic cosine noise with different delay times for left and right repetition is measured and given in § 3.2. These measurements are partly a reproduction of the measurements done by Zurek (1979) on this topic, who gives the threshold of harmonic cosine noise for which repetitions and direct signal of left and right signal stem from one source. As an extension to these tests, the coloration threshold is measured for uncorrelated harmonic cosine noise, for which the left and right ear signal come from uncorrelated sources. The expectation is, that the dichotic threshold will be higher than the diotic threshold. Since the models that are given in Chapter 6 describe that the binaural interaction depends on the interaural correlation of the signals, the prediction is that the amount of decoloration also depends on the interaural correlation.

Another test for the binaural decoloration is found in the measurement of the threshold for complementary cosine noise, a signal whereby harmonic cosine noise is presented to one ear and anharmonic cosine noise to the other ear. Also in this case the experiments are partly a reproduction of the measurements carried out by Zurek (1979), which measured the coloration thresholds for uncorrelated complementary cosine noise. An extension to these experiments is the coloration threshold measured for cosine noise in which the repetitions and direct noise for left and right ear stem from the same source.

It is expected that for complementary cosine noise the minimum of the signal spectrum at one ear may (partially) cancel the maximum of the signal at the other ear, which will diminish the coloration effect. This makes it a very suitable signal for the binaural decoloration test: when the binaural decoloration concept is correct, the threshold value for this listening situation should be significantly higher than for a situation of either harmonic or anharmonic cosine noise. In section 3.3.1 the coloration
thresholds are given for complementary cosine noise with equal delay times left and right and compared to harmonic cosine noise. Since the above assumption is, that there the binaural interaction depends on the interaural correlation, different thresholds are expected for the uncorrelated complementary cosine noise and the complementary cosine noise for which the interaural cross spectrum differs from zero.

In section 3.3.2 'complementary' cosine noise with different delay times for left and right ear signal is considered. When harmonic cosine noise with delay time $T_L$ is presented to one ear and anharmonic cosine noise with a different delay time, $T_R$, to the other ear, the expected cancellation of the coloration of the separate signals is disturbed. Since the cancellation is no longer optimal, the binaural decoloration is therefore expected to decrease. Strictly speaking, the left and right ear signals are no longer complementary, but since harmonic cosine noise is presented to one ear and anharmonic to the other ear, the signals will still be called complementary here for convenience.

In § 3.4 the binaural decoloration is measured of signals with multiple reflections at left and right ear (all stemming from the same source). Since the coloration in concert halls originates from multiple reflections, and the reflections that arrive at left and right ear often differ, this signal tries to mimic a natural situation. For the diotic case the reflections at one ear can weaken or increase the coloration, but in the dichotic case this interaction can also take place interaurally. The expectation is, that binaural decoloration takes place and the dichotic thresholds are higher than the diotic thresholds.

In § 3.5 is shown, that between two coloured dichotic signals having the same spectra, a difference can be detected based on interaural correlation. This difference can be present when dichotic harmonic cosine noise for which the gain of the repetition is altered, is compared to the same, but masked signal. Therefore, one should take care in comparing the thresholds of dichotic harmonic cosine noise expressed in the gain ‘g’ or the attenuation ‘a’.

Since these dichotic signals can give additional (unwanted) cues such as broadening of the image and lateralisation (see section 1.4), special care is taken in the experimental set-up to minimise their effect. For harmonic and complementary cosine noise the adjustment method is used, in which the subject is instructed to pay attention to the coloration only. For measuring the coloration threshold of the multiple reflections signal, the unwanted cues are randomised, so that these cannot be used in the forced choice method applied in this case.
3.2 Experiments with harmonic cosine noise with different delay times at left and right ear

The signals used in this experiment are harmonic cosine noise the repetition delay time at the left ear being 0.5 ms smaller than that at the right ear (compare Zurek, 1979). The cosine noise is generated from equal (correlated) or different (uncorrelated) input sources for left and right ear (compare Fig. 2.6). The delay time difference of 0.5 ms is comparable to the difference in arrival time of a reflection due to a reflecting surface placed near the left ear (Fig. 1.2). The thresholds are measured with the adjustment method as used earlier for (an)harmonic cosine noise with equal delay times at left and right ear (§ 2.4 and § B.2.4). Fig. 3.2 shows the results for all subjects, left for correlated cosine noise and right for uncorrelated cosine noise. The results show that when the delay times of left and right ear signal are equal (Fig. 3.2a), the thresholds are lower than in the situation where the delay times differ (Fig. 3.2b). This observation is valid for both correlated and uncorrelated harmonic cosine noise, although the difference between the thresholds is, for $T$ smaller than 10 ms, slightly smaller for the uncorrelated input sources. Apparently, the absence of interaural correlation weakens the decoloration effect, compare Fig. 3.2c1 with Fig. 3.2c2. This difference is smaller than that between the thresholds of correlated and uncorrelated harmonic cosine noise with the same delay times at left and right ear (Fig. 2.10). Irrespective of the difference between the delay times at left and right ear, the 'uncorrelated' thresholds are lower than the 'correlated' thresholds (compare Fig. 3.2a1 and Fig. 3.2a2 and Fig. 3.2b1 and Fig. 3.2b2).

![Graph](image)

*Fig. 3.1* Correlated harmonic cosine noise thresholds for $T_R = T_L$ subtracted from correlated harmonic cosine noise thresholds for $T_R = T_L + \Delta T$ compared to the results of Zurek, $\Delta T=0.5$ ms

Similar experiments have been carried out by Zurek (Zurek, 1979), for correlated harmonic cosine noise. In Fig. 3.1 the present results are compared to those of Zurek; the difference between the present results and those of Zurek are on average 1.9±2.4 dB, thus there is a good agreement.
Fig. 3.2 Measured coloration thresholds for harmonic cosine noise with equal ($g_{eq}$) (a) and different ($g_{diff}$) (b) delay time for left and right ear, and the difference $\Delta g = g_{diff} - g_{eq}$ (c), generated with correlated (1) and uncorrelated (2) noise input.
3.3 Experiments with complementary cosine noise

This section presents experimental results for both correlated and uncorrelated complementary cosine noise. The thresholds where obtained by comparing the coloured signals to the test signal that is obtained by putting $g=0$ in Fig. 3.3.

![Diagram of correlated and uncorrelated complementary cosine noise]

*Fig. 3.3 Correlated and uncorrelated complementary cosine noise*

In section 3.3.1 the coloration thresholds are given for complementary cosine noise with equal delay times left and right and compared to harmonic cosine noise. In section 3.3.2 'complementary' cosine noise with different delay times for left and right ear signal is given.

3.3.1 Experiments with equal delay times at left and right ear

The thresholds for correlated and uncorrelated complementary cosine noise, with equal delay times, are given in Fig. 3.4b1 and Fig. 3.4b2, respectively. For comparison the thresholds of correlated and uncorrelated harmonic cosine noise are depicted in Fig. 3.4a1 and Fig. 3.4a2. In Fig. 3.4c1 and Fig. 3.4c2 the differences between the thresholds of correlated and uncorrelated harmonic cosine noise and complementary cosine noise are given.

In Fig. 3.4 the individual results of the subjects are depicted as different symbols, while the results averaged over the subjects are connected by a solid line. As can be seen, the thresholds for complementary cosine noise are seen to be much higher than for harmonic cosine noise, for both the uncorrelated and the correlated cosine noise. The difference between the complementary and harmonic thresholds diminishes with increasing delay time, indicating a decrease of the binaural decoloration effect with delay time. This effect is more pronounced for correlated cosine noise than for uncorrelated cosine noise. The coloration of the correlated complementary cosine noise signal decreases as the delay time increases, but other binaural effects, such as perceptual broadening of the source, remain present as an additional cue in comparison to the diotic white noise.
Fig. 3.4  Individual coloration thresholds for harmonic ($g_h$) (a) and complementary ($g_c$) (b) cosine noise and the difference between those thresholds ($g_c - g_h$) (c), for correlated (1) and uncorrelated (2) input noise.
The thresholds for correlated complementary cosine noise therefore show less increase with delay time as might be expected (see Fig. 3.4b1). As a consequence, the difference between correlated complementary cosine noise and harmonic cosine noise decreases rapidly, and may for larger delay times even become negative, when the additional binaural cues outweigh the coloration effect.

For uncorrelated complementary cosine noise the perception is as broad as for uncorrelated white noise, hence broadening does not play a role. Also in this case the difference between harmonic and complementary cosine noise becomes smaller for increasing delay time, but not as fast as for correlated signal thresholds. The decreasing difference for large time delays is simply a result of the overall decrease of the coloration strength for both types of signals.

The coloration of correlated complementary cosine noise has been studied more extensively by Lassche (1990). In the experimental set-up of his experiments the subject was presented with two stimuli, one coloured stimulus and a white noise stimulus, in random order. The subject had to point out which stimulus was the coloured stimulus. The threshold was defined at 75% correct answers. In Fig. 3.5a the results for correlated complementary cosine noise are given, compared to harmonic cosine noise results.

![Graphs showing thresholds of correlated harmonic and complementary cosine noise](image)

*Fig. 3.5 Thresholds of correlated harmonic ($g_h$) and complementary cosine noise ($g_c$) (a) and the difference between those thresholds ($g_c - g_h$) as measured by Lassche (1990) (b) compared to the present results.*

The absolute threshold values found by Lassche are lower than those found in the present experiments, but the threshold differences between complementary cosine noise and harmonic cosine noise are in good agreement for delay times between 2 and 20 ms, as shown in Fig. 3.5b, the average difference is $1.5 \pm 1.6$ dB. In this figure the results are depicted as the mean value, averaged over the subjects; the error bar is the standard deviation among the results of individual subjects.
Fig. 3.7 Individual thresholds for complementary cosine noise with equal \( (g_0) \) (a) and different \( (g_d) \) (b) delay times for left and right ear and the difference between those thresholds \( (g_d-g_0) \) (c), for correlated (1) and uncorrelated (2) noise input.
Binaural Decoloration Experiments

Note the range of negative $\Delta g$ for large $T$, which is probably due to unwanted binaural cues, as discussed previously.

The thresholds for uncorrelated complementary cosine noise may be compared to those reported by Zurek (1979). Fig. 3.6a shows the difference of uncorrelated complementary cosine noise and uncorrelated harmonic cosine noise. Because the difference between the thresholds is given, the difference in experimental set-up plays a minor role. For small delay times, up to 10 ms, the difference in thresholds found by Zurek is somewhat higher, but in general the results agree (average difference: $3.5\pm4.0$ dB)

![Graph](image)

**Fig. 3.6** Harmonic cosine noise thresholds subtracted from complementary cosine noise for uncorrelated input, compared to the results of Zurek (a) and difference between thresholds of uncorrelated and correlated complementary cosine noise (b)

The thresholds found for uncorrelated harmonic cosine noise are consistently lower than those for correlated harmonic cosine noise, as shown in section 2.5.3. On the other hand, the thresholds of correlated and uncorrelated complementary differ not much, on average $0.3\pm1.9$ dB (Fig. 3.6b). As a result, the decoloration, e.g. the difference between harmonic and complementary signals, is slightly larger for uncorrelated input sources than for correlated input sources, compare Fig. 3.4c1 and Fig. 3.4c2.

### 3.3.2 Experiments with different delay times at left and right ear

Another interesting question is how the threshold of complementary cosine noise is affected when the delay times of the repetition left and right are not exactly the same. When the delay times are equal, peaks in the left ear signal spectrum are cancelled by troughs in the right ear spectrum, and the binaural decoloration is expected to be maximum.
Chapter 3

When the peaks of the left ear spectrum are somewhat shifted with respect to the troughs of the right ear spectrum, the cancellation cannot be at optimum, and one would expect less decoloration.

This experiment was carried out for both correlated (Fig. 3.7b1) and uncorrelated (Fig. 3.7b2) complementary cosine noise, with the delay time of the left ear repetition 0.5 ms larger than that of the right ear. The measurement procedure is again the adjustment method, and the subject is asked to concentrate on coloration only. With this type of signal this is rather difficult, since the left and right ear signals differ greatly. However, since the subjects were well trained in recognising coloration, the thresholds were quite consistent for repeated experiments.

The results of this experiment are compared to the thresholds for complementary cosine noise with equal delay times (Fig. 3.7a). As expected, the thresholds of the complementary cosine noise with different delay times are 3-5 dB lower than with equal delay times. The difference decreases with increasing delay time, as for harmonic cosine noise with different delay times left and right and with comparable amount (cf. Fig. 3.2c). Note also that the thresholds for harmonic and complementary cosine noise are comparable for unequal delay times (cf. Fig. 3.2b and 3.7b, the average difference is 0.4±1.2 dB). This indicates that the binaural decoloration effect of complementary with respect to harmonic cosine noise is much reduced when the delay times are no longer equal.

3.4 Coloration thresholds for a dichotic signal with multiple repetitions

In this section coloration thresholds are given for a (interaurally correlated) dichotic signal with four repetitions, as shown in Fig. 3.8. In this scheme \(n_1, n_2\) and \(n_3\) represent three independent white noise sources. The scheme is used for one and four repetitions. The level of each repetition is equal to that of the direct signal, so for harmonic cosine noise \(g_0=g_1=1/\sqrt{2}\) (with the other \(g_i\) zero), while for four repetitions for all \(g_i\) applies: \(g_i=1/\sqrt{5}\). The level of the coloured signals and the reference noise is the same in all situations, 60 dB (SL), so the subject cannot discriminate the signals by comparing the intensity level. The threshold of coloration is measured by masking the signal with white noise, in the scheme of Fig. 3.8 controlled by the parameter 'a'. The masking white noise signal is considered for two noise sources, \(n_2\) and \(n_3\), in order to be able to vary the interaural correlation, as is explained further below. The masked signal is compared to white noise in a two alternative forced choice method. According to the scheme: when the switches are set to position 1, (masked) coloured noise is produced, for reference white noise the switches are set to position 2.

The values of the delay times for the signal with four repetitions are given in Tab. 3.1. The difference of the delay time of the first repetitions is 0.1 ms, whereas for each following repetition the delay time difference is increased with 0.2 ms. The difference
in delay times is chosen such, that the first repetition arrives first at the right ear, the second repetition arrives first at left ear and so forth. This keeps the place of the percept of the signal near the middle. If all repetitions would arrive first at the left ear, the percept would be entirely at the left ear and this lateralisation would give an unwanted cue for the forced choice.

Fig. 3.8  Scheme for processing four different repetitions at left and right ear

In the previous section it appeared that for both correlated dichotic harmonic cosine noise and correlated complementary cosine noise the amount of decoloration depends on the delay time (Fig. 3.2c1 and Fig. 3.7c1), while this is less the case for the uncorrelated signals (Fig. 3.2c2 and Fig. 3.7c2). This might be caused by the fact that the subjects use other than coloration cues, especially broadening, for adjusting the threshold. In order to avoid this, the experiments described in this section are performed with a 2AFC-method in which interaural correlation is randomised. Cues other than coloration are randomised so the subject can no longer use these in the forced choice. The randomisation makes the experiment more difficult, since the subject must focus on differences in coloration instead of any difference whatsoever. Therefore, to obtain a good comparison between diotic and dichotic signals, this randomisation was used in all experiments, including the diotic signals. Since it is possible that the detectability of the coloration depends on the interaural correlation (comparable with the Binaural Masked Level Difference, see e.g. Langford and Jeffres, 1964) the randomisation of the interaural correlation of the masking noise is kept between narrow boundaries. The average interaural correlation of the reference and masking noise was chosen more or less equal to the interaural correlation of the coloured signal which was determined before the experiments, by comparing the broadness of the signals perceptually. This was done by a pilot match in which the 2 subjects matched the interaural correlation of white noise to the signals. The randomisation added to the average correlation is ±0.08.

In the signal scheme of Fig. 3.8 the interaural correlation for the masking noise is
equal to the parameter ‘c’.

The experiments were conducted with an four interval 2AFC up/down method (Levitt, 1971; see § B.2.2). In the experiments three experienced subjects took part. The coloration thresholds for the signal with four repetitions are given in Tab. 3.1, for the individual subjects. In the third column of this table the thresholds are displayed for the dichotic signal, in the first and second column the thresholds are given for the situation that the left, respectively right, signal is presented diotically.

**Tab. 3.1 Coloration threshold of diotic and dichotic signals with four repetitions**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Diotic Left/Left a (dB)</th>
<th>Diotic Right/Right a (dB)</th>
<th>Dichotic Left/Right a (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>-7.97 ± 0.44</td>
<td>-8.71 ± 0.37</td>
<td>-11.73 ± 0.77</td>
</tr>
<tr>
<td>BO</td>
<td>-6.19 ± 0.64</td>
<td>-7.77 ± 1.02</td>
<td>-8.55 ± 0.82</td>
</tr>
<tr>
<td>WK</td>
<td>-6.67 ± 0.72</td>
<td>-7.40 ± 0.85</td>
<td>-7.55 ± 0.48</td>
</tr>
</tbody>
</table>

Because the experimental set-up differs from that of section 3.2, the experiment is also carried out for one repetition. The thresholds as a function of the delay time are depicted in Fig. 3.9. In Fig. 3.9a the thresholds for diotic signals, in (b) the results for the dichotic signals with ΔT=0.5 ms. The standard error is about 0.8 dB. In Fig. 3.9c the difference between the diotic and dichotic thresholds is depicted. For the comparison in Fig. 3.9(d) with the results of section 3.2, equivalent g values have been calculated for the present results although the recalculation between ‘g’ and ‘a’ is not entirely valid for dichotic harmonic cosine noise (see next section).

**Fig. 3.9 Coloration thresholds for** $T_L=T_R$ (a) and for $T_L=T_R+\Delta T$, $\Delta T=0.5$ ms (b)
Fig. 3.9  (continued) Diotic thresholds subtracted from dichotic threshold (c), and this difference recalculated to equivalent gain level, compared to the results of section 3.2 (d)

From the thresholds given in Tab. 3.1 it appears that for the signal with four repetitions the thresholds of the dichotic signals are lower than that of the diotic signals, while on ground of the binaural decoloration a higher threshold would have been expected. For the signal with one repetition, from which the results are depicted in Fig. 3.9, the differences between the dichotic and diotic thresholds are much smaller than the results described in section 3.2 (see Fig. 2.9(d), on average $3 \pm 1$ dB difference). The diotic thresholds Fig. 3.9a are comparable with the results found in section 3.2, but the dichotic thresholds Fig. 3.9b are lower.

For both experiments the dichotic thresholds of the subjects display a larger scatter than the diotic thresholds. This is an indication that the experiment is difficult. On inquiry after the experiment, it appeared that the random interaural correlation distracts the attention from the coloration. The conclusion is therefore, that for measuring decoloration this experimental set-up is less suitable. It remains a challenge to measure the (de)coloration of a dichotic signal in the presence of cues like broadening of the image and lateralisation.

3.5 Conversion of gain in amplification for dichotic harmonic cosine noise

This section shows another example of the existence of other cues, besides coloration, in dichotic listening situations. To this end, dichotic harmonic cosine noise of which the gain of the repetition is weakened (Fig. 3.10a), is compared to the same signal, but masked with white noise (Fig. 3.10b). For each ear individually, both listening situations are equivalent, i.e. the signal spectra are identically, when (see section 2.2):
\[ a^2 = \frac{2g}{1+g^2} \]

This means that, when the input sources of the left and right ear signals were uncorrelated, or when the left and right ear delay times were the same, no difference would be perceived between both situations. However, when the delay times differ and the input sources are correlated, the cross spectrum of the left- and right ear signals is different (if 'g' and 'a' are smaller than 1.0). This difference can be perceived by the subject in the form of a broadening of the image.

Fig. 3.10  Dichotic harmonic cosine noise with correlated input signal, the repetitions are attenuated (a) or white noise added to coloured signal (b)

If the gain of the repetition is weakened, the cross spectrum of the left- and right ear signal is:

\[ (S_{LR}(f))_g = \left( \frac{1}{1+g^2} \right) + \left( \frac{g}{1+g^2} \right) \left( e^{j2\pi f_{T_L}} + e^{-j2\pi f_{T_R}} \right) + \left( \frac{g^2}{1+g^2} \right) e^{-j2\pi f(T_R-T_L)} \]

If on the other hand the dichotic harmonic cosine noise is masked, the cross spectrum is:

\[ (S_{LR}(f))_a = \left( \frac{1-a^2}{2} \right) + \left( \frac{a^2}{2} \right) \left( e^{j2\pi f_{T_L}} + e^{-j2\pi f_{T_R}} \right) + \left( \frac{a^2}{2} \right) e^{-j2\pi f(T_R-T_L)} \]

Eq. 3.3 can be made identical to Eq. 3.2 with substitution of Eq. 3.1, for \( T_L = T_R \), or for \( g = a = 1.0 \). When in Eq. 3.3 'a' is substituted by 'g' according to Eq. 3.1, the difference between the cross spectra is:

\[ (S_{LR}(f))_g - (S_{LR}(f))_a = m(g) \left( 1 - e^{-j2\pi f(T_R-T_L)} \right) \]
where the multiplying factor \( m(g) \) is given by:

\[
m(g) = \left( \frac{-g^2 + g}{1 + g^2} \right)
\]

and depicted in Fig. 3.11a as a function of \( g \).

\[ 3.5 \]

Fig. 3.11 The difference between the cross spectra for dichotic harmonic cosine noise for which the gain is weakened and 'masked' dichotic harmonic cosine noise

A listening experiment has been conducted to investigate the difference between schemes of Fig. 3.10a and Fig. 3.10b. For the scheme of Fig. 3.10b, a certain level ‘g’ is taken, and the equivalent parameter ‘a’ is calculated with Eq. 3.1, so the spectra of left (right) ear signal are equal for both schemes. Then the two signals are compared in a 2AFC-procedure. The threshold as a function of ‘g’ is determined with a trajectory method (see § B.2.3). The experiment was carried out for different delay times for left and right ear. Three subjects participated in these tests.

The duration of a stimulus was 750 ms, the silent interval between two stimuli 250 ms, the silent interval between the sets 500 ms. In Fig. 3.12 the percentage of correct answers is given as a function of gain ‘g’. In Fig. 3.12a the delay time left is 4.0 ms, and right the delay time is 4.5 ms, in Fig. 3.12b the delay times are respectively 4.0 and 8.0 ms. The percentage of correct answers does not depend on the difference of the delay times between left and right ears, and for both signals a difference is clearly audible between -20.0 and -10.0 dB, the percentage of correct answers being close to 100%. For \( g=0 \) dB the schemes are identical, namely correlated harmonic cosine noise, while for gains smaller than -25.0 dB, both schemes reduce to correlated white noise.

The difference that is perceived between the two signals is a difference in ‘broadness’ of the percept; if there is any difference in coloration, it is too weak to be
noticed. Note that the value of the delay time difference has hardly any effect on the results. The thresholds calculated from the results depicted in Fig. 3.12a and Fig. 3.12b both are: $g_1 = -4.3 \pm 0.2$ dB, $g_2 = -23.6 \pm 0.4$ dB.

![Graphs showing percentage correct vs. gain for different delay times](image)

Fig. 3.12 The percentage of correct answers as a function the gain when comparing dichotic harmonic cosine noise 'masked' with white noise and with equivalent gain level of the repetition, delay times $T_L = 4.0$ ms, $T_R = 4.5$ ms (a), delay times $T_L = 4.0$ ms, $T_R = 8.0$ ms (b)

If these results are related to the difference between the cross spectra (the only difference that is present between the two signals), the differences at the thresholds are (Eq. 3.5): $m(g_1) = 0.17 \pm 0.01$ and $m(g_2) = 0.06 \pm 0.01$. One might expect that both thresholds should lead to a single value, independent of the value of ‘g’. That the threshold depends on the level ‘g’ probably results from the nature of the cue. The broadening of the percept is more clear when the stimulus is nearly white (g near -25 dB), than when the stimulus is very coloured (g near 0 dB). The coloration distracts the attention from the broadening of the percept and accordingly the threshold is higher than expected.

