The modeling of auditory roughness for signals with temporally asymmetric envelopes

Ronnie P.N. Duisters
25th March 2005
Preface

The hearing system has evolved through millions of years and has adjusted itself to comprehend and withstand most sounds in nature. The hearing system is a complex structure. Modeling such a structure and its perception includes many disciplines like biology, physiology, signal processing, psychology and psychoacoustics. It is not a big surprise that the exact working of the hearing system is not completely understood. What has taken ages to develop cannot be fully analysed in a few decennia.

One small aspect of the properties of the hearing system is analysed in this report. Roughness perception belongs to the field of psychoacoustics and defines a measure of unpleasantness due to modulations in a sound. There is a difference in perception of roughness for signals with temporally asymmetric envelopes, such as rolling sounds, and their time-reversed counterparts.

As an electrical engineering student entering the complex world of psychoacoustics I had the opportunity to largely expand my knowledge of the hearing system, a subject that has fascinated me ever since music became an important part of my life. I believe an audio engineer can profit most from enlarging his knowledge of the auditory system as this system is the only way a human being can perceive sound. The companies that encorporate psychoacoustic knowledge in their audio systems are always first in line to present new ‘magical’ products. I will not give any names, but when I mention brilliant sound from very small speakers and surround sound from just two speakers I think I have said enough.

I owe my gratitude to Dik Hermes, for being a source full of information in a broad range of disciplines, Armin Kohlrausch, for many difficult questions that led to numerous ‘Aha-Erlebnisse’ and Leo Vogten for being a safety net and, by observing the project on a higher level, for providing practical remarks. Vincent Jourdes provided a fantastic basis for my research and suited me with a rolling start. Koen Steegs did a great job in recording the sounds of rolling balls and I enjoyed our many discussions on a wide range of topics. My other collegae at the subdepartment of Human Technology Interaction and at the department of Electrical Engineering were indespensable in bringing the project to a good ending. I want to thank Toshio Irino, Marc Leman and Renato Nobili for providing me their expert opinion on some implementation details and for insights. Last, but not least, I want to thank my parents and my girlfriend Ineke for their mental support during the project.
Abstract

To some extent, we are able to make observations on the size and the speed of rolling objects by just listening to its sound. Some properties of the sound the rolling object indicate that the object is either large or small and either rolling fast or slow. Recent research [15] suggested that auditory roughness might be an important temporal cue in our ability to differentiate, but it could not be underlined. In order to verify the suggestion, we need to apply a good auditory roughness model on the sounds of rolling objects.

Sounds of rolling objects often have temporally asymmetric envelopes. Literature [37] shows that there is a significant difference in roughness between ramped and damped signals. These ramped and damped signals have temporally asymmetric envelopes and are the exact time-reversed counterpart of each other. Ramped signals have a slowly increasing onset and a steep offset, whereas damped signals have a steep onset and a slowly decreasing offset. Current roughness models are not able to distinguish the difference between the roughness of ramped and damped signals. In order to estimate the roughness of rolling sounds, our model should be able to distinguish the roughness difference between ramped and damped signals.

In this report, an auditory roughness model is developed consisting of four modules: a peripheral filter, an auditory filterbank, an adaptation module and a roughness extraction module. The latter three modules each contain multiple different implementations and the modules are tested on the difference they ascribe to ramped and damped signals. With all the implementations, twelve roughness models are constructed and they are subjected to the ramped and damped stimuli, as well as to a larger set of stimuli for which perceptual data is available. Finally, the roughness trends with respect to the size and speed of rolling balls are predicted.