### 3.6 Summary and conclusions on binaural decoloration

This chapter discusses the decoloration of signals with the same spectrum for left and right ear signal with respect to signals with different spectra for left and right ear signal. For three types of signals the decoloration effect with respect to harmonic cosine noise is considered (with either correlated or uncorrelated input sources for the left and right ear signal):

- applying a delay time difference between the left and right ear signal (harmonic cosine noise with different delay times)
- complementary cosine noise with equal delay times
- a combination of both: ‘complementary’ cosine noise with different delay times
Binaural Decoloration Experiments

All three configurations act to decrease the spectral similarity between the left and right ear signals, and hence, are expected to result in a binaural decoloration effect. The results of the comparison are shown in Fig. 3.13a for correlated input signals, and in Fig. 3.13b for uncorrelated input signals. The symbols represent the mean difference, averaged over the subjects. As can be seen, the maximum decoloration occurs for uncorrelated complementary cosine noise. This is a consequence of the thresholds for uncorrelated harmonic cosine noise being lower than the correlated harmonic cosine noise thresholds, while the values of the thresholds of correlated and uncorrelated complementary cosine noise are comparable, see Fig. 3.6b. Also, the binaural decoloration strongly decreases if the left and right signals are no longer exactly complementary, that is, if harmonic cosine noise with delay time $T$ is presented to the right ear and anharmonic cosine noise with $T + \Delta T$ to the left ear.

![Graphs showing decoloration effects](image)

Fig. 3.13 The difference of dichotic signals with respect to harmonic cosine noise, averaged over the subjects, for correlated noise input (a) and uncorrelated noise input (b). Solid symbols: harmonic cosine noise with different delay times, open triangles: complementary cosine noise, open diamonds: ‘complementary’ cosine noise with different delay times, with $\Delta T=0.5$ ms.

Binaural decoloration is also observed for harmonic cosine noise signals if the delay times of left and right repetition differ. A comparison of Fig. 3.13a and Fig. 3.13b reveals that for correlated input signals the decoloration decreases with increasing delay time. This is probably due to cues in the signals other than coloration, such as broadening of the percept of the signal. Even when using the adjustment method, it is difficult for the subjects to pay attention to coloration alone when other cues are stronger than coloration.

For uncorrelated signals the percept is very broad, hence these effects do not occur here and the binaural decoloration decreases less. The decrease of binaural decoloration of uncorrelated complementary cosine noise thresholds can be explained from the fact that for both harmonic cosine noise and complementary cosine noise
thresholds increase with increasing delay time and approach zero threshold.

The decrease of binaural decoloration for harmonic cosine noise with different delay times can also be the result of the relative effect of the delay time $\Delta T$. This delay time is kept constant, so that when the delay time $T$ increases, the ratio $\Delta T/T$ decreases and with it the decoloration: the threshold approaches that of harmonic cosine noise with equal delay times.

The effect of $\Delta T/T$ is also reflected in the opposite trends of the thresholds for harmonic and complementary cosine noise with constant delay time difference $\Delta T$. If for complementary cosine noise the delay times of the left and right ear signal differ, the 'cancellation' of the coloration from the left ear signal by the right ear signal is disturbed. When the relative difference decreases, this 'cancellation' is less disturbed. Hence, for larger delay times the threshold approaches that of complementary cosine noise with equal delay times: the decoloration decreases.

In order to try to avoid that other cues beside coloration were affecting the measurements for interaurally correlated dichotic signals, an experiment was done with a 2AFC-method in which the interaural correlation of the 'masking' noise was randomised. This experimental set-up was applied on a signal with four, respectively one, reflections. For four reflections the threshold measured in the dichotic situation was lower than that for the diotic situation, in contradiction to what was expected on grounds of binaural decoloration. For one reflection the difference between the dichotic and the diotic thresholds was lower than measured with the adjustment method. The measurement method was strenuous for the subjects, since the randomised correlation distracted the attention from the coloration of the signal.

The influence of the interaural correlation for dichotic signals was investigated with an experiment in which only the interaural correlation of a coloured signal was changed. Dichotic harmonic cosine noise of which the gain of the repetition was changed, was compared to the same, but 'masked', signal. The two signals show identical spectra for the left and right ear, only the cross spectrum differed. From the results it appeared that if the gain of the repetition was between -10 and -20 dB, the subjects could easily perceive the difference between both signals, which could be correlated to the strength of the cross spectrum difference.
Chapter 4
Coloration of Concert Hall Signals

4.1 Introduction

In order to measure the strength of coloration of 'realistic' signals, experiments have been carried out with signals that were recorded in concert halls. These signals have also been employed by Potter (1993), for experiments investigating the spaciousness of halls. The reader is referred to his thesis for an extensive account of the equipment and recording methods. A list of the concert halls is given in Tab. 4.1 (from Potter, 1993). The choice of concert halls allows a comparison of a wide variety in shape and size.

The auditorium of the Delft University of Technology is equipped with an Acoustic Control System (ACS), with which the acoustics can be changed (Berkhout, 1988). Without the system active the hall has a short reverberation time, namely 1.2 s, and is suited for speech. To make the hall more suitable for music, extra reflections can be added by means of the ACS-system and the reverberation time increases. One can choose between several settings. The reverberation time can increase up to 7 s and extra early reflections can be added, for example. The setting most used for concerts is setting 4 with a reverberation time of 2.3 s. Recordings have been made without the ACS-system active (AU0) and the ACS-system in setting 4 (AU4).

The stationary noise source that was used for the concert hall recordings produced white noise with a frequency range from 70 to 8000 Hz. The recording was made with an artificial head, KEMAR, standing for Knowles Electronics Manikin for Acoustic Research, which mimics the average human in acoustic response for the free field situation (Burkhard and Sachs, 1975). The signals have been recorded at the rear and at the front of the hall. The locations in the hall where the recordings took place were chosen with respect to the dimensions of the hall, so that the recordings of different concert halls are mutually comparable. The hall signals were filtered with an inverse filter to remove the effects of the ear canal distortion of the artificial head (see Appendix C). During the listening tests the signals are presented with Beyer D770 headphones to subjects seated in a soundproof booth. The signals were judged on coloration by a paired-comparison and a triadic comparison, while also the masked threshold level was determined by an up/down-method.
Tab. 4.1 Concert halls, code in result plots, volume $V$, number of seats $N_A$ and seated area $S_T$

<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
<th>$V$ m$^3$</th>
<th>$N_A$</th>
<th>$S_T$ m$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concertgebouw Haarlem</td>
<td>CGH</td>
<td>10100</td>
<td>1290</td>
<td>815</td>
</tr>
<tr>
<td>Musis Sacrum Arnhem</td>
<td>MSA</td>
<td>7070</td>
<td>1001</td>
<td>762</td>
</tr>
<tr>
<td>De Doelen Concert Hall</td>
<td>DOG</td>
<td>27070</td>
<td>2230</td>
<td>1867</td>
</tr>
<tr>
<td>De Doelen Chamber M. Hall</td>
<td>DOK</td>
<td>4040</td>
<td>604</td>
<td>535</td>
</tr>
<tr>
<td>Concertgebouw Amsterdam</td>
<td>CGA</td>
<td>18700</td>
<td>2206</td>
<td>1135</td>
</tr>
<tr>
<td>Muziekcentrum Vredenburg</td>
<td>MCV</td>
<td>17020</td>
<td>1550</td>
<td>1230</td>
</tr>
<tr>
<td>Dr. Anton Philipszaal</td>
<td>APD</td>
<td>19200</td>
<td>1890</td>
<td>892</td>
</tr>
<tr>
<td>Auditorium TU Delft</td>
<td>±8000</td>
<td></td>
<td>977</td>
<td>610</td>
</tr>
<tr>
<td>ACS-system off</td>
<td>AU0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACS-system on</td>
<td>AU4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In § 4.2 the results are presented of a paired-comparison experiment carried out with the various hall recordings. The focus of interest is whether the strength of the coloration is comparable for all hall recordings, or that some halls show distinctly more coloration. Also it is investigated whether the coloration for the rear position is similar to that in the front position.

In § 4.3 the coloration thresholds of the hall signals are given and compared to the results of the paired comparison. The aim of this is to investigate if the coloration thresholds of the hall recordings provide a suitable measure of the quality of the coloration sensation experienced in the paired comparison.

Section 4.4 reports the thresholds of the hall signals when presented diotically. These thresholds are compared to the dichotic thresholds, in order to is investigate whether binaural decoloration can be measured for these realistic signals.

In § 4.5 the dependence of the coloration on the location in the hall is studied. Otherwise stated, is it possible to speak about the coloration of a hall, or does the coloration change much between different locations. In a triadic comparison the subject has to point out which pair of three recordings is most and which is least dissimilar. The results of this comparison are presented in a two-dimensional colour scale, which is related to the aspects of coloration. Also, the relation to the location of recording is addressed.

Finally, section 4.6 summarises the results and gives the conclusions.
4.2 Paired-comparison of concert hall signals.

In short, the paired-comparison scaling method comprises the relative scaling of a set of signals, based on the results of comparing each signal with every other signal. Further details can be found in Appendix B.3.1. In the paired-comparison experiment, a white noise signal was included for reference. The subject was told that all signals are hall signals and therefore he did not know that one of the signals is white noise. This signal is used as a reference when comparing the scale values. The signals recorded at the front of the halls and those recorded at the rear position were treated as two sets that are judged separately. The results are given in Fig. 4.1a for the signals recorded at the rear position, and Fig. 4.1b for the signals recorded at the front places. On the horizontal axis the code of the concert hall is given (see Tab. 4.1). In the coloration scale, which is plotted along the vertical axis, 0 is least coloured and 1.0 is most coloured. The different symbols denote different subjects (see Tab. B.2) while the solid line connects the mean values averaged over all subjects.

![Coloration of Concert Hall Signals](image)

Fig. 4.1 Results of paired-comparison, ref is the reference, white noise. Results for signals recorded in rear (a) and front (b) position

When the individual scale values of the concert halls signals are compared to the scale value of white noise it is remarkable that some subjects judge white noise more "coloured" than some concert hall signals (front position, symbol + and Δ). Possibly, as white noise signals are seldom encountered in concert halls, this signal is perceived as unnatural and therefore marked as coloured.

The mean results for front and rear positions, averaged over all subjects, are compared in Fig. 4.2a. The average scale values for the rear and front position agree very well. Only the scale value of ‘CGH’ is for the front position notably larger than for the back position.
When Fig. 4.1a and (b) are compared, it is apparent from the scatter of the results that the subjects agree more on scaling the signals recorded at the rear position than when scaling the signals recorded at the front position. In Fig. 4.2b the standard deviation among the subjects is depicted, and for all signals (except AU0) the standard deviation between the subjects is larger for the signals of the front position than for the rear position. A possible explanation is that at the front position, the direct signal is strong and the specific reflections of the halls have less influence on the total signal. The differences between the halls is smaller, and consequently scaling becomes more difficult and therefore shows more scatter.

![Graphs showing coloration and deviation between subjects for rear and front position](image)

**Fig. 4.2** Results for rear and front position, (a) mean value averaged over the subjects, (b) standard deviation between results of individual subjects

### 4.3 Coloration thresholds of hall signals

The threshold of coloration for the hall signals is measured with an up/down-2AFC-method (see § B.2.2). Uncorrelated white noise is added to the hall signal as to 'mask' the coloration. In this 2AFC-experiment three stimuli are given. The first stimulus is a reference consisting of white noise. The following stimuli are white noise and the 'masked' hall signal, in random order. The subject is asked which stimulus differs from the reference. Three correct answers must be given before the masking level increases and with one false answer the masking level decreases, so the threshold level of the up/down-method corresponds to 59% correct responses.

Recognition in the 2AFC procedure by another criterion than coloration, such as the "broadness" or the lateralisation of the signal, must be avoided. To this end, the masking and matching signal is white noise of which the interaural correlation is made perceptually equal to that of the hall signals. Also the loudness level of the white signal at the left and the right ear is made equal to the loudness level of the left
and right hall signal. Because it is difficult to make the correlation exactly equal, a small random variation of the correlation of the matching signals was applied. With this experiment five subjects participated in measuring the thresholds of the signals recorded in rear position (Fig. 4.3a) and for three subjects the threshold of signals recorded in the front position was measured (Fig. 4.3b). In this figure, each symbol represents a subject and the mean threshold (averaged over the subjects) is connected with a solid line. The standard deviation in measuring the threshold of an individual subject (the accuracy of the measuring method) is about 0.7 dB. In Fig. 4.4 the average thresholds for both recording positions are compared.

![Graphs showing coloration thresholds for rear and front position](image)

**Fig. 4.3** Coloration thresholds of hall signals for rear (a) and front (b) position

![Graph showing coloration threshold averaged over subjects](image)

**Fig. 4.4** Coloration threshold averaged over the subjects, for rear and front position

The difference between the thresholds measured in rear and front position is small, about 1±0.5 dB. The coloration threshold is -9±0.6 dB for most of the recordings, exceptions being 'DOK' with a 3 dB higher threshold and 'AU4' with a 4 dB lower
threshold. When comparing the thresholds with the scaling results, a low threshold is expected to correspond to a high scale value (near 1.0) and a high threshold correspond to a low scale value (near 0.0). The extreme values of Fig. 4.4 do not match with the scaling results in Fig. 4.2a. The hall “DOK” has a high threshold, but is judged quite coloured in the paired-comparison, comparable to hall “AU0”. This latter hall, on the other hand, has an average threshold. The lowest threshold is found for hall “AU4”, while the scaling result is smaller or equal to the result of “AU0”. In Fig. 4.5 for each concert hall the average threshold is plotted against the average scale value. The relative standard deviation between the thresholds is 25% for the rear position and 18% for the front position (see Fig. 4.3), so the relative standard deviation between the thresholds is only small. Also for this situation (compare to § 2.7), almost no correlation exist between the paired-comparison and the threshold value, a correlation coefficient of $R=0.07$ and $R=0.04$ respectively.

Fig. 4.5 Correlation between colour scale values and (mean) thresholds, for rear (a) and front (b) position

4.4 Binaural decoloration of hall signals

The signals used in the comparison and coloration threshold experiments, are dichotic signals. The signals of right and left ear are different, as well in physical sense as perceptual, as can be seen in Fig. 4.6 which shows the spectra (smoothed by “convolving” with the auditory filter, see § 5.3) for the left and right ear signal of hall “AU4”, recorded in rear position. The binaural decoloration assumption is that the central processor in the brain uses the different information provided by the left and right ear to diminish the coloration (see § 1.4 and Chapter 6 on this subject). To measure the effect of binaural decoloration, the threshold levels were also measured when the same signal is offered to the left and right ear: the left ear signal to both ears (“left”) or the right ear signal to both ears (“right”). In Fig. 4.7a the difference between the dichotic situation and diotic signal “right” is given, in 4.7b between “dichotic” and “left”, this for signals recorded in the rear position.
In Fig. 4.7 the difference between dichotic and diotic threshold is given as: the threshold of the diotic signals subtracted from the threshold of the dichotic signals. If binaural decoloration is present, lower thresholds are expected for the diotic signals. Hence, the difference is expected to have a positive value. As can be seen in Fig. 4.7, for most of the subjects and for most of the thresholds the value is negative. So for most of the situations no binaural decoloration is measured. This result can be the consequence of the broadness cue, since this cue is not easy to conceal (compare to the results of § 3.4). Nevertheless, more binaural decoloration was expected and certainly not lower dichotic thresholds.
Fig. 4.9  The stress percentage as function of the number of dimensions, as determined with the Kruskal method, for AU0 (a) and AU4 (b)

Fig. 4.10  Individual scaling solution (a), the numbers denote the receiver positions (see Fig. 4.8) and the condition space (b), the initials denote the subjects, for AU0 (1) and AU4 (2)
4.5 Dependence of coloration perception on location in a concert hall

For the experiments discussed in the preceding sections, the hall signals had been recorded in front and rear positions. An interesting question is how the perception in a hall changes from place to place. In this section this subject is addressed by comparing recordings made at various places in a single concert hall. For this purpose recordings were made at nine different receiver positions in the Auditorium of the Delft University of Technology (see Fig. 4.8), with and without the acoustic control system in operation (indicated by AU0 and AU4, respectively).

![Diagram of Auditorium](image)

**Fig. 4.8** Ground-plan of the Auditorium of the Delft University of Technology, with the receiver positions for the recordings

In the listening experiment the subject is presented with three stimuli (triadic comparison), and he has to point out which two of signals of a triad were most and which two signals were least dissimilar. The pair which is the most dissimilar, is allotted 2 points, the pair of least dissimilar stimuli 0 points, and the remaining pair 1 point. In this way a dissimilarity matrix is computed, and analysed with multidimensional scaling (see Appendix B). All signals were amplified when presented so as to make them equal in loudness, because this aspect of the signal will otherwise dominate all other aspects, such as coloration and spaciousness (see Vos, 1994).

The results are analysed with the Kruskal-method (Kruskal, 1964a, 1964b) in order to find an indication about the number of dimensions. The curve of the stress as function of the number of dimensions is given in Fig. 4.9, both for the signals recorded without the ACS system (AU0) and with the ACS system (AU4). According to the curves two or three dimensions would be adequate, since more than three
dimensions give only little extra information. There are two reasons why the results are further analysed in two dimensions, the first is found in the available number of stimuli, the second in the interpretation of coloration. With a number of \( n \) stimuli there are \( n(n-1)/2 \) independent measures. If one chooses a solution with \( r \) dimensions, one can estimate \( n-r \) parameters from these independent measures. A reasonable number of data is two times the number of estimated parameters, so the number of parameters should not exceed \( (n-1)/4 \) (Meerling, 1981). In the present results there were 9 stimuli, so a number of 2 dimensions is recommended. Furthermore, for reasons of interpretation, the coloration is defined as a combination of mainly two aspects, namely pitch and timbre (see § 1.2). An effort will be made to couple these attributes of coloration to the dimensions in the scaling method.

The results are given in Fig. 4.10, which shows the result of individual scaling (INDSCAL, Carroll and Chang, 1970). The condition space, Fig. 4.10b, depicts how each subject values the dimensions. The arc corresponds to the situation that all variance is accounted for. Averaged over all subjects, the signals recorded without the ACS-system (AU0) the variance accounted for is 72\% (dimension 1: 45\% and dimension 2: 27\%), and for those with the ACS system (AU4) 67\% (dimension 1: 46\% and dimension 2: 21\%) is accounted for.\(^\dagger\)

As can be seen from the condition space, the relative weight that is attached to each of the dimensions varies strongly from subject to subject. In situation AU0 (Fig. 4.10b1) subject IV attaches a relative weight of 0.91 to the first dimension and 0.12 to the second dimension, while for subject MF this is 0.37 and 0.86. The average relative value is 0.67±0.21 for the first dimension, and 0.45±0.28 for the second dimension. For AU4 (Fig. 4.10b2) the scatter is less extreme, the relative weight is 0.67±0.14 and 0.43±0.17 respectively. So, on average, the subjects attach a little more weight to the first dimension than to the second.

The most important evaluation of the multidimensional scaling is to interpret the axes of the scatter plot, and, in particular with regard to the interest of the present research, to investigate whether coloration plays a role in these scalings. On interviewing the subjects it appeared that important cues on which the subjects compared the signals were pitch sensation (low / high pitched), other aspects of coloration such as timbre and (possibly) rattle, and, to a lesser extent, spaciousness (broadening of the percept). Pitch is a very dominating characteristic of a sound and is therefore likely to be correlated to one of the dimensions. In order to investigate this, the multidimensional scaling was compared to the results of a paired-comparison, for two of the participating subjects (see Fig. 4.11).

\(^\dagger\) Since the INDSCAL method is metric, the results were not rotated on an axis.
In this experiment a unidimensional scaling is obtained, of which the scale value is near zero for weak or low pitched stimuli and near one for strong or high pitched stimuli. This pitch scaling is found to be correlated with the first dimension of the multidimensional scaling, with a correlation coefficient of 0.85 for subject IV, 0.90 for MS and for the averaged results 0.93, for AU0. For AU4 the correlation coefficient is 0.86 for IV, 0.53 for MS with an average 0.77. This outcome supports the assumption that the first dimension equals the pitch sensation for AU0, while for AU4 this is less certain. This is not surprising since for AU4 the subjects MS and IV attach less weight to the first dimension than was the case for AU0, respectively 0.68±0.17 and 0.84±0.09 (see Fig. 4.10b2).

![Graphs showing pitch perception results](image)

*Fig. 4.11 Paired-comparison for pitch perception, AU0 (a) and AU4 (b), solid dot subject IV, open dot subject MS, line: averaged over both subjects*

The second dimension is suspected to be related to the other aspects of coloration, such as timbre. This is investigated by comparing the differences between the spectra. The spectra of the signals are filtered with the auditory filters (see § 5.3) and normalised with the mean value. Then the standard deviation $d_{ij}$ between the spectra $i$ and $j$ is calculated as:

$$d_{ij} = \sqrt{\frac{\sum_{f} \left( \left| S_C(f)_{\text{norm},i} \right|^2 - \left| S_C(f)_{\text{norm},j} \right|^2 \right)^2}{\Delta f}}$$

4.1

These deviations are then put in a dissimilarity matrix and used to generate an individual scaling scatter plot, as if it were the outcome of a multidimensional scaling experiment. This was done for both left and right ear (as two different 'subjects'). The first (and dominant) dimension of this result was compared to the second dimension of the experimental results (Vos, 1994).
Chapter 4

The correlation between these dimensions was found to be 0.86 for AU0 and 0.40 for AU4. Hence, the conclusion that the second dimension of the original multidimensional scaling experiment equals the coloration (other than the pitch perception) seems likely for AU0, but for AU4 this is less certain. Probably the coloration (and also the pitch sensation) is less distinct with the acoustic control system in setting 4. This is supported by the paired-comparison results as given in §4.2.

Assuming that the above interpretation of the dimensions is correct, now the question is addressed if the results of Fig. 4.10 can be explained with regard to the receiver position (see Fig. 4.8). From the results it can be seen that for AU0 the signals 1, 7 and 8 have a strong pitch perception. The corresponding receiver positions for these signals are in the proximity of a concrete side (or -position 1- front) wall which is known to give strong reflections. A cosine noise-like coloration can be caused by these reflections, and may be the explanation for this result, although this is not directly supported by stimulus 4, which is also positioned near this wall. For signal 3 also a stronger tonal coloration was expected, since the glass rear wall (the operator cabin) also produces strong reflections. Probably the level of the receiver at position 3 is already too low to be reached by these reflections. Receiver position 2 gives more (non tonal) coloration than expected. Receiver position 5 and 9 are weakly coloured, and the results are comparable, which can be explained from the symmetry of the hall.

The results for AU4 and AU0 show much similarity. When comparing both results, it appears that for AU4 stimulus number 2 is shifted in the second dimension with respect to the other stimuli: presumably this stimulus seems to have a less distinct spectrum when the acoustic control system is turned on. Stimulus number 7 is shifted in both dimensions with respect to other stimuli. The conclusion is that mainly the tonal sensation in this particular location decreases when the reflections of the acoustic control system are added to the hall. When studying the results, one must be aware that the definition of coloration comprises both the pitch and the other aspects, such as timbre. When the tonal sensation wears off, generally the coloration as a whole diminishes as well.

4.6 Conclusions on coloration of concert halls

In this chapter, the coloration of a number of concert halls was considered. This was done by comparing recordings made by an artificial head and determining the coloration thresholds of these recordings.

The paired-comparison is in itself a straightforward method to compare the different recordings. Coloration consists of different aspects (the strength of coloration; is a certain pitch perceptible; is the pitch low or high; is a rattle perceptible; etc.), and therefore is it difficult to judge on the strength of coloration only. Therefore only
experienced subjects participated in these experiments. The scale values reveal a weak coloration for the recordings made in the concert halls of Haarlem, Amsterdam and Arnhem. Strong coloration was found for the Auditorium Delft and the chamber music hall of Rotterdam. The scale values for signals measured in the front and in the rear of the concert hall are comparable. An exception is the concert hall of Haarlem, in which the coloration in the front position is much higher than in the rear position.

In addition, coloration thresholds of the dichotic hall signals were measured with a 2AFC-method. To avoid recognition by correlation, the interaural correlation of the reference signal was made approximately comparable to that of the hall signals. Also the diotic coloration threshold was measured. In that case only the left ear signal or only the right ear signal was presented to both ears of the subject. The thresholds found for the dichotic signals are in most cases lower than the diotic thresholds. With this method of threshold measurement no binaural decoloration was detected for the hall signals.

There was hardly any correlation found between the threshold measurements and the scale values of the paired-comparison: the chamber music hall of Rotterdam has a high threshold and a high scale value, although a low scale value would be expected. The threshold of the Auditorium with ACS system is much lower than without ACS system, but the latter has a higher scale value. Since it is difficult to measure the threshold of these (dichotic) signals, the scale values are believed to give a better representation of the coloration than the threshold measurements. If so, the Auditorium benefits to some extent from the ACS-system as far as coloration is concerned. The relative coloration scale value decreases with an average 0.13 points for the rear position and 0.06 points for the front position for setting 4 with respect to 'ACS-system off', while the scale value averaged over all hall recordings is 0.54±0.14 for the rear position and 0.54±0.10 for the front position.

The coloration thresholds are also measured for the recorded signals, when the recordings are presented diotically: the left ear signal to both ears and the right ear signals to both ears. Contrary to what was expected, these thresholds were 1.2 dB higher than the dichotic thresholds. This result can be the consequence of the broadness cue in the dichotic signals, but even then certainly no lower thresholds were expected. The conclusion is, that no binaural decoloration is measured for these recordings.

The results of the multidimensional scaling of signals recorded at different positions in a hall show that the coloration perception varies from place to place (see also Vos, 1994). The most important features with which the subjects compare two stimuli are firstly the tonal percept of the coloration (higher or lower pitch), and secondly the
other aspects of coloration such as timbre. From the individual condition space it appeared that on average, the subjects gave a little more weight to the first (is pitch sensation) than the second dimension (timbre), 0.67±0.2 and 0.44±0.2 respectively.

Some of the results of the multidimensional scaling can be explained from the receiver position (strong reflections give strong tonal sensation). When the results with and without the acoustic control system are compared, it appears that the extremes in coloration diminish when the system is turned on. From the results one can conclude that there is no such thing as 'the coloration of a hall', since the coloration varies with the location in the hall. However, the result of Fig. 4.2 show that these differences are often smaller than the differences among different concert halls.
Chapter 5

Criteria for Coloration

5.1 Introduction

This chapter is devoted to the formulation of the criteria for coloration, with as aim to predict the strength of the coloration (i.e. the coloration threshold). The criteria under consideration are applied to the signal spectrum (or autocorrelation), irrespective of any possible binaural interaction. The criteria are tested on signals which possess the same spectrum for the left and right ear signal.

Two criteria for the perception of coloration have been proposed by Atal, Schroeder and Kuttruff (Atal et al, 1962 and Bilsen, 1968), one spectral criterion (\(A_0\)) and one autocorrelation criterion (\(B_0\)). Essential in the formulation of these criteria is the use of short-time analysis. Because of its non-ideal memory the ear cannot make a true Fourier analysis. Therefore, a weighing must be applied to the repetitions, depending on the value of the delay time (see § 5.2).

Atal et al derived this weighing function from experimentally determined coloration thresholds for different harmonic cosine noise signals. This weighing function provided an adequate mathematical formulation of the short-time analysis by the ear, but did not provide any insight in the corresponding properties of the auditory system, nor did it make a distinction between spectral processing for short delay times versus temporal processing for long delay times.

Therefore in § 5.3 the emphasis is placed on the coloration perception in relation to the frequency analysis by the ear. In this approach the auditory filters play an important role (Patterson and Moore, 1986). In § 5.4 and 5.5 the \(A_0\)-criterion and \(B_0\)-criterion will be reconsidered based on the auditory filtering approach. In these sections it will be shown that these criteria apply only to spectral coloration, in other words coloration as consequence of reflections with short delay times. Section 5.6 refers to models for coloration of harmonic cosine noise with long delay times. Section 5.7 discusses the application of the \(A_0\)- and \(B_0\)-criterion on signals with multiple reflections, and compares the predictions to the measured thresholds. In section 5.8 the conclusions are given.
5.2 The $A_0$ and $B_0$ coloration criteria

The assumption of Atal et al is that the size of the peaks in the spectrum ($A_0$-criterion) or the maximum value for the autocorrelation function ($B_0$-criterion) provides a measure for the coloration of a signal. If the peaks of the spectra or autocorrelation of harmonic cosine noise are considered, it is obvious that the size of the peaks is independent of the delay time of the repetition and depends only on the gain of the repetition. On the other hand, experimental observation reveals that the coloration threshold of harmonic cosine noise, hence the strength of the coloration, changes with the delay time (see Fig. 2.11). This apparent discrepancy was explained by Atal et al from the short-time memory of the ear. Repetitions that arrive early are considered to have a larger effect than repetitions that arrive later. This assumption was elaborated by Bilsen (1968) and can be accounted for by weighing the autocorrelation function:

$$\varphi_s(\theta) = \varphi(\theta) \rho(\theta)$$  \hspace{1cm} 5.1

where $\varphi(\theta)$ is the original (long-term) autocorrelation function, $\rho(\theta)$ the weighing function and $\varphi_s(\theta)$ is called the short-time autocorrelation function. The short time power spectrum is obtained from the Fourier-transform of the short-time autocorrelation:

$$\left| S_s(f) \right|^2 = \mathcal{F} \left\{ \varphi_s(\theta) \right\}$$  \hspace{1cm} 5.2

The $B_0$-criterion as defined by Atal et al states that coloration is perceptible if the ratio of the maximum value of the short-time autocorrelation function for any non-zero delay to its value at zero delay exceeds a certain threshold $B_0$:

$$B_0 = \frac{\varphi_s(\theta)_{\text{max}}}{\varphi_s(0)}, \theta \neq 0$$  \hspace{1cm} 5.3

The definition of the $A_0$-criterion according to Atal et al reads as follows: coloration is perceptible if the level difference between maxima and minima of the short-time power spectrum exceeds a certain threshold $A_0$:

$$A_0 = 10 \log \frac{\left| S_s(f) \right|_{\text{max}}^2}{\left| S_s(f) \right|_{\text{min}}^2}$$  \hspace{1cm} 5.4

The shape of the autocorrelation weighing function $\rho(\theta)$ was derived by Atal et al from experimentally determined coloration thresholds for harmonic cosine noise and is shown in Fig. 5.1a (after Bilsen). The form of the autocorrelation function was assumed to have an exponential decay, the parameters of which were derived from
the coloration thresholds for delay times larger than 10 ms. For smaller delay times the weighing function is extrapolated, with by definition ρ(0)=1. With this weighing function, the A₀ and B₀-definition, and the thresholds of harmonic cosine noise from Atal (Fig. 2.13b), the values found for A₀ and B₀ are: A₀=1.1 dB, B₀=0.063.

Fig. 5.1 Autocorrelation weighing function, (a) according to Atal et al (after Bilsen, 1968), (b) revised weighing function of Bilsen (after Bilsen, 1968)

Note that as the shape of ρ(θ) is fitted to the thresholds of harmonic cosine noise, the A₀- and B₀-criteria provide an adequate description for this type of signals, while furthermore the values of A₀ and B₀ are related, as shown below. For harmonic cosine noise to which uncorrelated white noise is added, the short-time autocorrelation and power spectrum are given by:

\[ ϕ_s(θ) = ρ(0)δ(θ) + \frac{a^2}{2} ρ(T)(δ(θ + T) + δ(θ - T)) \]  

5.5

\[ |S_s(f)|^2 = 1 + a^2 ρ(T) \cos(2πfT) \]  

5.6

where again a² denotes the relative contribution of cosine noise to the total signal (see Fig. 2.2b).

Fig. 5.2 Short-time autocorrelation function (a) and short-time power spectrum (b) of harmonic cosine noise, for T=4 ms and a=-1.5 dB
The autocorrelation weighing function reduces the size of the autocorrelation peaks (Fig. 5.2a) by a factor \( \rho(\theta) \). Identifying the peak in the autocorrelation function, application of the \( B_0 \)-criterion yields for the threshold value of \( a \):

\[
a_{th}^2 = 2B_0/\rho(T)
\]

When applying the \( A_0 \)-criterion, the maximum modulation of the spectrum must be found. As the effect of the weighing function in Eq. 5.6 is independent of the frequency, all harmonics are reduced by the same amount (Fig. 5.2b). For the threshold prediction it does not matter from which harmonic the modulation depth is calculated. The threshold value is:

\[
a_{th}^2 = \frac{1}{\rho(T)} \left( \frac{1}{10(A_0/10)} - 1 \right) / \left( \frac{1}{10(A_0/10)} + 1 \right)
\]

Comparing Eq. 5.7 and Eq. 5.8 it is clear that the \( A_0 \)-criterion can be calculated from the \( B_0 \)-criterion. When applying the criteria to arbitrary signals, the \( B_0 \)-criterion requires multiplying the autocorrelation with \( \rho(\theta) \), and for the \( A_0 \)-criterion convolving the spectrum with the Fourier transform of \( \rho(\theta) \).

For delay times below 10 ms, the \( A_0 \) and \( B_0 \)-criteria fail to predict the correct threshold. If the delay time of the repetition approaches zero, the weighing approaches one, for there is no reason the non-ideal memory of the ear should give a smaller weight to a repetition that arrives earlier. Nonetheless, the coloration threshold increases for smaller delay times, as can be seen from the results, Fig. 2.13b. The minimum threshold is found in the range 2.0 - 10.0 ms and decreases for smaller and larger delay times. Therefore, Bilsen proposed another criterion, the revised \( B_0 \)-criterion, with the modified autocorrelation weighing function \( \rho'(\theta) \) as given in Fig. 5.1b. The shape of the weighing function is fitted to the measurements, but without essentially adapting the reasoning of Atal et al.

Since the integration time of the ear, i.e. the time window, is assumed to be about 20 ms (see for instance: Haas, 1951), the theory of Atal cannot satisfy for short delay times. For short delay times another mechanism is influencing the perception of the coloration, namely the limited frequency analysing capability of the ear. This mechanism is discussed in the next section.

5.3 Frequency analysis by the ear in relation to the perception of coloration

The processing of a sound signal by the human auditory system starts in the cochlea. In this part of the inner ear the signal is analysed with respect to its frequency components. The auditory nerves in the cochlea respond to specific frequencies.
The most accurate frequency analysis would take place if an infinite number of auditory nerves would respond each to its own infinitely narrow frequency band. Of course, this is not the case: an auditory nerve responds to a fairly broad band of frequencies round a certain centre frequency for which the nerve’s response is maximum. This band of frequencies is called the ‘auditory filter’. These auditory filters evidently have a large influence on how the signal spectrum is perceived. Therefore, when attempting to judge a signal on coloration by analysis of its spectrum, the spectrum must be analysed in the same manner as in the ear: by filtering the original spectrum with auditory filters.

The auditory frequency analyser has been a topic of research since Helmholtz (1863). Recent research on the shape of the auditory filter was carried out by Patterson (a review on this work can be found in Patterson and Moore, 1986), who proposed the following shape of the auditory filter:

$$C(f, f_c) = \left(1 + \frac{4|f - f_c|}{W(f_c)}\right) \exp\left(-\frac{4|f - f_c|}{W(f_c)}\right)$$

where \(f\) is the frequency, \(f_c\) the centre frequency and \(W(f_c)\) the equivalent rectangular bandwidth. The equivalent rectangular bandwidth is defined as the bandwidth of a rectangular filter with unit amplitude, having the same area as the auditory filter:

$$\int_{-\infty}^{\infty} C(f, f_c) df = W(f_c)$$

The bandwidth of the auditory filter depends on the centre frequency and increases with increasing centre frequency. According to Patterson, the relation between centre frequency and bandwidth is given by:

$$W(f_c) = (6.23 f_c^2 + 93.39 f_c + 28.52) \times 10^{-3} \quad (W \text{ and } f_c \text{ in kHz})$$

An example of the auditory filter for \(f_c = 2\) kHz is given in Fig. 5.3.

In the cochlea, the incoming spectrum is convolved with auditory filters. To obtain the internal spectrum \(S_c\), the following applies:

$$|S_c(f_c)|^2 = \frac{\int C(f, f_c)|S(f)|^2 df}{\int C(f, f_c) df}$$

In order to compare this internal spectrum of a coloured signal with white noise the internal spectrum of the coloured signal is normalised with the internal spectrum of white noise (denominator of Eq. 5.12). The influence of auditory filters on rippled
spectra is studied by taking harmonic cosine as an example. For harmonic cosine noise with a delay time T (and gain one):

$$|S(f)|^2 = 1 + \cos(2\pi f T)$$  \hspace{1cm} 5.13

Thus the internal spectrum is:

$$|S_c(f_c)|^2 = 1 + \frac{1}{W(f_c)} \int_{-\infty}^{\infty} \left(1 + \frac{4|f-f_c|}{W(f_c)}\right) \exp\left(-\frac{4|f-f_c|}{W(f_c)}\right) \cos(2\pi f T) \, df$$  \hspace{1cm} 5.14

After some calculation the solution for this definite integral as a function of the (centre) frequency $f_c$ and delay time T appears to be (see Appendix D):

$$|S_c(f_c)|^2 = 1 + \left[1+\left[0.5\pi T W(f_c)\right]^2\right]^{-2} \cos(2\pi f_c T)$$  \hspace{1cm} 5.15

The result is that the cosine part of the spectrum is weighed by a function that depends on the delay time T and auditory filter width $W(f_c)$. As the latter increases with the frequency, the modulation depth decreases with increasing frequency. An example for harmonic cosine noise with delay time of 5.0 ms, convolved with auditory filters, is given in Fig. 5.3b.

![Fig. 5.3](image)

**Fig. 5.3**  (a) Shape of an auditory filter, $f_c= 2.0$ kHz, (b) Convolution of harmonic cosine noise ($T=5.0$ ms) with auditory filters
5.4 The spectral criterion

The new definition of the $A_0$-criterion is: “coloration is perceptible if the maximum modulation depth (i.e. the level difference between maxima and minima) of the spectrum convolved with auditory filters exceeds a certain threshold $A_0$.”

This criterion can be used for prediction of coloration due to spectral coloration, for instance caused by reflections with delay time differences smaller than 20 ms. Reflections with delay time differences below 20 ms arrive within the time window of the ear and are processed together as one signal. For delay time differences larger than 20 ms, the reflections produce a rhythm sensation, due to temporal recognition (Warren et al, 1980). The spectral features due to the large delay times are filtered out of the internal spectrum, hence the detection of large delay time differences is no longer a spectral effect. Since the delay time differences for the artificial coloured signals of the experiments described in Chapter 3 seldom exceed 20 ms, the present discussion is restricted to spectral coloration.

The prediction of the coloration threshold with this method is not as straightforward as with the use of the $A_0$-criterion as defined by Atal. The modulation depth of the harmonic cosine spectrum convolved with auditory filters is no longer dependent only on the delay time as is the case for the short-time spectra, but on the frequency as well. The modulation depth of the harmonics decreases with increasing number of the harmonic. The weighing function proposed by Atal depends on the delay time only, as every harmonic in the spectrum has the same modulation depth. This problem is not of much importance for cosine noise, since the modulation depth ratio for different harmonic peaks only results in a different value of $A_0$, as long as for different delay times the same harmonic is taken.

For more than one reflection, the maximum modulation depth is often found at very low frequencies. This problem is dealt with in § 5.7, now the estimation of $A_0$ is restricted to the definition as given above. Another problem is the definition of modulation depth, or the choice of the maximum and minimum in the spectrum. Is the modulation depth of the spectrum the modulation depth of the largest peak or is it calculated from the global maximum and minimum of the spectrum. Again, for one reflection this problem plays no role, since the maximum modulation depth is found for one single peak. From now on, the modulation depth is calculated from the global maximum and minimum of the spectrum. With this definition also the coloration can be predicted for spectra, which differ from white noise but have no distinct peaks, such as pink noise.

The calculation of the coloration thresholds with the $A_0$-criterion proceeds as follows. If the coloration of the signal is ‘masked’ by adding white noise, the spectrum
convolved with auditory filters is:

\[
|S_c(f_c)|^2_a = \frac{1}{W(f_c)} \int_{-\infty}^{\infty} C(f,f_c) \left[ (1-a^2) + a^2 |S(f)|^2 \right] df
\]

With the aid of Eq. 5.10 and Eq. 5.12 this can be written as:

\[
|S_c(f_c)|^2_a = (1-a^2) + a^2 |S_c(f_c)|^2
\]

wherein \(a^2\) is the relative strength of the coloured signal, as introduced in section 2.3.

Accordingly to the \(A_0\)-criterion the following at threshold \(a^2_{th}\) applies:

\[
A_0 = 10 \log \frac{1 + a^2_{th} \left| S_c(f_c) \right|^2_{\max} - 1}{1 + a^2_{th} \left| S_c(f_c) \right|^2_{\min} - 1}
\]

The value of \(A_0\) is calculated from masking level thresholds determined for harmonic cosine noise with different delay times. At the threshold level, which depends on the delay time, the coloration is no longer discernible from white noise. At this level the maximum modulation depth is such, that the \(A_0\)-criterion is satisfied. In this approach the \(A_0\)-criterion is assumed to be constant and independent of the delay time or any other feature of the spectrum than the modulation depth. As was explained in § 5.2, the maximum modulation depth of the convolved spectrum decreases with increasing delay time due to the auditory filters. The threshold should increase to the same extent as the modulation depth decreases so as to keep the \(A_0\)-criterion constant. For every delay time the maximum modulation depth at threshold, \(A(T)\), is calculated. If \(A(T)\) is independent of \(T\) the \(A_0\)-criterion provides a valid description of the coloration threshold level. The average value of \(A(T)\) (averaged over the results for all delay times) is taken as \(A_0\)-criterion and the standard deviation is a measure of the reliability of the criterion. The values for \(A(T)\) can be calculated form Eq. 5.15, Eq. 5.18 and the coloration thresholds for harmonic cosine noise:

\[
A(T) = 10 \log \frac{1 + a^2_{th} \left[ 1 + (0.5 \pi TW(0))^2 \right]^{-2}}{1 - a^2_{th} \left[ 1 + (0.5 \pi TW(1/2T))^2 \right]^{-2}}
\]

Since the difference in threshold of uncorrelated and correlated harmonic cosine noise is significant, the \(A_0\)-criterion was calculated for both uncorrelated and correlated harmonic cosine noise thresholds. The \(A_0\)-criterion calculated from the thresholds of uncorrelated harmonic cosine noise is: \(A_0 = 1.3 \pm 0.3\) dB and for correlated coloured signals is: \(A_0 = 1.7 \pm 0.3\) dB (Fig. 5.4a). The standard deviation of both criteria
is reasonably small. The difference between the criterion calculated from correlated harmonic cosine noise and uncorrelated harmonic cosine noise is apparent, although the values are not inconsistent. Since non-artificial signals are often neither fully correlated nor uncorrelated, the $A_0$-criterion for arbitrary signals is estimated to be about 1.5 dB.

![Graph](image)

Fig. 5.4 Calculation of $A_0$-criteria for correlated and uncorrelated harmonic cosine noise (a) and the prediction of the thresholds (b)

Apart from the standard deviation, another method to test the validity of the criterion is to predict the thresholds with the (average) $A_0$-value. Now the value of the parameter $\lambda_{th}$ in Eq. 5.18 is calculated, while the $A_0$-value is kept constant. In Fig. 5.4b the predicted thresholds (curves) are compared to the measurements, for the uncorrelated and correlated harmonic cosine noise thresholds. These figures reveal a standard deviation from the measured thresholds for small delay times (smaller than 2.0 ms) and for the larger delay times (larger than 20 ms). In the interval where the fit is best, the coloration perception is due to spectral pattern recognition.

5.5 The autocorrelation criterion

A new $B_0$-criterion is based on the autocorrelation $\varphi_c(\theta)$ of the spectrum 'convolved' with the auditory filters, i.e. the Fourier transform of spectrum $|S_c(f)|^2$:

$$\varphi_c(\theta) = \delta(\theta) + a^2 \left( \mathcal{F}^{-1} \left[ 1 + \left[ 0.5\pi T W(\omega) \right]^2 \right]^{-2} \right) \ast \frac{1}{2} \left( \delta(\theta - T) + \delta(\theta + T) \right)$$

where the inverse Fourier transform is denoted with $\mathcal{F}^{-1}$ and the convolution with an asterisk (*). In Fig. 5.5 the spectrum of harmonic cosine noise convolved with auditory filters and its autocorrelation are depicted.

The $B_0$-criterion cannot be derived from $\varphi_c(\theta)/\varphi_0(\theta)$ as was done in the original $B_0$-
criterion, since the autocorrelation peak at $\theta=0$ ms is a delta pulse, while at $\theta=T$ this is a finite peak with a certain broadness. Therefore, the definition is adapted as follows: "Coloration is perceptible if the area of the autocorrelation peak belonging to the delay time (difference) of the most dominant reflection (at say, $\theta_d$ ms) exceeds a certain threshold $B_0$, normalised on the area of the peak at zero delay time".

In other words:

$$B_0 = \frac{\int_{\theta_d-\varepsilon}^{\theta_d+\varepsilon} \varphi_c(\theta) d\theta}{\int_{-\varepsilon}^{+\varepsilon} \varphi_c(\theta) d\theta}$$

wherein the integral boundaries $\pm \varepsilon$ depend on the broadness of the peak.

**Fig. 5.5** Harmonic cosine noise convolved with auditory filters: spectrum $|S_c(\theta)|^2$ (a) and autocorrelation $\varphi_c(\theta)$ (b) for $T=2$ ms and $a=1$

This definition is a generalisation of the definition as posed by Atal: the area of the delta pulse at non-zero delay normalised with the area of the delta pulse at zero delay equals the relative difference.