The adaptation module appears to be the key module in explaining the roughness difference between ramped and damped signals. Furthermore, the twelve obtained roughness models are able to achieve a good fit with respect to the large set of available perceptual roughness data. Finally, our models predict that the roughness of rolling balls varies significantly with the size and speed parameters of the balls. With increasing speed, the estimated roughness increases and with increasing diameter, the estimated roughness decreases. Roughness may thus be a temporal cue in the perception of size and speed of rolling objects.
## Contents

1 Introduction .................................................. 1
   1.1 The concept of auditory roughness .......................... 1
   1.2 Signals with temporally asymmetric envelopes ............... 2
   1.3 Goal ...................................................... 3

2 Methods ....................................................... 4
   2.1 Modeling ................................................ 4
   2.2 Stimuli .................................................. 5
      2.2.1 Amplitude and frequency modulation .................. 5
      2.2.2 Ramped and damped signals ......................... 8
      2.2.3 Noise ............................................... 8
      2.2.4 Windowing ......................................... 9
   2.3 Perceptual data .......................................... 9
   2.4 Peripheral filtering .................................... 10

3 Auditory filtering ........................................... 12
   3.1 Filter bandwidth ....................................... 13
      3.1.1 Equivalent Rectangular Bandwidth (ERB) ............ 13
      3.1.2 Tonotopic scale ................................... 13
   3.2 Filter shape ............................................ 14
      3.2.1 Gammatone filter .................................. 14
      3.2.2 Gammachirp filter ................................ 17
   3.3 Auditory filtering and temporally asymmetric envelopes ..... 23

4 Adaptation .................................................... 25
   4.1 Meddis’ hair cell model .................................. 25
      4.1.1 Implementation .................................... 28
   4.2 Van Immerseel & Martens hair cell model ................... 29
      4.2.1 Implementation .................................... 30
4.3 Dau et al. adaptation model ........................................ 32
4.3.1 Implementation ............................................. 32
4.4 Adaptation and temporally asymmetric envelopes .............. 33

5 Roughness extraction module ........................................ 37
5.1 Aures’ roughness extraction model ................................ 37
5.1.1 Modifications .................................................. 42
5.2 Synchronisation index model (SIM) ............................... 44
5.2.1 Model I ......................................................... 45
5.2.2 Model II ......................................................... 47
5.3 Roughness extraction module and temporally asymmetric envelopes .... 48

6 Evaluation of the roughness models ................................ 49
6.1 Ramped and damped stimuli ...................................... 49
6.2 Larger dataset ..................................................... 51
6.2.1 Sinusoidally amplitude modulated sinusoids ................. 51
6.2.2 “Pseudo-AM” (pAM) sinusoids ................................ 54
6.2.3 Pure tones ...................................................... 56
6.2.4 Sinusoidally amplitude modulated noise .................... 58
6.2.5 Unmodulated bandpass-filtered white noise ................. 59
6.2.6 Sinusoidally frequency-modulated sinusoids ............... 60
6.3 Analysis .......................................................... 63
6.4 Rolling sounds ..................................................... 65

7 Conclusion .......................................................... 67

Bibliography .......................................................... 70

A Matlab code .......................................................... 75
A.1 Main program ...................................................... 75
A.2 Subfunctions ....................................................... 82
A.2.1 roughinit.m .................................................... 82
A.2.2 AdaptLevel.m ................................................ 83
A.2.3 rms.m .......................................................... 83
A.2.4 shiftcov.m ...................................................... 84
A.2.5 Other subfunctions ........................................... 85
A.3 Implementation example .......................................... 85
A.3.1 roughplotsasym.m ............................................ 85
B  Listing of used symbols  

88
from “Bernadette” by the Four Tops, bassline by James Jamerson
Chapter 1

Introduction

Nowadays, living in a modern society is an equivalent of living life in the fast lane, at least for most human beings. Visual and auditory cues are essential for people to make their way through the busy environment, although they mostly do not notice using them.

At the subdepartment of Human Technology Interaction of the Technische Universiteit Eindhoven, research is being conducted into how people make use of visual and auditory cues and how they perform if these cues are not as expected. The sort of auditory cues that are of interest here are the sounds of rolling objects. Houben [15] investigated the perception of rolling sounds, produced by rolling balls, with respect to the size and speed of the ball. To some extent, people are able to discriminate between sounds corresponding to different speeds or sizes of the rolling objects. One temporal aspect of the sound that appears to be important in this perceptual task is called ‘auditory roughness’.