The $B_0$-criterion for harmonic cosine noise is calculated by taking the integral over the peak at $T$ ms. Since the broadness of this peak depends on the delay time, the integral boundaries are set from $-\infty$ to $\infty$, for convenience. The area of the autocorrelation for harmonic cosine noise as a function of the delay time is (see Appendix E):

$$B(T) = a_{th}^2(T) \left[ 1 + \left[ 0.5\pi T W(0) \right]^2 \right]^{-2}$$

The right hand part of the function will be referred to as the autocorrelation weighing function $\rho''(\theta)$ (see the continuous curve in Fig. 5.6):

$$\rho''(\theta) = \left[ 1 + \left[ 0.5\pi \theta W(0) \right]^2 \right]^{-2}$$

If Eq. 5.22 is compared to the $A_0$-criterion (Eq. 5.19), and the approximation
$W(1/2T) = W(0)$ is applied, then the $B_0$-criterion can be related to the $A_0$-criterion as:

$$A_0 = 10 \log \frac{1 + 2B_0}{1 - 2B_0}$$

This agrees with the corresponding expression found by Atal et al, and the predictions for the thresholds are (almost) the same as for the $A_0$-criterion (Fig. 5.4b). Since the $B_0$-criterion is derived from the spectrum filtered with auditory filters, again this criterion can only be used for spectral coloration. The $B_0$-values calculated from the thresholds for uncorrelated harmonic cosine noise are: $B_0 = 0.072 \pm 0.014$, for correlated harmonic cosine noise: $B_0 = 0.098 \pm 0.011$.

**Fig. 5.6** The autocorrelation weighing function $\rho''(\theta)$ (a), compared to the (normalised) weighing functions $\rho(\theta)$ of Atal (1962) and $\rho'(\theta)$ of Bilsen (1968)

In Fig. 5.6b the weighing function $\rho''(\theta)$ is compared to the weighing functions $\rho(\theta)$ and $\rho'(\theta)$ derived earlier by Atal (1962) and Bilsen (1968), respectively. The weighing functions are normalised with the $B_0$-criterion belonging to the weighing function. The present weighing function $\rho''(\theta)$ is, for delay times larger than 5 ms, comparable to that of Atal, $\rho(\theta)$. For smaller delay times the curve of $\rho''(\theta)$ is less steep. For delay times up to 10 ms, there is a considerable discrepancy between the present weighing function $\rho''(\theta)$ and that of Bilsen $\rho'(\theta)$. This weighing function was based on only a few, relative high thresholds for small delay times. Although the present thresholds do not increase that much for small delay times (see Fig. 2.8), one would expect that $\rho''(\theta)$ shows a top for $T = 3$ ms. But since the function $\rho''(\theta)$ is fairly flat for small delay times, this discrepancy between measurements and model is only small.
5.6 Coloration of harmonic cosine noise with long delay times.

If the delay time increases, the modulation depth of harmonic cosine noise convolved with auditory filters decreases. For delay times of about 25 ms the modulation is so small, that both the $A_0$ and $B_0$ criteria predict no coloration. When listening to these signals, a rattle (rhythm) sensation is perceived. Apparently the time structure in the signal is perceived instead of the spectral structure, so the $A_0$-criterion cannot be used. Since the autocorrelation of a signal depends mathematically on the spectrum of the signal, the $B_0$-criterion is not suitable either. For long delay times another criterion must be defined, based on time structure perception, instead of the spectral structure.

A possible mechanism is that the correlation between the incoming signal (reflection) and the signal in auditory memory (direct signal) is scanned. For small delay times the time difference between direct signal and reflection is too small to detect, and in that case the direct signal and reflection are treated as one signal with a clear spectrum structure. For larger delay times no spectral structure is perceived, but the time difference is clear. Of course, if the time difference between direct signal and reflection becomes too long, no rhythm sensation is perceived any longer. Temporal domain analysis can be used to model the coloration of long delay times, see Warren et al, (1980) and Warren, (1982).

5.7 Threshold prediction for signals with multiple reflections

As a further test of the $A_0$ and $B_0$-criterion on more realistic signals, the predicted thresholds of signals with multiple reflections ($§$ 2.2) will be investigated. The spectrum of signals with multiple reflections convolved with auditory filters is the sum of the separate cosine functions each convolved with auditory filters (Eq. 2.5 and Eq. 5.15 combined), for the signal of Fig. 2.4:

$$\left| S_c(f_c) \right|^2 = 1 + a^2 \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} \frac{2g_ig_j}{\sum_{k=0}^{N} g_k^2} \frac{1}{1 + \left[ 0.5 \pi (T_i - T_j) W(f_c) \right]^2} \cos \left( 2 \pi f_c (T_i - T_j) \right)$$

where $T_i - T_j$ is the delay time difference ($T_0 = 0$ ms) and $g_i$ the amplitude of reflection $i$.

The masking level can be found by determining the maximum and minimum of the unmasked spectrum. Because all cosines add at frequency zero, the maximum is found at this point, independent of the delay time values. Therefore, the spectra are high pass filtered (cut off frequency: 60 Hz). Then overall maximum and minimum of these spectra are taken to predict the threshold, see Fig. 5.7a and Fig. 5.7b. Secondly, to investigate the use of the $B_0$-criterion consider the autocorrelation of spectra with multiple reflections (Fourier transformation of Eq. 5.25):

78
Criteria for Coloration

\[
\varphi_c(\theta) = \delta(\theta) + a^2 \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} \frac{2g_i g_j}{N} \int \left[ \left( 0.5 \pi \left( T_i - T_j \right) W(f_c)^2 \right)^2 \right]^{-\frac{1}{2}} \delta \left( \theta \pm \left( T_i - T_j \right) \right)
\]

With multiple reflections several peaks appear in the autocorrelation function, which can partially overlap, thus complicating the interpretation of the most significant peak.

![Graphs showing autocorrelation peaks](image)

**Fig. 5.7** Internal spectrum for four regular (a) and irregular (b) spaced repetitions

For a given repetition strength the width of autocorrelation peaks increases as the delay time increases, while the maximum of the autocorrelation decreases. With multiple reflections the most dominant autocorrelation peak, which is the autocorrelation peak with the largest area, determines the autocorrelation criterion. Since the autocorrelation peaks overlap, the boundary of the integration has to be chosen carefully, as is depicted in Fig. 5.8a for a signal with four irregular spaced repetitions with equal strength. For this signal the eighth peak is the most dominant peak.

In Appendix E (Fig. E.2) it is explained that for a single repetition the form of the autocorrelation peak is such, that applies:

\[
\int_{-\infty}^{\infty} \varphi_c(\theta) d\theta = c \left( \theta \varphi_c(\theta) \right)_{\text{max}}
\]

This is also a good approximation for multiple reflections. Since it is simpler to find the maximum of \( \theta \varphi_c(\theta) \) than to find the peak with the largest area, the \( B_0 \)-criterion is tested with this function:

\[
B_0 = a^2 (\theta \varphi_c(\theta))_{\text{max}}
\]
In Fig. 5.8b the autocorrelation function multiplied with the autocorrelation time is depicted, and as can be seen, the maximum is found at the eighth peak, and this maximum is (almost) equal to the area of the eighth peak.

Fig. 5.8 Internal autocorrelation of a signal with four reflections of equal strength ($T_1=0.94$, $\Delta T_m=0.99\pm0.36$), with dominant shaded peak

Once the $A_0$ and $B_0$-criterion have been defined for signals with more than one reflection, the coloration thresholds can be predicted. The predictions are compared to the mean threshold, averaged over the subjects. The error bar is the standard deviation between the results of the individual subjects.

In Fig. 5.9 the thresholds predicted with the $A_0$-criterion (solid symbol) and $B_0$-criterion (open symbols) are compared to the measured threshold (cross marks), for the three sets of measurements with multiple repetitions of equal strength (see § 2.6). The predictions are generally satisfying, sometimes the $A_0$-criterion gives a better result (Fig. 5.9a, N=2, 4 and 16), sometimes the $B_0$-criterion (the predictions of Fig. 5.9c). Generally there is no large difference between both criteria. The conclusion is therefore, that for artificially coloured signals, both the $A_0$-criterion and the $B_0$-criterion will give a reasonable prediction of the coloration. In the table (Fig. 5.9d) the t-values are given for the test of the hypothesis ‘the difference between prediction and threshold do not differ significantly’. For Fig. 5.9(c), for the $A0$-criterion this hypothesis is rejected.
Fig. 5.9 Predictions compared to the measured thresholds for multiple reflections with equal strength. $T_1 = 0.9$, $\Delta T_m = 1.0 \pm 0.4$ ms (a), $T_1 = 8.5$, $\Delta T_m = 8.9 \pm 3.3$ ms (b), $T_1 = 8.5$, $\Delta T_m = 8.5$ ms (c) and table for the t-values, the critical value is for $\alpha = 0.05$: $t_{\alpha} = 2.57$ (d)

5.8 Conclusions on criteria for coloration

In this chapter two criteria on coloration are discussed, a spectral criterion ($A_0$) and an autocorrelation criterion ($B_0$). These criteria were defined by Atal et al (1962) and were based on the non-ideal memory of the ear. The reflections were weighed accordingly; reflections that arrive later in time have a lower weighing factor than those arriving early. The shape of the weighing function was based on threshold measurements for harmonic cosine noise. Since the weighing assumption is not correct for small delay times (the memory of the ear, thus its time-window, plays a role above 20 ms, Moore et al, 1988), an attempt has been made to give another explanation for the weighing function, which also gives a better prediction for thresholds of coloration for small delay times. The criteria remain essentially the same, but the weighing
function is based on the non-ideal frequency processing of the ear: the spectra are convolved with auditory filters. The values of the criteria are derived from harmonic cosine noise thresholds and tested by predicting the thresholds of artificially coloured signals.

The new definition of the $A_0$-criterion is related to the maximum-minimum ratio of the spectrum filtered with the auditory filters (internal spectrum). The conclusions are that the $A_0$-criterion gives a good description of the coloration threshold as a function of the delay time for delay times between 2 and 20 ms. The value of the $A_0$-criterion derived from thresholds measured of uncorrelated harmonic cosine noise (1.3±0.3 dB) is smaller than when derived from thresholds of correlated harmonic cosine noise (1.7±0.3 dB). Although this difference is not insignificant† one must realise that signals measured in real-life situation are not fully correlated or fully uncorrelated, for which an average $A_0$-criterion of about 1.5 dB is proposed.

The autocorrelation criterion $B_0$ is related to the area of the most dominant peak of the autocorrelation function of the spectrum filtered with the auditory filters (internal autocorrelation function). The area of the most dominant peak is taken to be proportional to the maximum of the autocorrelation multiplied with the autocorrelation time. This criterion is easier to apply and used to predict the thresholds for coloration. The value of the $B_0$-criterion found with the thresholds for cosine noise is $B_0 = 0.072±0.014$ for uncorrelated and $B_0 = 0.098±0.011$ for correlated harmonic cosine noise. The prediction with this criterion of thresholds of signals with multiple reflections gives satisfying results.

For delay times larger than 20 ms another mechanism is used to judge a signal on its coloration. The spectrum of a signal with (difference in) delay times larger than 20 ms is flat when convolved with auditory filters: the ear can no longer detect the coloration by analysing the spectrum and the $A_0$-criterion predicts no coloration. Also the autocorrelation of the spectrum convolved with auditory filters, multiplied with the autocorrelation time, displays no peaks so the $B_0$-criterion also predicts no coloration. The coloration that is present in the signal has another character than the coloration caused by small delay times. Instead of spectral coloration a rhythm sensation is present. The conclusion is therefore, that the $A_0$- and $B_0$-criterion should only be used for the prediction of spectral coloration.

When using the $A_0$- and $B_0$-criterion one must be aware that these criteria give only a prediction of the coloration threshold of the signal, which is a representation of the

† For 15 degrees of freedom and $\alpha=0.05$ the t-test gives a critical value $t_{0}=2.13$, the t calculated from the measured points is $t=4.43$: the hypotheses 'A_0-criterion for correlated and uncorrelated harmonic cosine noise are equal' is rejected.
Criteria for Coloration

strength of the coloration. The criterion gives no information about the appreciation of the coloration. Two signals can have the same threshold, but can have a distinctly different character, yielding different scale values when comparing the signals in a paired-comparison. Hence, this difference can not be predicted from these criteria. Since the threshold gives important information, applying the $A_0$- and $B_0$-coloration is a first and important step in describing the coloration of a signal.
6.1 Introduction

Coloration is more clearly audible when listening with only one ear than when listening with both ears. The sounds appear to have a more hollow quality and the sensation of the sound is less natural when listening with only one ear. Investigation of this binaural decoloration effect has two major aims. Firstly, this effect must be incorporated for the prediction of the strength of coloration for dichotic signals. When the coloration of a dichotic signal is predicted from a monotic signal, for instance the left ear signal, the coloration is overpredicted in general. Secondly, binaural decoloration can possibly help to further understand the mechanism of the binaural interaction.

In this chapter binaural interaction is studied and binaural interaction models are tested for their suitability to predict the binaural coloration sensation and decoloration. To this end, in § 6.2 a (simplified) scheme is used to describe how the perception emanates from the left and right sound signals. This model is based on the physiology of neural networks, which is elaborated in § 6.3 (after Pickels, 1982). The binaural interaction is described with the Central Spectrum (CS) model in § 6.4 (after Bilsen, 1977; Raatgever, 1980; Raatgever and Bilsen, 1986). In § 6.5 dichotic coloration criteria are defined based on the CS-model, in § 6.5.1 for interaurally correlated signals and in § 6.5.2 for interaurally uncorrelated signals. The criteria are tested by predicting the coloration thresholds for dichotic signals and comparing these to the thresholds obtained from the experiments (Chapter 3). In § 6.6 the criteria are applied to the concert hall recordings, in order to test the criteria on realistic signals.

Since for interaurally complementary cosine noise the coloration prediction does not agree with the experiments (no coloration predicted, though coloration was perceived), this signal is investigated more closely in § 6.7. In § 6.7.1 the model proposed by Zurek (1979) is tested on the results. In § 6.7.2 the internal spectrum for complementary cosine noise is scanned with a sine probe (van Keulen, 1990; Put, 1992). The results from this latter experiment raise the surmise that a subject can concentrate on the ear which gives the most important information. In order to determine whether the subject uses this strategy with the coloration experiments, this
Chapter 6

'best-ear' model is tested with the coloration thresholds for complementary cosine noise. Since this model gave no satisfactory results for the coloration thresholds, in § 6.7.3 a new model is proposed which is based on the pitch sensation which complementary cosine noise generate at left and right ear. Finally, in § 6.8 the conclusions are given for the binaural hearing models.

6.2  Physiological background of perception

The model describing binaural hearing processes is based on results gained from experiments with subjects and is supported by physiological knowledge about the working of the ear. The binaural hearing process is shown schematically in Fig. 6.1.

The cochlear filters, the neural network for binaural interaction and the monaural channels in this model are deduced from physiological knowledge of the human ear, which will be discussed in the next section. The result of the binaural interaction and the information of the monaural channels is transferred to a higher order process, which derives the most important information on the properties of the sound source. With this information a perceptual image of the sound is made, including aspects such as the direction of the source, and spaciousness and coloration. The higher order processes are described with the Central Spectrum model (Raatgever and Bilsen, 1986), which describes how a central processor scans the output of a neural network.

![Diagram](image)

Fig. 6.1  Scheme describing the perception which originates from the left and right ear signal

86
The function of the monaural channels in this model is twofold. Firstly, for binaural listening situations the interaction of the left and right ear signal disturbs the original left and right ear signal. To maintain this information, the monaural information is also given directly to the higher order process. Secondly, it is an extension for the model for monaural listening situations when the higher order process receives information from one monaural channel only.

6.3 The physiology of neural networks for binaural interaction

The physiological mechanism of hearing has been described, among others, by Pickels (Pickels, 1982; Johnson et al., 1990; Dabak and Johnson, 1992). The sound enters the ear by the ear canal, and reaches the cochlea via the eardrum and earbones. In the cochlea the auditory nerves are activated, for each ear separately. The nerves of left and right cochlea which participate in the binaural interaction come together in the so-called Superior Olivary Complex, located in the brain stem. The main nuclei in this complex are the Lateral Superior Olive (LSO) and the Medial Superior Olive (MSO), and the medial nucleus of the trapezoid body (MTB). The MSO is supplied with binaural information via the contra-lateral and ipsi-lateral nerves from the cochlea, the Antero Ventral Cochlear Nucleus (AVCN). The MTB carries information from the opposite cochlea to the ipsi-lateral LSO, and in this way the LSO receives information from both ears, from the ipsi-lateral cochlea (via the AVCN) and the contra-lateral MTB, see Fig. 6.2. The nuclei respond in different ways to the binaural input. Most of the nuclei in the LSO have high-frequency characteristics and are mainly sensitive to frequencies above 3 kHz. The majority of binaural nerve fibres in the LSO are excited by ipsi-lateral stimuli and inhibited by contra-lateral ones; this type of nerves is called E/I-type. With these characteristics, the LSO can respond to a difference in intensity between the ears, defined on a spectral basis. Since level differences give a cue on the direction of the sound source, this binaural interaction can be used for sound localisation. Whether the LSO (the level differences between the right and left spectrum) plays a role in binaural decoloration is discussed in section 6.7 (see also Put, 1992). Most of the nuclei found in the MSO have low frequency characteristics, less than 10 kHz, and are of E/E-type: the units are excited by both contra-lateral and ipsi-lateral stimuli.

From electro-physiological measurements it appeared that for low frequencies (up to 1.5 kHz) the nuclei in the MSO are sensitive for interaural time differences (Pickels, 1982). The time differences are detected as the difference in arrival time for the signals of each ear arriving at a nucleus. The excitation of the nucleus is maximum when the signal from one ear coincides with the delayed signal from the other ear. The difference in time of arrival of the signals is a result of a property of the cells, and each cell has its characteristic delay.
When the signals arrive at the same time at the ears, maximum output is found at the location where the delays of the ipsi-lateral and contra-lateral nerves are equal. When an interaural time difference is present, the maximum output occurs at the nuclei where the difference in the neuronal delays from both ears equals the interaural time difference of the incoming signal. With this property the sound localisation based on interaural time difference is possible. In addition the interaural correlation and phase relation between the left and right signal can be detected, which might offer a basis for an explanation for binaural decoloration. The binaural interaction that takes place in the MSO cells is modelled as the Central Activity Pattern, CAP, which is described in the next section.

![Binaural interaction network: LSO and MSO](after Pickels, 1982)

**Fig. 6.2**  Binaural interaction network: LSO and MSO (after Pickels, 1982)

### 6.4 The Central Spectrum model

The Central Spectrum (CS-) model is based on physiological knowledge and on psycho-physical measurements (Bilsen, 1977; Raatgever, 1980; Raatgever and Bilsen, 1986). The model is based on earlier models as formulated by Jeffres (1948), among others. The CS-model describes the activity of the excitatory-excitative type cells in the MSO as a function of the frequency $f$ and the interaural delay time $\tau_i$. The activity of the cells is 'translated' in the model into power as a function of the frequency for a certain delay time, the Central Activity Pattern (CAP) $P(f, \tau_i)$.

The model assumes that a nerve fibre from a certain peripheral filter of the ipsi-lateral ear interacts with a nerve fibre from the contra-lateral ear of the corresponding peripheral filter. The signal from the contra-lateral ear is delayed, and the signals of the ears add up as illustrated in Fig. 6.3 (Raatgever, 1989).
Modelling of Binaural Decolororation

![Diagram of binaural interaction network]

Eqs. 6.3 The binaural interaction network (After Raatgever, 1989). The input from the right side is delayed with respect to the left side (positive $\tau_i$) and vice versa (negative $\tau_i$)

The binaural interaction is calculated for low frequencies, which agrees with the characteristic frequency of the MSO cells that are sensitive for interaural time differences. From the system of Eq. 6.3 the power as a function of the frequency and interaural delay time can be calculated. The input signals of the system are: $s_L(t)$ and $s_R(t)$, $c_c$ is the critical band filter with centre frequency $f_c$, and the output signal of the system for one peripheral filter is:

$$p_{c}(t, \tau_i) = c_c(t) * s_L(t) + c_c(t) * s_R(t) * \delta(t - \tau_i)$$  \hspace{1cm} 6.1$$

From the Fourier transformation of the output signal the power function is determined, and integrated over all peripheral filters:

$$P(f, \tau_i) = \int_{-\infty}^{\infty} C(f, f_c) \left[ |S_L(f)|^2 + |S_R(f)|^2 \right. \left. + 2 w_1(f) w_2(\tau_i) \text{Re}\left\{S_L(f)S_R^*(f)\exp[j2\pi f \tau_i]\right\}\right] df$$  \hspace{1cm} 6.2$$

The effect of the peripheral filter is a 'convolution' with the transfer function $C(f)$, as is explained in § 5.3.

The cross spectrum part of the CAP is weighed with a frequency weighing function $w_1(f)$ and a internal delay weighing function $w_2(\tau_i)$. The form of these weighing functions for both interaural delay time and frequency dominance has been calculated from psycho-physical data (Raatgever 1980).
The weighing function for the frequency dominance is given by (with \( f \) in kHz):

\[
w_1(f) = \begin{cases} 
\exp\left(-\left(\frac{f}{0.3} - 2\right)^2\right) & f < 0.6 \text{ kHz} \\
\exp\left(-\left(\frac{f}{0.6} - 1\right)^2\right) & f \geq 0.6 \text{ kHz}
\end{cases}
\]

and is based on psycho-physical lateralisation data.

The weighing function for the delay time is (with \( \tau_i \) in ms):

\[
w_2(\tau_i) = \begin{cases} 
1, & \tau_i \leq 1 \text{ ms} \\
\exp\left(-\frac{1}{|\tau_i| - 1}\right), & \tau_i > 1 \text{ ms}
\end{cases}
\]

Internal delays between 1 and 3 ms are of importance for some dichotic pitch phenomena (Bakkum, 1986). In testing the CAP for prediction of coloration, the CAP is reduced to the interval between -1 and 1 ms, the range where most of the important information is present. The CAP yields a power spectrum for each value of the internal delay. Examples of CAPs are shown in Fig. 6.4 and Fig. 6.5. Fig. 6.4a shows the CAP for diotic white noise, and Fig. 6.4b for correlated white noise for which the right signal is delayed 0.4 ms with respect to the left signal. The internal spectra of Fig. 6.4b are shifted with respect to those of Fig. 6.4a. In Fig. 6.4b the internal spectrum which belongs to \( \tau_i = -0.4 \) ms equals the internal spectrum for \( \tau_i = 0.0 \) ms of Fig. 6.4a.

**(a)**

CAP for correlated white noise (a) and correlated white noise with right signal 0.4 ms delayed with respect to left signal.
In Fig. 6.5a the CAP for diotic harmonic cosine noise with $T_L = T_R = 2$ ms is depicted, and in Fig. 6.5b for dichotic harmonic cosine noise ($T_L = 2.0$, $T_R = 2.5$ ms). As a result of the peripheral filters, the modulation depth of harmonic cosine noise decreases rapidly with the frequency, so for this type of signal only the low frequencies are important. The CAPs are restricted to 2 kHz, the frequency region where the frequency weighing function is larger than zero.

**Fig. 6.5** The CAP of correlated harmonic cosine noise, diotic: $T_L = T_R = 2.0$ ms (a) and dichotic: $T_L = 2.0$ ms, $T_R = 2.5$ ms (b)

**Fig. 6.6** The CAP of correlated complementary cosine noise, $T_L = T_R = 2.0$ ms (a) and $T_L = 2.0$ ms, $T_R = 2.5$ ms (b)
6.5 Binaural decorolation modelled with the Central Spectrum model

In this section the Central Spectrum Model is used to predict the binaural decorolation. For most of the binaural phenomena, the central processor scans the Central Activity Pattern, and selects one (or more) of the internal spectra. This 'central spectrum' is processed for information as lateralisation or dichotic pitch (Raatgever and Bilsen, 1986). It is not very likely that the same mechanism applies for coloration. The colour of a signal consists of multiple cues, such as pitch and timbre, and those cues can be present in every internal spectrum of the CAP; none of the internal spectra possesses information on coloration that is a priori more important than the others. Therefore a binaural decorolation model is proposed in which all internal spectra play an equal part. In section 6.5.1 the model is tested for correlated signals with the aid of the deviation-criterion \( D_0 \) (which uses the mean deviation of a coloured CAP from a flat CAP) and the deviation-criterion \( D_0' \) (the mean deviation of a coloured CAP from a white CAP). In section 6.5.2 the model is tested for uncorrelated signals. Since in this case the CAP reduces to one internal spectrum (summation of left and right spectrum), the \( A_0^- \) and \( B_0^- \)-criterion are used to predict the binaural decorolation. The prediction of the binaural decorolation is tested with the threshold measurements described in Chapter 3.

6.5.1 Model for interaurally correlated signals

An interesting question is whether it is possible to predict the binaural decorolation using the Central Activity Patterns as given in Fig. 6.5 and Fig. 6.6. At first glance, like in the diotic spectra criterion, the coloration strength may appear to be correlated with the modulation strength of the CAP: strongest for diotic harmonic cosine noise, and weakest for complementary cosine noise. However, when the CAP of white noise (Fig. 6.4) is observed, it is apparent that just any modulation present in the CAP not necessarily leads to a coloration sensation. Nevertheless, in this section two models are presented in which the modulation of the internal spectra is considered.

In the first model, for every internal spectrum of the CAP the deviation from a flat spectrum is calculated, and averaged over these deviations. The idea is that if the coloration of the signal decreases, the deviation decreases as well. The minimum level of the mean deviation necessary to perceive coloration, is called the \( D_0^- \)-criterion.

\[
D_0 = \frac{1}{N_\tau} \sum_{i=1}^{N_\tau} \left( \frac{1}{N_f} \sum_{j=1}^{N_f} P(f_j, \tau_i)^2 - \left( \frac{1}{N_f} \sum_{k=1}^{N_f} P(f_k, \tau_i) \right)^2 \right)
\]

(6.5)
In order to test this model, the deviation is calculated for the signals for which the thresholds are given in Chapter 3, namely for correlated harmonic and complementary cosine noise with equal $(T_R = T_L)$ and different $(T_R = T_L + 0.5 \text{ ms})$ delay times for left and right ear. The deviations are calculated at the threshold level and if the model is correct, this will lead to a constant $D_0$-criterion. In Tab. 6.1 the results are given. In this table the value quoted for ‘white’ stands for the deviation when the coloration is minimum, at zero gain of the repetition. From this table it follows that for most of the dichotic signals (shaded in Tab. 6.1) it is impossible to predict a threshold with the $D_0$-criterion found for diotic harmonic cosine noise ($D_0 = 0.113$). For these signals the deviation at threshold (and higher values of the gain) is already smaller than the deviation at minimum gain. For these signals the deviation increases with decreasing gain, and reaches its maximum if the coloration is minimum, while for diotic harmonic cosine noise the deviation decreases with decreasing gain. This can be understood when Fig. 6.4a is compared to Fig. 6.6a: the deviation of white noise is larger than for complementary cosine noise.

**Tab. 6.1** The deviation at threshold for harmonic and complementary cosine noise, with equal and different delay times for left and right ear

<table>
<thead>
<tr>
<th>T (ms)</th>
<th>harmonic, $T_R = T_L$</th>
<th>harmonic, $T_R = T_L + \Delta T$</th>
<th>compl., $T_R = T_L$</th>
<th>compl., $T_R = T_L + \Delta T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>0.127</td>
<td>0.122</td>
<td>0.090</td>
<td>0.120</td>
</tr>
<tr>
<td>4.0</td>
<td>0.113</td>
<td>0.112</td>
<td>0.067</td>
<td>0.109</td>
</tr>
<tr>
<td>8.0</td>
<td>0.115</td>
<td>0.101</td>
<td>0.061</td>
<td>0.106</td>
</tr>
<tr>
<td>12.0</td>
<td>0.110</td>
<td>0.099</td>
<td>0.061</td>
<td>0.103</td>
</tr>
<tr>
<td>16.0</td>
<td>0.110</td>
<td>0.095</td>
<td>0.076</td>
<td>0.101</td>
</tr>
<tr>
<td>20.0</td>
<td>0.109</td>
<td>0.0908</td>
<td>0.072</td>
<td>0.101</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.113±±0.007</td>
<td>0.104±±0.011</td>
<td>0.083±±0.007</td>
<td>0.107±±0.007</td>
</tr>
</tbody>
</table>

Since it is not possible to predict all thresholds, the model is adapted in such a way that at minimum gain the deviation is identically to zero. Instead of the deviation from a flat spectrum, the deviation of the coloured CAP $(P(f, \tau))$ from the CAP of white noise $(P_w(f, \tau) = 1 + w_1(f)w_2(\tau) \cos(2\pi f \tau))$ is used to predict the threshold.

$$D_0 := \frac{1}{N_T} \sum_{i=1}^{N_T} \left( \frac{1}{N_f} \sum_{j=1}^{N_f} \left[ P(f_i, \tau_i) - P_w(f_i, \tau_i) \right]^2 - \left( \frac{1}{N_f} \sum_{k=1}^{N_f} \left[ P(f_k, \tau_i) - P_w(f_k, \tau_i) \right] \right)^2 \right)$$  

6.6

93
For diotic harmonic cosine noise, the $D_0'$-criterion is found to be $D_0' = 0.0047 \pm 0.0056$. With this criterion the thresholds are predicted. The results are given in Fig. 6.7. As can be seen, for diotic harmonic cosine noise (Fig. 6.7a) the predictions are inferior to that of the diotic $A_0'$-prediction. For the dichotic harmonic cosine noise signals the predictions are reasonable (the predictions are too steep as a function of the delay time $T$), for complementary cosine noise with equal delay times, the predictions are quite good. For complementary cosine noise with different delay times the prediction is less good, but the trend of the measurement and the prediction as a function of the delay time is comparable.

![Graphs showing predictions and measurements for diotic and dichotic harmonic cosine noise](image)

Fig. 6.7 Predictions with the adapted $D_0'$-model, compared to the measurements for harmonic with equal (a1) and (a2) different delay times, and complementary cosine noise with equal (b1) and (b2) different delay times.

Although the predictions with the $D_0'$-criterion are reasonable for the dichotic signals discussed here, one must be aware that the model not necessarily predicts the coloration threshold. When the model is used on lateralised white noise (Fig. 6.4b), a
threshold can be predicted, since lateralisated white noise differs from non-lateralised white noise. However, both signals are perceived as white. The deviation model uses all the differences occurring in the CAP, coloration differences and lateralisation differences etc. Unfortunately, the coloration in the CAP cannot be isolated from the other aspects.

6.5.2 Model for interaurally uncorrelated signals

For interaurally uncorrelated signals, the cross spectrum term in Eq. 6.2 vanishes. Thus the CAP is reduced to the information which arrives at left and right ear: the left and right internal spectrum are simply added:

\[ P_u(f_c) = \int_{-\infty}^{\infty} C(f, f_c) \left( |S_L(f)|^2 + |S_R(f)|^2 \right) df \]

6.7

The average of left and right spectrum can be regarded as an ‘effective diotic’ spectrum (Eq. 6.7). Since for diotic signals the \( D_0 \)-criterion gave inferior results with respect to the \( A_0 \)- and \( B_0 \)-criterion, the ‘effective diotic’ spectrum is modelled with the \( A_0 \)- and \( B_0 \)-criterion. The predictions for the binaural decoloration are tested with the threshold measurements as given in § 3.2.

In Fig. 6.8 the spectra of diotic harmonic cosine noise with equal and different delay times left and right are depicted.

![Effective diotic spectra](https://via.placeholder.com/150)

Fig. 6.8 Effective diotic spectra for diotic (top) and dichotic (bottom) harmonic cosine noise, delay times \( T_L=2.0 \), \( T_R=2.5 \) ms (a), \( T_L=4.0 \), \( T_R=4.5 \) ms (b)
For the dichotic $A_0$-criterion the following statement applies: coloration is perceptible if the modulation depth of the effective spectrum exceeds a certain threshold $A_0$ (where $A_0$ is the diotic criterion: Eq. 5.18):

$$A_0 = 10 \log \frac{1 + a^2_{th}(P_u(f_c) - 1)_{\max}}{1 + a^2_{th}(P_u(f_c) - 1)_{\min}}$$

Fig. 6.8 shows that for the effective diotic spectrum the modulation depth is smaller and the regularity less when the left and right spectrum are not equal. For a given delay time difference, the 'distortion' of the effective diotic spectrum with respect to the dichotic spectrum decreases with increasing delay time. The results of the measurements (see Chapter 3, Fig. 3.1) support this observation: the larger the delay time, the smaller the difference between the diotic and dichotic thresholds. Whether the reduction of the maximum modulation depth obtained by this approach is sufficient to satisfy the $A_0$-criterion is tested by predicting the difference between the dichotic and the diotic coloration threshold.

The predictions and the measurements are compared in Fig. 6.9. It is clear that the dichotic $A_0$-criterion applied to the simplified coloration model does not provide a very good prediction of the difference for delay times between 2 and 10 ms. Since the $A_0$-criterion works on the first peaks of the spectrum, binaural decoloration is predicted for small delay times only; the predicted decoloration decreases faster with the delay time than the measured decoloration.

Fig. 6.9  Prediction of binaural decoloration with the dichotic $A_0$-criterion compared to the threshold measurements

The weighed autocorrelation criterion applied to the dichotic situation is defined as: 'coloration is perceptible if the maximum of the sum of the left and right ear autocorrelation (of the spectrum convolved with auditory filters) multiplied with the
Modelling of Binaural Decoloration

autocorrelation time, is larger than $B_0$ (compare the diotic criterion in Eq. 5.28). The threshold of coloration $\alpha_m$ can be calculated from Eq. 6.9.

$$B_0 = a_{\alpha_m}^2 \left[ \frac{1}{2} (\phi_L(\theta) + \phi_R(\theta)) \right]_{\text{max}}$$  \hspace{1cm} 6.9

An example for the sum of weighed autocorrelation is given in Fig. 6.10.

![Graphs showing weighed autocorrelation for harmonic cosine noise with delay times](image)

**Fig. 6.10** Weighed autocorrelation for harmonic cosine noise with delay times $T_L = T_R = 2.0$ ms / $T_L = 2.0$, $T_R = 2.5$ ms (a) and $T_L = T_R = 8.0$ ms / $T_L = 8.0$, $T_R = 8.5$ ms (b)

In Fig. 6.10a the result is given for harmonic cosine noise for $T_L = T_R = 2.0$ ms and for $T_L = 2.0$, $T_R = 2.5$ ms, in Fig. 6.10b for $T_L = T_R = 8.0$ ms and for $T_L = 8.0$, $T_R = 8.5$ ms. If the delay time is small, the autocorrelation peaks are so small that the peak at $T$ and that at $T + \Delta T$ (with $\Delta T = 0.5$ ms) are discernible. Then, the area of the 'most dominant peak' for the dichotic situation is roughly half the size of that of the diotic situation. If the delay time increases, the width of the peaks increases and the peaks become less and less discernible. Hence, the most dominant-peak area increases. In terms of the 'weighed autocorrelation' $\phi(\theta)$, the difference between the maxima of the dichotic and diotic weighed autocorrelation decreases with increasing delay time (compare Fig. 6.10a and b). This is in qualitative accordance with the experimental results (Chapter 3, Fig. 3.1).

The prediction with the dichotic $B_0$ criterion is compared to the measurements in Fig. 6.11. The conclusion drawn from Fig. 6.11 is that the $B_0$-criterion gives a fairly good prediction of the binaural decoloration. For $4 \leq T \leq 20$ ms the prediction gives a difference which is comparable to the measured difference (the absolute difference is $0.9 \pm 0.4$ dB). For 2 ms delay time the measured difference is 3 dB smaller than predicted.
6.6 Criteria of coloration applied to concert hall recordings

To test both the $A_0$-criterion and $B_0$-criterion on natural signals, the white noise masking thresholds of the concert hall signals which were reported in Chapter 5 are predicted. The values of the criteria are calculated with the thresholds measured for these signals using the effective spectra for the dichotic signals. The values found for the criteria are compared in Tab. 6.2 to those found for harmonic cosine noise. As can be seen, the values found for the $A_0$-criterion deduced from the measurements on the diotic hall signals are larger than those for harmonic cosine noise, and for the dichotic hall signals is comparable to that of correlated harmonic cosine noise. The $B_0$-values are much smaller for the hall signals.

Tab. 6.2 Comparing of $A_0$ and $B_0$ criterion calculated from harmonic cosine noise and from the hall signals

<table>
<thead>
<tr>
<th>signal</th>
<th>$A_0$-criterion</th>
<th>$B_0$-criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>uncorrelated harmonic cosine noise</td>
<td>1.30±0.30 dB</td>
<td>0.072±0.014</td>
</tr>
<tr>
<td>correlated harmonic cosine noise</td>
<td>1.70±0.30 dB</td>
<td>0.098±0.011</td>
</tr>
<tr>
<td>dichotic hall signals</td>
<td>1.65±0.45 dB</td>
<td>0.031±0.008</td>
</tr>
<tr>
<td>diotic hall signals</td>
<td>2.51±0.45 dB</td>
<td>0.048±0.012</td>
</tr>
</tbody>
</table>

A possible explanation for the deviation in the $A_0$-criterion is that the values calculated for harmonic cosine noise are deduced from the theoretical spectrum, while the spectra of the hall signals are calculated numerically. The difference found in the $B_0$-criterion (factor 2) is too large to be attributed to the numerical differences only. A difference in value of the criterion ($A_0$ or $B_0$) results in a shift of the predicted thresholds, but as this effects all hall signals in a similar manner the trend among the
halls can be predicted correctly (within ±2 dB), as can be seen in Fig. 6.12. Only in the 'diotic right ear, rear position' the thresholds for halls 'APD' and 'AU4' are much lower than predicted. Still, the prediction that hall 'AU4' has the lowest threshold is correct. The conclusion is that the diotic predictions for the effective diotic spectra provide a reliable prediction of the masking threshold, as can be seen from Tab. 6.3, where the hypothesis 'no significant difference between prediction and threshold' is tested, for the shaded table cell the hypothesis is rejected.

Fig. 6.12 Thresholds compared to A₀- and B₀-prediction for dichotic rear position (a) and front position (b), diotic "left" (c) and diotic "right" (d) (rear position)

Tab. 6.3 Results of the t-test for n=9 and α=0.05: tₐ=2.31

<table>
<thead>
<tr>
<th>signal</th>
<th>t value for A₀</th>
<th>t value for B₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>dichotic, rear</td>
<td>1.71</td>
<td>1.68</td>
</tr>
<tr>
<td>dichotic, front</td>
<td>2.52</td>
<td>1.77</td>
</tr>
<tr>
<td>diotic, rear, left</td>
<td>0.83</td>
<td>0.23</td>
</tr>
<tr>
<td>diotic, rear, right</td>
<td>0.64</td>
<td>0.79</td>
</tr>
</tbody>
</table>
6.7 The coloration of uncorrelated complementary cosine noise considered

Although the $B_0$-model applied to the effective diotic spectrum, as defined in section 6.5.2, gives satisfactory predictions for binaural decoloration for uncorrelated harmonic cosine noise and hall signals, it does not predict the correct amount of binaural decoloration for uncorrelated complementary cosine noise. The 'effective diotic spectrum' obtained from adding the left and right spectrum, is flat and hence the predicted threshold is zero. As can be seen from Fig. 3.4, this is clearly not the case. This means that for uncorrelated complementary cosine noise the CAP as defined in Eq. 6.2 does not predict this coloration. A possible reason for this discrepancy between model and measurement, is that the CAP does not take into account the difference in coloration strength of harmonic and anharmonic cosine noise (see § 2.5.1). The cancellation of harmonic and anharmonic cosine noise therefore can not be optimum. The difference of 2 dB between harmonic and anharmonic cosine noise thresholds (Fig. 2.10a) is insufficient to explain the coloration strength of complementary cosine noise (threshold about 15 dB, see Fig. 3.4). Another reason can be that no exact summation of the signals of left and right ear occurs, and that 'leakage' of left- and right signal produces the colour sensation. Yet another possibility is the use of the monaural channels (Fig. 6.1), which may result in a more or less independent colour sensation for the left and right ear signal. However, the signals cannot be entirely separated in this model, otherwise the coloration strength of uncorrelated complementary cosine noise would be the same as for harmonic cosine noise. Instead, a certain amount of binaural interaction takes place.

6.7.1 Zurek's binaural decoloration model

The binaural interaction of uncorrelated left- and right signals was also topic of research by Zurek (Zurek, 1979). In the relevant research uncorrelated cosine noise was attenuated at one side with a factor $k$, so that at maximum attenuation a monotonic signal remained. This experiment was done for both harmonic and complementary cosine noise, where in the latter signal the anharmonic side was attenuated. In the experiment the delay time was kept constant ($T=3$ ms) and the coloration thresholds were measured as a function of $k$, using a 2AFC-procedure. The results of these experiments are given in Fig. 6.13. For harmonic cosine noise, the threshold is constant as a function of $k$, for complementary cosine noise the threshold decreases linearly with $k$.

Zurek proposed as model a non-linear summation, with a compression exponent $p$ to explain these results:

$$P_z(f) = |S_L(f)|^{2p} + |S_R(f)|^{2p}$$  

6.10
Then the thresholds are predicted by applying the $A_0$-criterion to $P_z(f)$. In Fig. 6.13 the prediction is given as function of $k$, and the prediction suits the measurements very well. The $A_0$-criterion was determined from the measurements for harmonic cosine noise as measured by Zurek, $A_0=1.38$ dB, while the resulting compression exponent is $p=0.34$.

![Diagram showing coloration thresholds as function of the attenuation $k$ of the left ear signal, for uncorrelated harmonic and complementary cosine noise (anharmonic cosine noise is attenuated), and threshold prediction with the compression model (After Zurek, 1979)](image)

Fig. 6.13 Coloration thresholds as function of the attenuation $k$ of the left ear signal, for uncorrelated harmonic and complementary cosine noise (anharmonic cosine noise is attenuated), and threshold prediction with the compression model (After Zurek, 1979)

Zurek did not use auditory filtering in his model. Because the results of Fig. 6.13 are measured for only one delay time, this hardly plays a role. In order to predict the coloration threshold as function of the delay time, the auditory filters are introduced in his model as follows:

$$P_z(f_c) = \left[ \int_{-\infty}^{\infty} |C(f, f_c)|^2 d(f) \right]^p + \left[ \int_{-\infty}^{\infty} |C(f, f_c)|^2 d(f) \right]^p$$

6.11

It appears that when the $A_0$-criterion is applied to this model the coloration thresholds, predicted as function of $T$, are much too high, as can be seen in Fig. 6.14a. For these predictions the $A_0$-criterion is determined from the present measurements, $A_0=1.27$, and the prediction thresholds are given for different values of $p$. Comparison of these results to those depicted in Fig. 6.13 shows that for $k=0$ dB the threshold found by Zurek is much higher than the present result. This can be due to Zurek using a roving level in his 2AFC-procedure, in order to avoid that the subjects use intensity discrimination instead of coloration. This roving level, however, can make the experiments more laborious for the subject, resulting in a higher threshold.

In the experiments described in Chapter 3, the adjustment method was used with the explicit instruction to the experienced subjects to find the thresholds of coloration.
And although the subjects agreed that the coloration sensation of complementary cosine noise was usually notable less than for harmonic cosine noise, a distinct coloration was still perceptible, leading to a lower threshold than found by Zurek.

\[ \text{dB} \]
\[ g \]
\[ T \text{ ms} \]

(a)  

(b)  

Fig. 6.14 The thresholds of uncorrelated complementary cosine noise, with the same (a) and different (b) delay times at left and right ear signal, and the prediction with the compression model

In order to test the measurements further, the model is applied to the thresholds for 'complementary' cosine noise with different delay times for left and right ear (Fig. 6.14b). In this figure the compression factor is set to the value found by Zurek, p=0.34. These predictions increase to fast as function of T. The conclusion is that in spite of the promising results for the prediction of the thresholds as function of the attenuation of anharmonic cosine noise, the model is less suitable to explain all aspects of the binaural decoloration of uncorrelated complementary cosine noise.

6.7.2 'Best ear'-model

Further proof that complementary cosine noise can be perceived as quite coloured, is found in the following measurements from Put (1992), where the internal spectrum is scanned with a sine probe. Correlated complementary cosine noise is presented to the subject (delay time at both ears: 2.0 ms), together with a sine of a variable frequency. The correlated complementary cosine noise is used to mask the sine. The sine level has to be decreased until it is no longer perceptible. The required attenuation of the sine probe is supposed to be related to the power of the masking signal at the frequency of the probe. The more activity of the masking signal at the probe frequency, the less the probe has to be decreased before it is masked. In these masking experiments three different experimental set-ups were used.
Correlated complementary cosine noise is produced by presenting anharmonic cosine noise to the left ear and harmonic cosine noise to the right ear. In the first experiment the sine probe is presented to the left ear only, in the second experiment the sine probe is presented to the right ear, in the third experiment to both ears. In Fig. 6.15 the results are given for the first and second experiment.

Since the attenuation of the sine at masking level is the reverse of the gain of the signal at the related ear, the vertical axis of the results has been reversed, so the results can be compared directly to the signal level at the corresponding ear, which is depicted below the masking results.

Fig. 6.15 Sine probe masked by correlated complementary cosine noise, for probe in respectively left (a) and right (b) ear. Top: the masking results (sine probe attenuation), bottom: the signal spectrum (filtered with auditory filters) at the corresponding ear. The points in the spectra correspond with the frequencies at which the experiments were performed.

When comparing the results with the masking level, it is clear that the inverse of the masking level at a certain frequency closely follows the signal strength at that frequency. This indicates that to a large extent both ears operate independently. When presenting complementary cosine noise, coloration of anharmonic cosine noise is perceived at the left ear and harmonic cosine noise at the right ear, as would be expected with the monaural channel model. The only deviating measurement is that for 0.25 kHz at the right ear, where the attenuation at masking level of the sine is expected to be higher, since the signal level at this point is far lower than at 0.75 kHz, but the masking level at 0.75 kHz equals the masking level at 0.25 kHz.
Fig. 6.16 All results of masking a sine probe with correlated complementary cosine noise (a), the attenuation at threshold follows the level of the ear which is least masked (b).

In Fig. 6.16a the results of all three experiments are given. When the attenuation of the probe applied to both ears is compared to the attenuation for the single ear probe, the attenuation at masking level is observed to depend on the lowest signal level. When for a certain frequency the lowest signal level occurs at the left ear, the attenuation of the probe at masking level matches the attenuation when the probe is presented to the left ear only. When for another frequency the lowest signal level is found at the right ear, the masking level of the dichotic probe matches the attenuation of the probe at the right ear only. Where the signal level at both ears are equal, the masking level attenuation is comparable for all three situations (left only, right only, both). In Fig. 6.16b the attenuation at the threshold for the probe in both ears is compared to the minimum of the left and right ear signals. The results correspond with this minimum signal level, except for the point at 0.25 kHz. At this frequency the attenuation at masking level is lower than expected, like in the case when the probe is presented to right ear only. In both cases the attenuation is measured to be 5 dB, while about 15 dB was to be expected from the signal level.

Similar measurements have also been performed for uncorrelated complementary cosine noise (van Keulen, 1990). In Fig. 6.17a the results are given for the probe presented to one ear (harmonic cosine noise) and to both ears. The results are very well comparable to the results of correlated complementary cosine noise, as is depicted in Fig. 6.17b for frequencies up to 1 kHz.
Fig. 6.17 Results for uncorrelated complementary cosine noise with sine probe in both ears (a) compared with the results of correlated complementary cosine noise (b)

When the CAP of correlated complementary cosine noise is considered (Fig. 6.6a), it is obvious that this model cannot provide an explanation for the results found by Put. The peaks and valleys of the internal spectra of the CAP do not match the attenuation of the sine at the threshold level. For uncorrelated complementary cosine noise the CAP is flat for all internal spectra, and is identical to that for uncorrelated white noise. Therefore, if uncorrelated complementary cosine noise is used to mask a sine probe, the masking level should then be the same for all frequencies. But since the results of uncorrelated complementary cosine noise match the results of correlated complementary cosine noise, the same mechanism is used by the ear for both signal types and apparently the interaural correlation does not play an important role in this experiment. A plausible explanation of this result is that the monaural channels of both ears are scanned, and the channel of the ear with the most important information determines the masking level:

$$P_{\text{min}}(f_c) = \min \left\{ \int_{-\infty}^{\infty} C(f, f_c) S_L(f)^2 \, df, \int_{-\infty}^{\infty} C(f, f_c) S_R(f)^2 \, df \right\}$$  

This model, in combination with the $A_0$-criterion, is tested to predict the amount of decoloration of complementary cosine noise with respect to harmonic cosine noise. As can be seen from Fig. 6.18, the predicted amount of decoloration is too small: the coloration of complementary cosine noise according to this model is stronger than measured. From this it can be concluded that the results as found by Put do not explain the strength of coloration perception. But if for the sine masking measurements the monaural channels are determining the outcome, it is likely that the monaural channels play an important role in coloration perception as well.
resemblance between the results of the sine masking and coloration experiments is that in both experiments the threshold of correlated complementary cosine noise is comparable to that of uncorrelated complementary cosine noise. This suggests that also for coloration a similar mechanism is used for both uncorrelated and correlated complementary cosine noise, and that interaural correlation is of less importance.

![Graph showing dB vs. ms with prediction and threshold lines.]

**Fig. 6.18** The amount of decoloration for complementary cosine noise with respect to harmonic cosine noise, predicted with the 'most important information'-model and the $A_0$-criterion

The results of Put give rise to the surmise that it is possible to concentrate on one ear while scanning for the (masked) sine. The concentration on one ear while judging coloration is not impossible, but it is not very likely to happen. When masking the sine, the concentration is on one specific frequency, namely the probe frequency. With coloration the modulation depth of the spectrum is important for all frequencies, because the perception of the coloration is an overall spectral perception. If the subject would be concentrating on one ear only when measuring coloration (that with the most dominant coloration), one should expect the same threshold for harmonic cosine noise and complementary cosine noise. The difference between harmonic and complementary cosine noise thresholds is sufficiently large to reveal that the subject is not concentrating on one channel. The information of both monaural channels is considered and fused to a single overall perception of coloration. However, the question remains how the monaural channels work together to bring about this overall perception, and how this effect can be incorporated in a model for coloration perception.
6.7.3 Pitch sensation model

The conclusion drawn from §6.7.1 and §6.7.2 is that models based on the spectral combination of harmonic and anharmonic spectra do not lead to a proper prediction of the coloration threshold of uncorrelated complementary cosine noise. Therefore, in this section an attempt is made to predict these coloration thresholds with the aid of the separate pitch perception of harmonic and anharmonic cosine noise.

In the case that the ears receive signals that are uncorrelated, but equal in pitch sensation, the threshold is found to be lower than when the pitch perception of the left and right ear differ. This is for instance the case for harmonic cosine noise with different delay times for the left and right ear signal, and this difference could be predicted with the $B_0$-criterion (see § 6.5.2 and Fig. 6.11). For harmonic cosine noise the pitch perception equals the fundamental tone of the spectrum, namely $1/T$, see § 2.2 and Bilsen, 1968. The pitch perception of harmonic cosine noise is characterised in the following way:

$$\varphi_{\text{pitch}}(\theta, T) = \varphi_h(\theta, T)$$  \hspace{2cm} (6.13)

where $\varphi_h(\theta, T)$ indicates the autocorrelation of the harmonic cosine noise spectrum for delay time $T$. For anharmonic cosine noise the pitch perception is weaker and more ambiguous. The central pitch processor 'calculates' a best fitting harmonic spectra and the pitch 'matched' to these harmonic spectra are 0.89/T and 1.14/T (§ 2.5 and Bilsen, 1977 among others). So if harmonic and anharmonic are combined to make complementary cosine noise, the pitch from left and right ear differ. The pitch perception of anharmonic cosine noise is now described as the autocorrelation of the harmonic fits to the anharmonic spectrum:

$$\varphi_{\text{pitch}}(\theta, T) = f_{\text{anharm}}[\varphi_h(\theta, T/0.89) + \varphi_h(\theta, T/1.14)]$$  \hspace{2cm} (6.14)

The factor $f_{\text{anharm}}$ is used to account for the weaker strength of the pitch sensation, and is calculated from the coloration thresholds for anharmonic cosine noise. (see Fig. 2.8b), which gives $f_{\text{anharm}}=0.70\pm0.12$. The motivation of this model is that, as suggested by the measurements by Swets and by Vos (see § 2.5 and § 4.5), pitch is a very important aspect of the coloration sensation. The underlying assumption can then be that for uncorrelated signals first a pitch perception is formed for each monaural channel, before these are combined in a total coloration perception.

The thresholds of uncorrelated harmonic cosine noise were calculated by applying the $B_0$-criterion on the sum of the autocorrelation of left and right ear signal (Eq. 6.9), with $B_0=0.072$ (§5.5). The threshold of uncorrelated complementary cosine noise is now calculated by adding the autocorrelation of harmonic cosine noise and the autocorrelation of the harmonic fits to the anharmonic spectrum:

$$\varphi_{\text{pitch}}(\theta, T_L, T_R) = 1/2[f_{\text{anharm}}[\varphi_h(\theta, T_L/0.89) + \varphi_h(\theta, T_L/1.14)] + \varphi_h(\theta, T_R)]$$  \hspace{2cm} (6.15)
Fig. 6.19 Dichotic $B_0$-criterion applied to the 'pitch autocorrelation' of harmonic cosine noise with equal (a1) and different (a2) delay times for left and right ear, on anharmonic cosine noise (b) and complementary cosine noise with equal (c1) and different (c2) delay times for left and right ear.
In Fig. 6.19 the results of the predictions with the $B_0$-model are depicted. These figures show that the threshold for uncorrelated harmonic and anharmonic cosine noise can be predicted correctly with this model. For uncorrelated complementary cosine noise the thresholds are predicted about 2.5 dB too low. For uncorrelated complementary cosine noise with different delay times the predicted thresholds increase somewhat faster than measured. In Tab. 6.4 the hypothesis 'no significant difference between predictions and thresholds' is tested with the t-test. For the shaded cell this hypothesis is rejected, namely for complementary cosine noise with equal delay times. However, the trend in the complementary cosine noise predictions is promising and is definitely better than when adding the autocorrelation functions of harmonic and anharmonic cosine noise directly. In future models the pitch aspects of the coloration should be incorporated.

<table>
<thead>
<tr>
<th>signal</th>
<th>$t_\alpha$</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>harmonic, TR=TL</td>
<td>2.57</td>
<td>0.62</td>
</tr>
<tr>
<td>harmonic, TR=TL+$\Delta$T</td>
<td>2.31</td>
<td>1.18</td>
</tr>
<tr>
<td>complementary, TR=TL</td>
<td>2.16</td>
<td>3.43</td>
</tr>
<tr>
<td>complementary, TR=TL+$\Delta$T</td>
<td>2.57</td>
<td>2.38</td>
</tr>
</tbody>
</table>

**6.8 Conclusions on binaural hearing models**

Binaural decoloration has been clearly demonstrated in various experimental observations. The thresholds of dichotic signals are often significantly higher than those of diotic signals. The modelling of this binaural decoloration which can predict those threshold differences is still unsatisfactory. The Central Activity Pattern, which describes the interaction between left and right signal was designed originally to predict dichotic pitch (Bilsen, 1977) and laterisation (Raatgever and Bilsen, 1986). In most cases, the CAP provides a satisfactory explanation for these phenomena. The cues, both laterisation and coloration, are related to the shape (i.e. the modulation depth of the peaks) of the internal spectra. It is not always clear whether the shape of an internal spectrum will be interpreted as a laterisation cue or a coloration cue. If two stimuli are perceived as both having the same colour, this does not imply that the CAPs need to be equal. If the first stimulus has a central perception, while the other is shifted laterally, the CAPs are evidently different. It is therefore too complex to distinguish in general which part of the CAP provides the information about the perceived colour and colour strength of the stimulus. This becomes even more clear if harmonic cosine noise with different delay times is
studied (Fig. 6.5a and b). Here the coloration is altered, but also the percept is 
lateralised with respect to harmonic cosine noise. As far as the binaural decoloration 
is concerned, the CAP suggests the activity to be less coloured when left and right 
signal differ (compare Fig. 6.5a and b).

For correlated signals the $D_0$-criterion and the $D_0'$-criterion were tested, respectively 
the deviation of a coloured CAP from a flat CAP and the deviation of a coloured CAP 
from a correlated white noise CAP. With the $D_0$-criterion it was not possible to predict 
most of the dichotic signal thresholds. The $D_0'$-criterion gave a satisfactory result for 
the threshold prediction of correlated complementary cosine noise, a moderate result 
for dichotic harmonic cosine noise. For diotic harmonic cosine noise the results of the 
$D_0'$-criterion were inferior to those of the $A_0$-criterion.

For uncorrelated signals, the CAP reduces to an ‘effective diotic spectrum’, equal to 
the sum of the left and right spectrum. To this ‘effective diotic spectrum’ the $A_0$-
criterion and $B_0$-criterion were applied. This resulted in reasonable binaural 
decoloration predictions of the thresholds for the $B_0$-criterion for harmonic cosine 
noise with different delay times left and right and for the $A_0$-criterion for the concert 
hall signals. Also, the binaural $B_0$-criterion correctly predicts the relative levels of the 
thresholds for the concert hall signals.

For uncorrelated complementary cosine noise the threshold could not be predicted 
with a linear combination of harmonic and anharmonic spectrum. A non-linear 
summation, as proposed by Zurek (Zurek, 1979), could predict how the threshold 
decreases as function of the attenuation of anharmonic cosine noise, but for the 
prediction as function of the delay time the model was less effective. From sine-
probe measurements (Put, 1992) it appeared that for an analytic listening situation 
the most important information of either ear is used. However, also this model is less 
suited to predict the coloration thresholds. The most effective prediction was made 
with the assumption that the coloration threshold is found by using the pitch 
information of both the harmonic and anharmonic signal. In this case the pitch 
information of anharmonic cosine noise is described with the autocorrelation of the 
harmonic fits. This result stresses the role of the pitch sensation in the coloration 
perception, which is also supported by the measurements done by Swets (see § 2.5 
and Swets, 1994) and, when comparing the coloration of hall signals, by Vos (see 
§ 4.5 and Vos, 1994).

The conclusion is that the Central Activity Pattern alone is less suited for the 
prediction of the amount of binaural decoloration than was assumed by Zurek. The 
development of models which predict binaural decoloration should include the effect 
of the monaural channels, as well as the pitch perception.
Chapter 7

General Conclusions

7.1 Introduction

In this thesis experiments on (de)coloration were described for a wide variety of signals. The coloration of a signal has been defined in Chapter 1 as "that attribute with which a listener can judge that two sounds similarly presented and having the same loudness are dissimilar" and which involves a spectral deformation with respect to either the original signal or a reference signal such as white noise.

As was said in Chapter 1, it poses quite a challenge to create an experimental set-up where one can be confident about which aspect of the coloration the subject uses in his judgement, the timbre or pitch sensation (or both). This question was encountered in the difference in thresholds between harmonic and anharmonic cosine noise (§ 2.5), and in the comparison of profile analysis with coloration thresholds (§ 2.8). With the multidimensional scaling the subjects used both aspects, in which it depended on subject to which extent the aspects were valued, but on average there was a little more weight attached to the pitch sensation (0.67 versus 0.44 for timbre sensation).

In the subsequent sections of this chapter, the results of the experiments are given, and the conclusions are drawn.

7.2 Coloration of artificial signals

Thresholds have been measured for harmonic cosine noise, i.e. white noise to which one reflection is added. The threshold is determined by comparing the coloured signal to white noise. The coloration of the signal is weakened until no difference is detected between the coloured signal and the white noise. The coloration of harmonic cosine noise can be weakened by either altering the gain of the repetition (g-threshold) or by masking the signal with white noise (a-threshold). A low threshold corresponds to a signal with strong coloration.

The results (§ 2.4.2, Fig. 2.8) closely match results of measurements that have been reported previously, for instance by Atal et al (1962), Bilsen (1968) and Zurek (1979). Clearly the strength of the coloration decreases with increasing delay time. If more
reflections are added to the white noise the threshold strongly depends on the pattern of the delay time differences between the reflections (§ 2.6 Fig. 2.16, Fig. 2.17). The threshold decreases (i.e. stronger coloration) if the reflections are equally spaced. For randomly spaced reflections the threshold increases (i.e. weaker coloration), where for delay times of about 8 ms this effect is more pronounced than for small delay times of about 1 ms. For delay times of about 8 ms, the signal is perceived nearly white for twelve repetitions. For delay times near 1 ms, the signal is still strongly coloured for sixteen repetitions.

The threshold measurements are compared to scale values derived from paired-comparisons (§ 2.7, Fig. 2.19). In this method a set of signals is ordered on a (subjective) coloration scale, based on successive pairwise comparison of the signals. This latter method takes the appreciation of the coloration into account, and can therefore provide an indication to which extent the threshold measurements are representative of the coloration perception, which is essentially a sensation occurring above threshold level. A distinct correlation between both experimental approaches was obtained only when the signals differ significantly in threshold. For threshold with small relative differences, with a relative standard deviation between the thresholds of about 25%, the scale values lie more apart than expected; the correlation coefficient between paired-comparison and the thresholds is R=0.1 or less (see § 2.7 and § 4.3). The conclusion is that coloration (above threshold) is indeed stronger if the threshold is lower, but even if two signals have comparable threshold values one may still judge the coloration to be quite different in strength. Since the coloration criteria (A₀- and B₀-criterion) were derived from the threshold measurements, these criteria should therefore be used with care. Consequently, to ensure that a signal is not perceived as very strongly coloured, a high threshold is a necessary but not sufficient condition.

For anharmonic cosine noise higher coloration thresholds were found than for harmonic cosine noise (§ 2.5). This agrees with earlier reported observations that for anharmonic cosine noise the pitch strength is less than for harmonic cosine noise (Yost and Hill, 1978). Experiments by Swets (1994) show that the difference between harmonic and anharmonic coloration thresholds is of the same order as the difference in pitch strength. The conclusion is that the pitch sensation of the signal plays an important role in the coloration perception.

7.3 Binaural decoloration experiments

Binaural decoloration occurs if different signals are applied to each ear. For small delay time differences, about 0.5 ms which is a realistic value to occur in practical situations, decoloration could indeed be detected. For harmonic cosine noise with 0.5 ms delay time difference between right and left ear, the decoloration was found to
depend on the repetition delay time of the cosine noise and varies between 0 and 5 dB (see § 3.2, Fig. 3.1).

The binaural decoloration effect increases if in a signal the left ear spectrum (partially) cancels the right ear spectrum, as in the case of complementary cosine noise. However, the coloration does not vanish completely as would be expected if the right and left ear signal spectra would simply add. The decoloration measured for complementary cosine noise varies between 5 and 10 dB for uncorrelated complementary cosine noise and between 0 and 10 dB for correlated complementary cosine noise, for the latter the decoloration shows a clear dependence on the delay time of the repetition (see § 3.3.1, Fig. 3.4). Comparable results were found by Zurek (1979) and Lassche (1990). If the signals are not entirely complementary due to a difference in delay time for the left and right ear signal, the decoloration decreases. The decoloration then varies between 3 and 5 dB if the left and right signal are interaurally uncorrelated, and between 0 and 5 dB if the signals are interaurally correlated (Fig. 3.13).

A complication encountered in measuring the decoloration of dichotic signals is the broadening of the percept. If the coloration threshold of a dichotic signal is measured with a 2AFC method, it is possible that the subject employs the broadness of the percept to discriminate between the stimuli, instead of the difference in coloration. If the threshold for the broadness of the percept is lower than the coloration threshold, the experiment yields a threshold that is too low, resulting in an underestimation of the decoloration. This problem caused that in the experiment described in § 3.4 the decoloration was smaller than expected (smaller than 2 dB difference). Therefore, it is recommended that coloration thresholds of dichotic signals should preferably be measured with an adjustment method (as in § 3.3) or a paired-comparison experiment, in which the subject is instructed to listen to coloration only.

7.4 Coloration of concert hall signals

Coloration is always present in signals that are produced in concert halls, but in a good concert hall the coloration is only weak. Coloration measurements were carried out in a number of concert halls in the Netherlands (Chapter 4). The halls differ significantly in coloration, as revealed by the results of the paired-comparison (§ 4.2, Fig. 4.1), although the differences between the coloration thresholds are relatively small (§ 4.3, Fig. 4.3). The binaural decoloration determined from threshold measurements is small for all signals, which is probably due to the extra dichotic cues (§ 4.4, Fig. 4.7).

The multidimensional scaling experiments by Vos (1994) show that the pitch sensation and the timbre aspects of the coloration play an important role in the comparison
of the stimuli recorded in a concert hall (§ 4.5). The results of these experiments also demonstrate that the coloration of the hall varies from place to place (Fig. 4.9). Paired-comparison of signals recorded at front and rear position in various concert halls (§ 4.2, Fig. 4.1) show that the place-to-place difference is smaller than that between different concert halls.

According to the paired-comparison the Delft Auditorium gives, both with and without ACS-system, more coloration than average. For the threshold experiments, the ACS-system gave no improvement, the threshold was 3 dB lower for AU4 with respect to AU0 (§ 4.3). The paired-comparison did show some improvement with the ACS-system, the coloration scale values were for 0.72 AU0 and 0.61 for AU4 (§ 4.2). In the experiments in which the coloration was measured as function of the seat location (§ 4.5) the results for AU0 and AU4 were comparable. The conclusion is, that the ACS-system does not induce a place dependent coloration, but that the coloration does not improve much with the use of the system.

7.5 Criteria for coloration for diotic signals

Coloration perception is modelled in this thesis on basis of the spectrum of the signal, to which a filtering with auditory filters is applied (Chapter 5). The revised \( A_0 \)-criterion determines the maximum modulation depth of the spectrum convolved with auditory filters (Eq. 5.18), while the \( B_0 \)-criterion is based on the area of the most ‘dominant’ autocorrelation peak, i.e. the autocorrelation peak with the largest area, see Eq. 5.21. This area can be approximated by the peak value weighed with the delay time (Eq. 5.28). This greatly facilitates the implementation in the case when multiple peaks occur in the autocorrelation function, and where the dominant peak may be difficult to define unambiguously.

Both criteria are derived from the criteria proposed by Atal et al (1962), although in their original form the criteria were based on the wide-band short time autocorrelation instead of the auditory filters approach used here. Both the \( A_0 \)-criterion and the \( B_0 \)-criterion perform reasonably well in predicting the coloration thresholds of the synthetic signals (Fig. 5.4 and Fig. 5.10), that is for coloration due to spectral deformation of the signal. For the concert hall signals the \( B_0 \)-criterion predicts higher thresholds than were measured. For coloration due to large delay times (larger than 20 ms) both criteria fail, because the peaks of the spectrum are spaced too closely together to be resolved by the auditory filters. The spectrum convolved with auditory filters is flat and no coloration is predicted, although coloration can still be perceived as a rattle sensation. The coloration of these type of signals can be explained from a mechanism that recognises time structures in signals (see Warren et al, 1980), in contrast to the spectral coloration to which the above criteria apply.
7.6 Modelling of binaural decoloration

Binaural decoloration is confirmed by coloration thresholds for the majority of signals. An attempt to model this coloration was carried out by means of the Central Spectrum Model (Bilsen, 1977; Raatgever and Bilsen, 1986) which describes the interaction between the left and right ear signals to form a Central Activity Pattern (CAP). In order to predict the binaural decoloration with this model, the $D_0^\ast$-criterion and $D_0^\ast$'-criterion were proposed: the deviation of a coloured CAP from a flat CAP, or from a (correlated) white noise CAP, respectively. The $D_0^\ast$-criterion failed to predict thresholds for most of the dichotic signals tested (Tab. 6.1). The $D_0^\ast$'-criterion gave satisfactory results for dichotic signals, especially for the correlated complementary cosine noise (Fig. 6.8 b-d). For diotic signals, the $D_0^\ast$'-criterion performed less adequate (Fig. 6.8a). A drawback to this criterion is that any modulation, whether due to coloration or lateralisation, affects the threshold prediction. Therefore, the prediction for the dichotic thresholds is not necessarily the coloration threshold.

For interaural uncorrelated signals the CAP forms an 'effective diotic' spectrum, as the internal spectra of left and right ear add directly. Since the $D_0^\ast$'-criterion was found inferior to the $A_0$-criterion for diotic signals, the binaural decoloration of uncorrelated signals was predicted with the $A_0$-criterion and $B_0$-criterion applied to the 'effective diotic' spectrum. With the $B_0$-criterion this model gave a satisfactory prediction of the decoloration for harmonic cosine noise with different delay times at the left and right ear (Fig. 6.12); with the $A_0$-criterion the measured decoloration was larger than predicted (Fig. 6.10).

For uncorrelated complementary cosine noise the CAP-based predictions yielded no satisfactory result. An uncorrelated complementary cosine noise CAP and an uncorrelated white noise CAP are both flat, hence, no coloration is predicted. For this signal the model as proposed by Zurek (Zurek, 1979) was tested, namely a non-linear adding of left and right ear spectra (Eq. 6.12 in § 6.7.1). For the $A_0$-prediction, the thresholds of complementary cosine noise were predicted about 7 dB too high (less coloration predicted than was present) which means that the decoloration is overestimated (Fig. 6.14). That still quite some coloration can be perceived for complementary cosine noise was shown by Put (1992) (§ 6.7.2). In his research a sine probe was masked by correlated complementary cosine noise. The results of this experiment were satisfactory explained with the model that the most significant information of left and right ear is used by the monaural channels when masking the sine probe. For the coloration thresholds, however, this 'best-ear' model underpredicts the thresholds for complementary cosine noise with about 7 dB (Fig. 6.18). The most effective prediction was made with the assumption that the coloration threshold is found by using the pitch information of both the harmonic and anharmonic signal. In this case the pitch information of anharmonic cosine noise is described with the autocorrelation of the harmonic fits. This result stresses the role of the pitch sensation in
the coloration perception, which is also supported by the measurements done by Swets (see § 2.5 and Swets, 1994) and, when comparing the coloration of hall signals, by Vos (see § 4.5 and Vos, 1994).

The conclusion is that the Central Activity Pattern can provide an indication of the decoloration effect. However, the CAP alone is less suitable for predicting the exact amount of decoloration. In view of the results for complementary cosine noise, future models should include the effect of the monaural channels, which provide the 'higher order processes' in the brain with the left and right signals directly. How the information of the monaural channels and the central activity pattern is combined to form a total percept of coloration should be an important theme for future investigations.
Appendix A

Delay Times of Multiple Repetitions

Tab. A.1 the repetition delay times of the artificial signals as described in Chapter 2. In the first row the first delay time is given, in the second row the mean delay time difference, as used in the caption of the figures in § 2.6. In the following rows the delay times are given as function of the number of the repetition. All delay times are in milliseconds.

Tab. A.1  Delay time of the repetitions as function of the number of the repetition

| $T_1$ (ms) | 0.94 | 8.46 | 8.46 |
| $\Delta T_m$ (ms) | 0.99±0.36 | 8.88±3.33 | 8.46 |
| 1 | 0.94 | 8.46 | 8.46 |
| 2 | 2.26 | 20.34 | 16.92 |
| 3 | 2.78 | 25.02 | 25.38 |
| 4 | 3.82 | 34.38 | 33.84 |
| 5 | 5.02 | 45.18 | 42.30 |
| 6 | 5.62 | 50.58 | 50.76 |
| 7 | 6.98 | 62.82 | 59.22 |
| 8 | 8.18 | 73.62 | 67.68 |
| 9 | 8.86 | 79.74 | 76.14 |
| 10 | 9.34 | 84.06 | 84.60 |
| 11 | 10.46 | 94.14 | 93.06 |
| 12 | 12.14 | 109.26 | 101.52 |
| 13 | 13.06 | 117.54 | 109.98 |
| 14 | 14.38 | 129.42 | 118.44 |
| 15 | 15.22 | 136.98 | 126.90 |
| 16 | 15.74 | 141.66 | 135.36 |
Appendix B

Psycho-physical Methods to Measure Coloration

B.1 Introduction

In this appendix the experimental methods as used in the experiments are described and the experimental environment is given. The experimental methods comprise the methods with which a quantitative measure for the coloration is obtained (the threshold of perceptibility of coloration), as well as methods which provide a qualitative measure (the comparison methods). The threshold methods contain the 2AFC (two alternative forced choice) up/down-strategy and the 2AFC trajectory method, and the adjustment method; the comparison methods contain the paired-comparison and multidimensional scaling.

In addition, details are given of the experimental environment such as a key to the subjects who participated in the experiments, as well as some remarks on the experimental equipment.

B.2 Threshold of coloration

B.2.1 The Two Alternative Forced Choice method

In a Two Alternative Forced Choice (2AFC) method the subject has to judge stimuli by choosing between two alternatives, of which one is correct. The subject will probably give the correct answer if he can detect the difference, but if he cannot detect any difference, he may give the correct answer by chance. The course of the experiment is influenced by the correctness of his choice.

In the implementation of this method several strategies can be followed, concerning the number of stimuli provided per decision event. In coloration experiments, coloured noise is compared to white noise. In the simplest form of the experiment only two stimuli, a white and a coloured stimulus are presented in random order, and the subject has to indicate which stimulus is coloured. This experimental set-up is hampered by the absence of a reference: near the threshold a slight difference between the stimuli can be detected, but it may not be possible to tell which stimulus is the coloured stimulus. The experiment can be made easier by adding a reference
stimulus. In this case the first reference stimulus is white noise, and the next two stimuli are coloured noise and white noise in random order. Now the subject has to indicate which stimulus differs from the reference stimulus. The only remaining problem of this method is that the subject has to remember the reference stimulus when responding.

In yet another method four stimuli are presented to the subject, two sets of two stimuli. In one set the stimuli differ, in the other set the stimuli are equal, e.g. the first set is white noise/white noise and the other set is coloured noise/white noise. The subject has to point out in which set the stimuli differ. The subject only has to detect the difference, without the need of a reference signal. A disadvantage of this method is that it is more time consuming. It can be tiresome for the subject to listen to two more stimuli when (s)he is certain there is (no) difference in the first set of stimuli.

B.2.2 The up/down-strategy

With the up/down-strategy the strength of the coloration is altered step by step, as a function of the perception of coloration by the subject, for instance by altering the gain g of the repetition with respect of the direct signal. The strategy starts at maximum coloration which is then subsequently decreased until the subject can no longer hear the difference between coloration and white noise. Then the strength of the coloration is increased, until the subject again perceives coloration, then the step is reversed again, and so on.

![Graph](image)

Fig. B.1 2AFC Up/down-method, two correct: step down, one incorrect: step up

The experiment starts at a gain of 0 dB (maximum coloration), and a large step is taken until the first reversal has to be made. Then the gain level steps back to the former level, and the step size is decreased. After a certain number of reversals, the experiment is stopped. The threshold is calculated from the gain levels which are passed after the third reversal. If this method is used in combination with a two-
alternative forced-choice, the level of coloration is adjusted according to the correctness of the answers. The subject has to give two or more correct answers before the coloration level is decreased, while one incorrect answer increases the coloration level again, an example is given in Fig. B.1a. For an extended description of this method the reader is referred to Levitt (1971). Additional information on thresholds and bias estimators of this method can be found in Green (1990) and Schlauch and Rose (1990).

In Tab. B.1 the experiment information for the up/down strategy is given in detail.

**Tab. B.1  Experiment information for the up/down-strategy**

<table>
<thead>
<tr>
<th>four interval 2 AFC up/down-strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong># Stimuli</strong></td>
</tr>
<tr>
<td>2 sets of two stimuli:</td>
</tr>
<tr>
<td>- white white</td>
</tr>
<tr>
<td>- white coloured</td>
</tr>
<tr>
<td>- white coloured</td>
</tr>
<tr>
<td>- white white</td>
</tr>
<tr>
<td><strong>Question</strong></td>
</tr>
<tr>
<td>In which set do the stimuli differ: 1 or 2?</td>
</tr>
<tr>
<td><strong>Stimuli repeated</strong></td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td><strong>Stimulus Duration</strong></td>
</tr>
<tr>
<td>set 1: gate 650 ms open, 200 ms silence, gate 650 ms open</td>
</tr>
<tr>
<td>set 2: gate 650 ms open, 200 ms silence, gate 650 ms open</td>
</tr>
</tbody>
</table>

**Experimental Set-Up**

The experiment starts at 0 dB attenuation, the step is 4 dB, in about 6 steps the level is near threshold. At one incorrect answer the attenuation is reduced, for two correct answers the attenuation is increased. After the first reversal the step is set to 2 dB. After four reversals follow 15 level changes before the experiment is halted, if the fourth reversal is not reached, the experiment is halted after 25 level changes. The threshold is the averages of all levels after the fourth reversal. A threshold measurement is conducted three times, the final result is averaged over these three measurements.

**B.2.3  The trajectory method**

In the up/down strategy the (gain) levels which are presented are determined by the correctness of the given answers. Many measurements are made at some levels, while at others only a few. Also, the next level will be near the former level, so the subject knows what level he can expect at next trial. Especially near the threshold this can be tiresome. A more sophisticated method is the trajectory method, which is an adaptation of the method described by Lopes Cardozo, (1962). In this method the experiment leader determines beforehand at what levels the subject has to do the trials, and present these levels randomly to the subjects. In Tab. B.2 the possible number of correct and incorrect answers that can be given with the 'stop criterion', with probability whether the level is above the threshold.

The levels have to be chosen carefully, so that the threshold will be near the middle level: as many levels below and above the threshold. At each level a number of correct/incorrect answers has to be given before it is certain that at that level the subject can or cannot detect difference. The number of trials per level is such, that the
maximum number of trials per level is 12 and that the majority of the trials is concentrated around the threshold.

Tab. B.2 Stop criterion for trajectory method

<table>
<thead>
<tr>
<th>% correct</th>
<th>incorr.</th>
<th>total</th>
<th>Probability level above threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>100.0</td>
<td>0</td>
<td>5</td>
<td>0.835</td>
</tr>
<tr>
<td>97.5</td>
<td>1</td>
<td>8</td>
<td>0.714</td>
</tr>
<tr>
<td>81.8</td>
<td>2</td>
<td>11</td>
<td>0.621</td>
</tr>
<tr>
<td>75.0</td>
<td>3</td>
<td>12</td>
<td>0.436</td>
</tr>
<tr>
<td>66.7</td>
<td>4</td>
<td>12</td>
<td>0.238</td>
</tr>
<tr>
<td>63.6</td>
<td>4</td>
<td>11</td>
<td>0.196</td>
</tr>
<tr>
<td>60.0</td>
<td>4</td>
<td>10</td>
<td>0.158</td>
</tr>
<tr>
<td>55.6</td>
<td>4</td>
<td>9</td>
<td>0.125</td>
</tr>
<tr>
<td>50.0</td>
<td>4</td>
<td>8</td>
<td>0.098</td>
</tr>
<tr>
<td>42.9</td>
<td>4</td>
<td>7</td>
<td>0.075</td>
</tr>
<tr>
<td>33.3</td>
<td>4</td>
<td>6</td>
<td>0.057</td>
</tr>
<tr>
<td>20.0</td>
<td>4</td>
<td>5</td>
<td>0.042</td>
</tr>
<tr>
<td>0.0</td>
<td>4</td>
<td>4</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Fig. B.2 Example of the trajectory method

two trajectories, with its likelihood

In Fig. B.2 an example of an experiment is given, namely the percentage of correct answers of the total number of trials at certain level. With the aid of a maximum likelihood method the threshold is determined: for each interval the probability is calculated that the interval is above the threshold for given correct percentage. With these probabilities the likelihood as a function of the interval is calculated, from which the thresholds follows.

Because the number of trials at a certain level are independent of the results at another level, the trials are not committed to a certain sequence. Per trial different levels or different signals can be presented to the subject. Consequently, a small difference between two threshold can be easily detected, for instance the difference in the thresholds of coloration for harmonic and anharmonic cosine noise (see § 2.5.1). A drawback of this method is that the experiment leader must have some prior knowledge about the level of the threshold.

In Tab. B.3 the experimental set-up for this method is described in detail.
Tab. B.3 Experiment information for the trajectory method

<table>
<thead>
<tr>
<th>Four interval trajectory method</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># Stimuli</td>
<td>2 sets of two stimuli:</td>
</tr>
<tr>
<td></td>
<td>- white white white coloured</td>
</tr>
<tr>
<td></td>
<td>- white coloured white white</td>
</tr>
<tr>
<td>Question</td>
<td>In which set do the stimuli differ: 1 or 2?</td>
</tr>
<tr>
<td>Stimuli repeated</td>
<td>No</td>
</tr>
<tr>
<td>Stimulus Duration</td>
<td>set 1: gate 650 ms open, 200 ms silence, gate 650 ms open, 650 ms silence</td>
</tr>
<tr>
<td></td>
<td>set 2: gate 650 ms open, 200 ms silence, gate 650 ms open</td>
</tr>
</tbody>
</table>

Experimental Set-Up
The maximum and minimum coloration levels are determined beforehand, symmetric around the expected threshold, the levels are 2 dB apart. The coloration levels at which the trials are presented, are randomised. A threshold measurement is conducted three times for one measurement, the threshold is determined by a likelihood method.

B.2.4 The adjustment method

In both the up/down strategy and the trajectory method the measurement is based on forced choice experiments and the subject is not free to decide at which level the coloration reaches the threshold. With the up/down method the subject can give correct answers by chance and is tired by the number of stimuli he has to react to without hearing the difference. The same problem applies to the trajectory method, if the trajectory is not chosen carefully. To avoid this problem, the adjustment method can be used in which the subjects can search for the threshold themselves. Two stimuli are presented to the subject, a white stimulus and a coloured stimulus, in random order. The subject can either increase or decrease the white noise level in order to find the level at which the coloration is just detectable. The subject is asked to step round his threshold, from detection of coloration to no detection of coloration. Just as in the up/down method, several reversals between 'detection/no detection' (like in §B.2.2 correct/incorrect) are made. This experiment can start at maximum coloration, after which the subject has to increase the white noise level until no coloration is perceived (top-down) or start at maximum white noise level, followed by decreasing the white noise level until coloration is detected (bottom-up). In both cases, the initial step with which the white noise level changes is randomly chosen and not revealed to the subject, so the subject does not know how many steps need to be taken before the threshold is reached. Every experiment is repeated once. Again, the threshold is averaged from the mean level after the fourth reversal from both experiments. In Tab. B.4 the experimental set-up for the adjustment method is described in detail.
Appendix B

For signals with a sharp threshold and with experienced subjects this type of experiment is easy, accurate and very fast when compared to up/down-methods. Drawback of this method is (in situations with fuzzy thresholds or inexperienced subjects) that the experiment leader is not certain whether the subject has reached his threshold or only thinks he has. These thresholds tend to be too low. More information on estimating the threshold with the adjustment method can be found in Green (1993).

Tab. B.4 Experiment information for the adjustment method

<table>
<thead>
<tr>
<th>Adjustment method</th>
</tr>
</thead>
<tbody>
<tr>
<td># Stimuli</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Question</td>
</tr>
<tr>
<td>Stimuli repeated</td>
</tr>
<tr>
<td>Stimulus Duration</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Experimental Set-Up</td>
</tr>
<tr>
<td>The experiment start at maximum coloration, attenuation 0 dB. The step is chosen randomly from 4/5/6/7/8 dB, so that the threshold is not reached after a fixed number of steps. The attenuation increases if the subject indicates that he hears difference. After a reversal the step is halved, with a step minimum of 2 dB. The step minimum is chosen thus, that only 1 or 2 steps are necessary to step from 'difference perceptible' to 'no difference perceptible'</td>
</tr>
<tr>
<td>The experiment is ended after 15 reversals. The threshold is the average of all levels after the fourth reversal. A threshold measurement is conducted two times, the final result is the average over the three measurements.</td>
</tr>
</tbody>
</table>

B.3 Scaling of coloration by comparison

B.3.1 Paired-comparison

Even if the strength of the coloration of two signals as revealed by the threshold level is comparable, the nature of the colour of both signals can be quite different: the threshold does not provide an unambiguous description of all qualities of the coloration. With the ‘paired-comparison’ method different qualities can be compared. In these experiments all signals of a set are compared pairwise and the subject has to judge on coloration. If the coloration is comparably strong, the subject will use other qualities of the colour when rating the signals, for instance the presence of rattle, or high pitched coloration. Using this method the signals can be arranged on a subjective scale of coloration.

The experimental procedure is as follows. The subject starts the experiment with an introduction to the signals: all the signals are presented, preceded by white noise. In the experiment itself all signals are compared to each other, and to white noise. Two
signals are presented each time, and the subjects is asked to indicate which signal is considered the most coloured signal and the amount of difference. The subject can repeat the combination of signals as often as needed to make a good comparison. There are four categories for the amount of difference:

- almost no difference
- moderate difference
- reasonable difference
- large difference

Note that there is an even set of categories, to avoid that the subject chooses for the middle category for most of the time. The subject is forced to make a choice: no difference is not a category. Two signals can differ in timbre, but if the coloration of the first signal is as strong as for the second signal, the answer should be in the first category. The two answers result in one value: if signal j is compared to signal k, the answer \( a_{j,k} \) is \( \pm 1, \pm 2, \pm 3 \) or \( \pm 4 \). If signal j is the most coloured signal \( a_{j,k} \) is positive, otherwise \( a_{j,k} \) is negative. Both combinations \( j,k \) and \( k,j \) are presented to the subject. If the subject is consistent in his answers, the answer will be anti symmetric: \( a_{j,k} = -a_{k,j} \). The mean answer is averaged over several measurements \( i \) and the combinations \( j,k \) and \( k,j \), leading to the scale value \( s_i \) (compare Thurstone, 1927). The results can be arranged from minimum to maximum coloration, by the scale value \( s_j \), where the smallest \( s_j \) is the least coloured signal with the lowest coloration. When white noise is included in the scaling it usually has the smallest value, but in case signals with very weak coloration are scaled, the smallest result can belong to another signal.

Tab. B.5 Experiment information for the paired-comparison

<table>
<thead>
<tr>
<th>Paired-comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td># Stimuli</td>
</tr>
<tr>
<td>Question</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Stimuli repeated</td>
</tr>
<tr>
<td>Stimulus Duration</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Experimental Set-Up</td>
</tr>
</tbody>
</table>

B.3.2 Multidimensional scaling

With the paired-comparison the subject compares each time two stimuli and is asked to concentrate on coloration only when judging which signal appears as most coloured. This is not an easy task, since the question assumes coloration to be an one-dimensional sensation. But coloration has more than one aspect, namely timbre
Appendix B

and pitch and a determination of these aspects can be determined by using a multi-dimensional scaling method. For a detailed description of multidimensional scaling the reader is referred to Kruskal (1964a, 1964b), Carroll and Chang (1970) and Vos (1994). For the measurements described in this thesis (§ 4.5, from Vos 1994), the subject is presented with three stimuli and is asked to point out which two stimuli of these three are most alike and which two differ most. The subject is free to choose any aspect of the sound when comparing the stimuli.

In short, the multidimensional scaling can be explained as follows. The number of the dimensions is, in first approach, equal to n-1, where n is the number of stimuli. To test whether there are relations between the number of dimensions is reduced. This is done by the Kruskal method, which states that the dissimilarities and the distances between the stimuli are monotonically related. The ‘goodness of fit’ of this monotonically relation, is calculated by the residual sum of squares and is called the ‘stress’. The amount of stress of the scaling shows whether the relationship between the results can be represented correctly in this reduced number of dimensions. The stress is given as a percentage: a zero stress indicates a perfect relation between the number of dimensions and relations, while a stress of 10% is still acceptable.

A configuration found with the Kruskal method is based on the sum dissimilarity matrices of all subjects. In order to do right to all individual results, the final results are elaborated in an individual scaling (see Carroll and Chang, 1970), in which a ‘condition space’ shows how a subject values the dimensions. In a two dimensional case, a subject may value the first dimension for 30% and the second dimension for 70%, then obviously the second dimension is more important to the subject than the first dimension.

Tab. B.6 Experiment information for the multidimensional scaling

<table>
<thead>
<tr>
<th>Multidimensional scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong># Stimuli</strong></td>
</tr>
<tr>
<td><strong>Question</strong></td>
</tr>
<tr>
<td>2. which pair of stimuli the least dissimilar</td>
</tr>
<tr>
<td><strong>Stimuli repeated</strong></td>
</tr>
<tr>
<td><strong>Stimulus Duration</strong></td>
</tr>
<tr>
<td>stimulus 2: gate 1300 ms open</td>
</tr>
<tr>
<td>at first presentation: 700 ms silence, stimulus 3: gate 1300 ms open</td>
</tr>
<tr>
<td><strong>Experimental Set-Up</strong></td>
</tr>
</tbody>
</table>

The most difficult aspect of multidimensional scaling is to interpret the dimensions: which quality of the signal matches which dimension. Or, stated differently, if two
signals are on the same axis, in which aspect do these signal match. A method to
determine this is to perform a paired-comparison with an explicit instruction for the
subjects to pay attention to one particular aspect of the signal only. The results of the
paired-comparison are then correlated with the dimensions of the multidimensional
scaling. A high correlation (near one) then reveals that the aspect under considera-
tion corresponds to this particular dimension.

B.4 Experimental Environment

The experimental set up is as follows. The signals are generated or presented with a
Digital Signal Processor (DSP 16 of Ariel). The signals are presented to the subjects
by means of a headphone. The subject is seated in a soundproof booth and can in-
teract with the experimental program through a terminal. The artificial signals were
generated with the DSP, using separate analog white noise sources (Chapter 2 and
3). The white noise was filtered between 20 Hz and 10 kHz before it was used as in-
put of the DSP. The output of the DSP was also filtered between 20 Hz and 10 kHz.
The signals which were recorded in concert halls (Chapter 4) are downloaded in and
presented with the DSP, the output is filtered between 20 Hz and 8 kHz, this because
of the frequency reach of the sources used in the hall recordings. All signals were
presented to the subjects at 60 dB SL.

Tab. B.7 Symbols and information on the subjects

<table>
<thead>
<tr>
<th>symbol</th>
<th>initials</th>
<th>experience</th>
<th>gender</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>○</td>
<td>MS</td>
<td>+</td>
<td>♀</td>
<td>20..30</td>
</tr>
<tr>
<td>▼</td>
<td>BO</td>
<td>0/+</td>
<td>♂</td>
<td>30..40</td>
</tr>
<tr>
<td>◊</td>
<td>FB</td>
<td>++</td>
<td>♀</td>
<td>50..60</td>
</tr>
<tr>
<td>△</td>
<td>JR</td>
<td>++</td>
<td>♂</td>
<td>40..50</td>
</tr>
<tr>
<td>+</td>
<td>WK</td>
<td>++</td>
<td>♀</td>
<td>20..30</td>
</tr>
<tr>
<td>□</td>
<td>ES</td>
<td>0/+</td>
<td>♀</td>
<td>20..30</td>
</tr>
<tr>
<td>□</td>
<td>DB</td>
<td>0</td>
<td>♀</td>
<td>20..30</td>
</tr>
<tr>
<td>□</td>
<td>PJ</td>
<td>0</td>
<td>♀</td>
<td>20..30</td>
</tr>
<tr>
<td>●</td>
<td>IG</td>
<td>0/+</td>
<td>♀</td>
<td>20..30</td>
</tr>
<tr>
<td>●</td>
<td>LP</td>
<td>0/+</td>
<td>♀</td>
<td>20..30</td>
</tr>
</tbody>
</table>

→ Averaged over the individual subjects
Appendix B

In the experiment more or less experienced subjects took part. The less experienced subjects took a training on recognising coloration with an up/down-method before taking part in an adjustment method or a paired-comparison test. For the last type of test experienced subjects are a prerequisite, because the subject is free to choose and the experiment leader cannot verify whether the subject reacts on the right cue. In this thesis often the results of the individual subjects are presented in order to stress the individual differences. If no other legend is given with the results, the symbols denote the subjects, where each subject has his or her individual symbol. The symbols are given in Tab. B.7, together with other information on the subjects: initials, experience (o: average, +: good), age and gender. The latter can be important in some auditory task (Roberts and Sanders, 1992).
Appendix C

Manikin Recordings and Coloration

In Chapter 4 the results are presented of the strength of coloration measured for concert hall recordings. The coloration perceived in these recordings (made with a manikin such as KEMAR and presented afterwards in the listening experiments by a headphone) is a good approximation of the sound as it is perceived in the hall itself. Nevertheless, the recording and presentation techniques may add coloration to the signal, that is not present in the hall. The coloration that is perceived by the subject when listening to the recording is the sum of the coloration of the source (which is also perceived in the hall), the coloration of the hall (this is the point of interest) and the coloration that is added by KEMAR and the headphone. The coloration of KEMAR is present in all recordings and also the coloration of the headphone is a constant factor, but this additional coloration should be avoided if possible since it can distract from the coloration of the concert hall itself.

The coloration from the sources results from the imperfection of the source. The source consist of two loudspeakers, one for the low frequencies ('RCF-Espace') and an omnidirectional sound source for the mid-range frequencies. The latter loudspeaker was 'custom made'. Both loudspeakers overlap at 250 Hz and the frequency range is from 70 Hz to 8 kHz. The recorded signals are band filtered between these boundaries. The omnidirectionality of the mid-range loudspeaker is good up to 2 kHz, above which the level of the signal becomes direction dependent. Therefore the recording of the signal should be done in the free field situation so that the effect of the directionality of the signal is reduced.

The coloration which KEMAR adds to the signal results from the form of its head and body. KEMAR has microphones at the end of each ear canal, at the location of the ear drum in the human head. The artificial head and ear are made to resemble the human head and ears in order to simulate the distortion of the sound fields by the human head and ears. This form of coloration is experienced as quite natural, it gives this quality to a recording which misses from a recording made with microphones only. The coloration can be used by the listener to gain extra information about the source, such as the elevation with respect to the listener (Blauert, 1983). An additional distortion is caused by the ear canal: the frequencies round 2700 Hz are made more prominent with respect to the other frequencies (up to 15 dB, see Burkhard and Sachs, 1975 and Killion, 1979), see Fig. C.1a.
Fig. C.1  Transfer function: white noise presented to KEMAR with a Beyer D770 headphone, not corrected (a) and corrected (b) for ear canal distortion

Fig. C.2  Filter for ear canal distortion

Fig. C.3  Frequency response of white noise signal recorded in (a) the non-reverberant room and (b) the reverberant room
Hence, the ear canal adds extra coloration. The headphone used in the experiments (Beyer D770) is outside the ear canal and does not correct for this coloration (this headphone does not have a flat eardrum-pressure frequency). This results in a signal that at the eardrum of the listener has twice the effect of the ear canal coloration: from the ear canal of KEMAR (from the recording) and of the ear canal of the listener. This problem could be avoided by placing the microphones at the beginning of the ear canal, at the outside of the ear. Since the recordings have been made with the microphones at the ear drums, the extra coloration has to be removed by filtering the recorded signals before presenting them to the subjects. The filter which has been used (a custom made filter, Fig. C.2) removes most of the effect of the ear canal, this coloration is no longer prominent. The coloration of the ear canal which is not removed by the filter is present in all recordings and a constant factor when comparing the signals. The transfer function of the head phone used in the experiment, Beyer D770, is not straight. When white noise is presented to KEMAR with this headphone and the signal is corrected for the ear canal distortion, the remaining signal is not white as it should be, see Fig. C.1b. The transfer function is band filtered between 70 Hz and 8 kHz like the hall recordings. The dip in the spectrum at 5 kHz must be attributed to the headphone since this dip is not present in the recordings.

When comparing the hall signals on coloration it would be convenient to have a reference signal. An ideal reference signal embeds the coloration due to the source and KEMAR. When comparing a hall recording to this reference signal with the same headphone the difference in coloration is the coloration added by the hall. This signal is a better reference than a white noise signal, since white noise will exaggerate the coloration added by the halls since the coloration of KEMAR and the source is still present in the signal. The recording of this ideal reference signal should be made in a room which generates no coloration (no reflections or frequency-dependent absorption) and wherein a diffuse field condition can be made with the source that is used in the hall recordings. The first demand requires a recording in a non-reverberant room, but in a non-reverberant room no diffuse field can be created using the experimental source. A diffuse field can only be created in a reverberant room.

In Fig. C.3a the power spectrum of the signal recorded in the non-reverberant room is given. This situation is extreme: the sound comes from one direction, so the distortion of the non-omnidirectionality of the source is important and also the reflections of the head and ear lobe of KEMAR come from only one direction. The coloration due to the source and KEMAR is maximum. In Fig. C.3b the power spectrum is given of a recording made in an extremely reverberant room. The sound comes from many directions and the distortions of the source and reflections from KEMAR are diminished with respect to the non-reverberant room. In this power
spectrum the coloration of the reverberant room itself plays an important role. Both power spectra are not corrected for the ear canal distortion.

The coloration of both the recordings in the non-reverberant room and of the reverberant room is too strong to be suitable as a reference signal. Therefore, in case a reference signal is needed, white noise is used although is not the ideal reference signal.
Appendix D

Harmonic Cosine Noise Convolved with Auditory Filters

The aim of this appendix is to calculate the auditory filtering (after Patterson and Moore, 1986) of harmonic cosine noise. For white noise the solution of the integral is:

\[ |S_c(f_c)|^2 = \int_{-\infty}^{\infty} \left(1 + \frac{4|f-f_c|}{W(f_c)}\right) \exp\left(-\frac{4|f-f_c|}{W(f_c)}\right) df = W(f_c) \]  \hspace{1cm} D.1

with

\[ W(f_c) = \left(6.23f_c^2 + 93.39f_c + 28.52\right) \times 10^{-3} \text{ kHz} \]  \hspace{1cm} D.2

For harmonic cosine noise spectrum the auditory filtered (and normalised) spectrum is:

\[ |S_c(f_c)|^2 = \frac{\int C(f,f_c)(1+\cos(2\pi f T)) df}{\int C(f,f_c) df} \]  \hspace{1cm} D.3

thus:

\[ |S_c(f_c)|^2 = 1 + \frac{1}{W(f_c)} \int_{-\infty}^{\infty} \left(1 + \frac{4|f-f_c|}{W(f_c)}\right) \exp\left(-\frac{4|f-f_c|}{W(f_c)}\right) \cos(2\pi f T) df \]  \hspace{1cm} D.4

The integral of Eq. D.4 is equal to:

\[ 2\cos(bf) \int_{x=0}^{\infty} (1 + ax) \exp(-ax) \cos(bx) dx \text{ with } a = \frac{4}{W(f)} \text{ and } b = 2\pi T \]  \hspace{1cm} D.5

For the calculation of integral of D.5 the following definite integrals are used (Carmichael and Smith, 1931, Eq. 338 and 339):

\[ \int_{0}^{\infty} \exp(-ax) \sin(bx) dx = \frac{b}{a^2 + b^2} \]  \hspace{1cm} D.6
\[ \int_0^\infty \exp(-ax) \cos(bx) \, dx = \frac{a}{a^2 + b^2} \]  

D.7

With the rules:

\[ \int_a^b f'(x)g(x) \, dx = \left[ f(x)g(x) \right]_a^b - \int_a^b f(x) g'(x) \, dx \]  

with \( g(x) = x \) and \( f'(x) = \exp(-ax) \cos(bx) \)

it can be shown that the following applies:

\[ \int_0^\infty (1 + ax) \exp(-ax) \cos(bx) \, dx = \frac{2a^3}{(a^2 + b^2)^2} \]  

D.8

If this is combined with Eq. D.4, the result is (see Eq. 5.15):

\[ |S_c(f_c, T)|^2 = 1 + \frac{\cos(2\pi Tf_c)}{\left( 1 + \frac{1}{2} \pi TW(f_c) \right)^2} \]  

D.9
Appendix E

The Calculation of the ‘Internal’ Autocorrelation

The spectrum of a (coloured) signal is ‘convolved’ with auditory filters in order to calculated the internal spectrum. The internal autocorrelation is the Fourier transform of the internal spectrum. When calculating the internal autocorrelation the interesting part of the signal spectrum is the part which originates from the reflections. As an example, the autocorrelation of one reflection, ‘convolved’ with auditory filters is calculated (see Eq. 5.15):

\[ |S_c(f_c)|^2 = \left[ 1 + \left[ 0.5\pi T W(f_c) \right]^2 \right]^{-2} \cos(2\pi f_c T) \]  

E.1

The autocorrelation is the Fourier transform of this function:

\[ \varphi_c(\theta) = \int_{-\infty}^{\infty} \left[ 1 + \left[ 0.5\pi T W(f_c) \right]^2 \right]^{-2} \cos(2\pi f_c(\theta - T)) df_c \]

E.2

with the bandwidth: \( W(f_c) = (6.23 f_c^2 + 93.39 f_c + 28.52) \times 10^{-3} \) kHz

The integral over this peak (important when deriving the B0-criterion) is calculated as:

\[ \int_{-\infty}^{\infty} \varphi_c(\theta) d\theta = \int_{-\infty}^{\infty} \left[ 1 + \left[ 0.5\pi T W(f_c) \right]^2 \right]^{-2} \cos(2\pi f_c(\theta - T)) df_c = \]

\[ \int_{-\infty}^{\infty} \left[ 1 + \left[ 0.5\pi T W(f_c) \right]^2 \right]^{-2} \delta(f_c) df_c = \left[ 1 + \left[ 0.5\pi T W(0) \right]^2 \right]^{-2} \]

E.3

The exact form of the autocorrelation as given by Eq. E.2 cannot be derived analytically, but a close approximation is:

\[ \varphi_a(\theta) = f(T) \left( 1 + \frac{\alpha |\theta - T|}{T} \right) \exp \left( -\frac{\alpha |\theta - T|}{T} \right) \]

E.4

where \( \alpha \) is a constant to be determined later, and \( f(T) \) is calculated by demanding that the approximation and the exact expression, Eq. E.2, agree in the integral sense, that is Eq. E.3 applies for the approximation as well:
\[ \int_{-\infty}^{\infty} \varphi_a(\theta) d\theta = \left[ 1 + (0.5\pi T W(0))^2 \right]^{-2} \]

For the approximate expression the integral is calculated as follows:

\[ \int_{-\infty}^{\infty} \varphi_a(\theta) d\theta = f(T) \int_{-\infty}^{\infty} \left( 1 + \frac{\alpha |\theta - T|}{T} \right) \exp \left( - \frac{\alpha |\theta - T|}{T} \right) d\theta = \frac{4T}{\alpha} f(T) \]

If Eq. E.5 and Eq. E.6 are combined the result is:

\[ f(T) = \frac{\alpha}{4T} \left[ 1 + (0.5\pi T W(0))^2 \right]^{-2} \]

The value of \( \alpha \) is chosen such that the top value of the approximation agrees with that of the exact expression, Eq. E.2. Numerical evaluation of the latter value yields \( \alpha = 31.7 \) approximately, and the approximation expression of \( \varphi_c(\theta) \) becomes:

\[ \varphi_a(\theta) = \frac{\alpha}{4T} \left[ 1 + (0.5\pi T W(0))^2 \right]^{-2} \left( 1 + \frac{\alpha |\theta - T|}{T} \right) \exp \left( - \frac{\alpha |\theta - T|}{T} \right) \]

In Fig. E.1 the approximation \( \varphi_a(\theta) \) is compared to the numerically derived autocorrelation of \( \varphi_c(\theta) \).

\[ \begin{align*}
\text{Fig E.1} & \quad \text{Approximation of the autocorrelation and the numerical calculated autocorrelation for } T=2 \text{ ms (a) and } T=10 \text{ ms (b)}
\end{align*} \]
For multiple reflections, the internal autocorrelation is the sum of the autocorrelation of the internal spectra which originate from the separate repetitions:

\[
\varphi_a(\theta) = \sum_{i=0}^{N-1} \sum_{j=i+1}^{N} \frac{\alpha/4}{(T_i - T_j)} \left[ 1 + 0.5\pi(T_i - T_j) W(0) \right]^{2} \left[ 1 + \frac{\alpha |\theta - T_i + T_j|}{(T_i - T_j)} \right] \exp \left( -\frac{\alpha |\theta - T_i + T_j|}{(T_i - T_j)} \right)
\]

Comparison of Eq. E.5 and Eq. E.8 reveals that the area under the autocorrelation \( \varphi_a(\theta) \) is proportional to the maximum of the autocorrelation, multiplied with the location of the maximum \( (\theta_{max} = T, c = 4/\alpha) \):

\[
\int_{-\infty}^{\infty} \varphi_a(\theta) d\theta = c \theta_{max} \varphi_a(\theta_{max}) = c \left( \theta \varphi_a(\theta) \right)_{max}
\]

The relation of E.10 will play an important role for the autocorrelation of signals with multiple reflections (§ 5.7).

In Fig. E.2 shows the maximum of the autocorrelation, multiplied with the location of the maximum for the approximation \( \varphi_a(\theta) \), which equals the area under the autocorrelation \( \varphi_c(\theta) \), and the numerically derived \( \left( \theta \varphi_c(\theta) \right)_{max} \). It shows both the accuracy of the of the approximation of \( \varphi_a(\theta) \), as the assumption that E.10 also applies for \( \varphi_c(\theta) \).

Fig. E.2  The maximum of the autocorrelation, multiplied with the location of the maximum for the approximation \( \varphi_a(\theta) \) and for the numerically derived \( \varphi_c(\theta) \)
References


Bremer, T.B.J. (1992), De toonhoogtesprong bij de overgang van harmonische naar anharmonische kamfilterruis, masters thesis, Delft University of Technology (unpublished)


Cardozo, B.L. (1962), Een sequente-keuze methode voor het bepalen van psychofysieke drempelwaarden, I.P.O. rapport nr. 22, Eindhoven

References


Chernyak, R.I. and Dubrovsky, N.A. (1968), Pattern of the noise images and the binaural summation of loudness for different interaural correlation of noise, Proceedings 6th Int. congress on acoustics, vol. 1, Tokyo, A-3-12


Haas, H. (1951), Uber den Einfluss eines Einfachechos auf die Horsamkeit von Sprache, Acustica, 1, 49-58


Keulen, W. van, (1990) Personal communications


140
Kruskal, J.B., (1964a), Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika, 29, no 1
Kruskal, J.B., (1964b), Nonmetric multidimensional scaling: a numerical method, Psychometrika, 29, no 2
Meerling (1981), Methoden en technieken van psychologisch onderzoek, Drukkers drukkerijen bv, Meppel
Raatgever, J. (1989), Binauraal horen gebaseerd op spectrale patrononherkenning?, NAG journaal 97, 15-26

141
References


Thurstone L. L., (1927), Law of comparative judgement (case V), *Psychological Review*, 34, 273-286

Vos, I.M., (1994), The perception of coloration at different places in a concert hall, analysed with multidimensional scaling techniques, masters thesis, Delft University of Technology (unpublished)


COLORATION AND BINAURAL DECOLORATION
OF SOUND DUE TO REFLECTIONS

Summary

The object of the research presented in this thesis is to provide insight in the psycho-
physical aspects of coloration and binaural decoloration of sound signals. In this
thesis the colour of a sound is described as that attribute with which a listener can
calculate that two sounds similarly presented and having the same loudness are
dissimilar. Accordingly, the coloration is defined as an audible distortion which alters
the (natural) colour of the sound. A major originator for coloration are reflections
which interfere with the direct sound, and in this research emphasis is put in
measuring the relation between reflections and coloration, for both synthetic DSP
generated signals and signals recorded in concert halls.

Binaural decoloration is defined as a decrease in coloration when listening binaurally
compared to monaural listening. This means that when the left and right ear signals
differ (dichotic listening situation), the coloration is less than when both signals are
equal (diotic listening situation) or when the signal is presented to one ear only.
Binaural decoloration will occur, for instance, if reflections at left and right ear differ in
delay times with respect to the direct signal. The main goal was to find a relation
between the amount of difference between left and right ear signals and the amount
of binaural decoloration.

In this research various experimental methods were used, namely threshold
measurements which give an absolute value of the strength of the coloration and
comparison methods which give a relative scale for the subjective prominence of the

coloration. In the threshold methods we can distinguish the ‘forced choice’ method in
which a subject has to indicate an absolute difference between two stimuli, and the
adjustment method in which the subject is free to adjust the stimulus to the coloration
threshold. These methods are applied to both the synthetically generated coloured
signals and to the signals recorded in concert halls.

From various experiments in this research it appeared that the pitch sensation plays
an important role in the coloration appreciation. A strong pitch sensation is usually
accompanied by a low threshold. In the multidimensional scaling, where the
coloration is analysed in its underlying components, it appears to be the dimension
which (on average) the subjects attach the most importance to. If the pitch sensation is less clear, the timbre of the signal starts playing a role.

The results found in these experiments were used to test prediction criteria, such as the $A_0$- (spectral) and $B_0$- (autocorrelation) criterion. These criteria are an adaptation of those defined by Atal et al., 1962. Atal defined the $A_0$-criterion with the maximum modulation depth of the short-time spectrum, on grounds of the limited memory of the hearing organ. The present $A_0$-criterion is defined with the maximum modulation depth of the internal spectrum of the signal, i.e. the signal filtered with the auditory filters which are present in the cochlea. The original $B_0$-criterion was related to the maximum of the autocorrelation of the short time spectrum, in the present model this criterion is related to the area of the dominant peak of the autocorrelation of the internal spectrum. These criteria give satisfactory results for the synthetic generated diotic signals, the $A_0$-criterion gives good results for the signals recorded in the concert halls.

The modelling of the binaural decoloration was attempted by using the Central Spectrum-model (CS-model, Raatgeber and Bilsen, 1986). This model gives the interaction of left and right signal as a pattern of neural activity (Central Activity Pattern, CAP), which is considered as central spectra depending on the internal delay. These central spectra are scanned by higher order processes in the brain. As a criterion for the coloration for interaurally correlated signals the deviation of a coloured CAP from a diotic white noise CAP, the $D_0$-criterion, was defined. This model gave reasonable results for most of the dichotic signals, for diotic signals the results were less satisfying. This can be a consequence of the fact that the deviation model uses all the differences between a coloured CAP and a white CAP to predict the threshold, so lateralisation differences as well as coloration differences. The CAP can give a fair indication of the decoloration effect, but not exactly predict the amount of the decoloration. In view of the results for complementary cosine noise, future models should include monaural channels, which provide the 'higher order processes' with the left and right signals directly. How the information of the monaural channels and the central activity pattern is combined to form a total percept of coloration should be an important theme for future investigations.

A. M. Salomons
KLEURING EN BINAURALE ONTKLEURING VAN GELUID TEN GEVOLGE VAN REFLECTIES

Samenvatting

Het in dit proefschrift beschreven onderzoek heeft als doel inzicht te geven in de psychofysische aspecten van kleuring en binaurale ontkleuring van geluidssignalen. In dit proefschrift is de kleur van een geluid omschreven als dat kenmerk waarmee de luisteraar kan beoordelen dat twee geluiden, op dezelfde manier gepresenteerd en met dezelfde luidheid, verschillen. Overeenkomstig is de kleuring gedefinieerd als een hoorbare verstoring van de natuurlijke kleur van het geluid. Een belangrijke veroorzaker van kleuring zijn reflecties die met het directe geluid interfereren. In dit onderzoek is nadruk gelegd op het meten van de relatie tussen reflecties en kleuring, voor zowel synthetische signalen gegenereerd met een DSP, als voor signalen opgenomen in concertzalen.

Binaurale ontkleuring is gedefinieerd als een afname in kleuring wanneer binauraal geluisterd wordt, ten opzichte van monauraal luisteren. Dit betekent dat wanneer linker en rechter signaal verschillen (dichotische luistersituatie) de kleuring minder is dan wanneer beide signalen gelijk zijn (diotische luistersituatie), of indien het signaal aan slechts één oor wordt aangeboden. Binaurale ontkleuring zal bijvoorbeeld optreden als de reflecties aan linker- en rechter oor verschillen in vertragingstijd ten opzichte van het directe signaal. Het hoofdthema van het onderzoek was een relatie te vinden tussen het verschil tussen linker- en rechteroor signalen en de mate van binaurale ontkleuring.

In dit onderzoek zijn verschillende experimentele methoden gebruikt, namelijk drempelmetingen die een absolute waarde geven aan de sterkte van kleuring, en vergelijkingsmethoden die een relatieve schaal geven aan de kleuringssterkte. De drempelmethoden kunnen worden opgedeeld in ‘gedwongen keuze’ methode, waarin een proefpersoon een absoluut verschil tussen twee stimuli aan moet geven, en in een instelmethode waarbij de proefpersoon vrij is de stimulus in te stellen op het drempelniveau. Deze meetmethoden zijn toegepast op zowel de kunstmatig gegenereerde gekleurde signalen, als op de signalen opgenomen in concertzalen.

Uit verschillende experimenten in dit onderzoek is naar voren gekomen, dat de toonhoogtegewaarneming een belangrijke rol speelt bij de kleuringsbeoordeling. Een
sterke toonhoogtegewaarwording gaat meestal gepaard met een lage kleuringsdrempel. Bij de multidimensionale schaling, waarbij de kleuring in zijn onderliggende componenten wordt ontleed, blijkt het de dimensie te zijn waarbij (gemiddeld gezien) de proefpersonen het meeste gewicht aan geven. Bij minder duidelijke toonhoogtegewaarwording blijkt de timbre van het signaal een rol te gaan spelen.

De resultaten gevonden met de drempel-methoden zijn gebruikt om predictie criteria te testen, het \( A_0 \)-(spectraal) en \( B_0 \)-(autocorrelatie) criterium. Deze criteria zijn een aanpassing van de criteria gedefinieerd door Atal et al, 1962. Atal defineerde het \( A_0 \)-criterium met de maximum modulatie diepte van het beperkte tijd-spectrum, op grond van het beperkte geheugen van het gehoororgaan. Het huidige \( A_0 \)-criterium is gedefinieerd met de maximum modulatie diepte van het interne spectrum van het signaal, d.w.z. het signaal gefilterd met de auditieve filters die in de cochlea aanwezig zijn. Het originele \( B_0 \)-criterium was gerelateerd aan het maximum van de autocorrelatie van het beperkte tijd-spectrum, in het huidige model is dit criterium gerelateerd aan de oppervlakte van de dominante piek van de autocorrelatie van het interne spectrum. Deze aangepaste criteria geven bevriddigende resultaten voor de synthetisch gegenereerde diotische signalen, het \( A_0 \)-criterium geeft goede resultaten voor de signalen die opgenomen zijn in de concertzalen.

De modellering van de binaurale ontkleuring wordt gezocht in het Centraal Spectrum-model (CS-model, Raatgever en Blisen, 1986). Dit model geeft de interactie van linker- en rechterssignaal als een patroon van neurale activiteit (Centraal Activiteiten Patroon, CAP), dat wordt beschouwd als centrale spectra die afhangen van de interne vertraging. De centrale spectra worden door hogere orde processen in de hersenen afgetast. Als maat van kleuring voor interauraal gekorreleerde signalen werd de deviatie van het gekleurde CAP ten opzichte van diotische witte ruis CAP, het \( D_0 \)-criterium, gedefinieerd. Dit criterium gaf voor diotische signalen redelijke resultaten, voor diotische signalen waren de resultaten minder bevriddigend. Dit kan het gevolg zijn van het feit dat het deviatie model alle verschillen tussen een gekleurd en een wit CAP gebruikt om de drempel te voorspellen, dus zowel lateralisatieverschillen als kleuringsverschillen. Het CAP kan een goede indicatie geven voor het ontkleuringseffect, maar niet exact de hoeveelheid ontkleuring voorspellen. In het licht van de resultaten voor complementaire cosinusruis, zou in toekomstige modellen de monaurale kanalen moeten worden opgenomen, die de 'hogere orde processen' direct voorzien van het linker- en rechtersignaal. Hoe de informatie van de monaurale kanalen en het centrale activiteiten patroon gecombineerd wordt tot een totale waarneming van de kleuring zou een belangrijk thema kunnen zijn voor toekomstig onderzoek.

A. M. Salomons

146
Acknowledgements

I want to thank Frans Bilsen and Johan Raatgever, both for supporting the research and for being (sometimes more, sometimes less) enthousiastic subjects.

Thanks also go to Jan Potter, for all the labour he put in the recordings of the concert hall signals.

Further, I am very grateful for the electronics support given by Wim Willems and Wilfred van Wees.

I want to thank the students who participated in the project, particularly Ingrid Vos, with whom it was good to co-operate.

Also many thanks go to Bas, as subject, for his valuable comments on my manuscript and, last but not least, as supporting husband.

And I want to thank all my patient subjects, who undertook the sometimes labourious experiments.
Biography

Aurelia Maria Salomons was born in Alphen aan den Rijn, the Netherlands, on June 23rd, 1966. She received her M.Sc. degree in Physics from the Delft University of Technology in 1989. The work on which her masters thesis was based, was carried out in the field of spectroscopy of magnetic resonance signals. In August 1989 she worked at the Laboratory of Perceptual Acoustics (Delft University of Technology), in a Ph.D. research on the subject of the perception of coloration in sound signals.