1.1 The concept of auditory roughness

In principle, sound can be described by its signal properties in the time or frequency domain. These properties can be described mathematically in a well-defined way. This allows the calculation of spectra, spectrograms, temporal envelopes, etc. These mathematically defined signal properties must be distinguished from perceptually defined sound properties such as pitch and loudness. These perceptual attributes of a sound are the result of auditory information processing in the central nervous system. They have no simple relation with mathematically defined signal properties. Officially, according to the American National Standards Institute [1], there are four perceptual attributes of sound: pitch, loudness, duration and timbre.

Pitch is defined as “that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low”, loudness as “that intensive attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud” and timbre as “that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different”. Timbre can be subdivided into percepts such as roughness, sharpness and tonality.

When the envelope of a pure tone is varied with yet another sinusoidal signal with increasing frequency, different percepts occur. Varying the envelope of a sinusoid with another sinusoid is called ‘sinusoidal amplitude modulation’ and the frequency of the sinusoid in the enve-
Figure 1.1: Oscillogram of a rolling sound. The envelope of this signal is damped as it contains very steep onsets and slowly decreasing offsets.

lope is called ‘modulation frequency’. For very low modulation frequencies the sound will be perceived as a single sinusoid whose amplitude increases and decreases with the modulation frequency. At a modulation frequency of about 0.5 Hz, the human listener starts to order the beats, produced by the modulations, in units of two, three or four. This is called ‘rhythm perception’. This effect gradually turns into a percept of roughness. The beats cannot be separated in groups anymore and the modulations start to sound unpleasant. The unpleasantness has a maximum for a modulation depth between 20 and 70 Hz and then it decreases. Roughness is thus a measure for a kind of impurity or unpleasantness in a sound. Hermann von Helmholtz [14] was the first to address this issue and used the word ‘sensory dissonance’, to extend the concept of musical dissonance to non-musical sounds. Later, auditory roughness was systematically studied for stimuli such as amplitude-modulated tones [44], frequency modulated tones [23] and amplitude-modulated noise [9]. A computational model to estimate perceived roughness was developed by Aures [2] and optimised by Daniel & Weber [5]. This model was able to obtain a good fit to perceptual roughness data for the stimuli mentioned. Jourdes [22] preceeded this project and improved the Daniel & Weber model by applying more up-to-date components in the roughness model.

1.2 Signals with temporally asymmetric envelopes

Figure 1.1 shows the amplitude of a rolling sound plotted against time. The temporal envelope of this signal is asymmetric. Very abrupt onsets go over into more gradual offsets. If this signal is reversed in time, the resulting sound is perceived different than the original sound. This is essential in determining the roughness of rolling sounds as the rolling sound will probably be judged rougher than the time-reversed rolling sound.

The effect of this temporal asymmetry within the auditory system has been the topic of various research papers, e.g. [31], [32], [18] and [36]. In these papers, two stimuli with temporally asymmetric envelopes were defined. The first stimulus had a slowly increasing onset and a steep offset and was called ‘ramped’. The second stimulus had a steep onset and a slowly decreasing offset and was called ‘damped’. Throughout this report, this terminology will be used.
That there is a difference in roughness perception between signals with an asymmetric envelope and their time-reversed counterparts was demonstrated by Pressnitzer & McAdams [37]. They also constructed two stimuli with temporally asymmetric envelopes, similar to the ramped and damped stimuli mentioned earlier. However, Pressnitzer & McAdams defined their stimuli as ‘sawtooth’ and ‘reversed sawtooth’, respectively. The signals were constructed so that the amplitude spectrum of these two signals was identical. The only difference lay in their phase spectrums. The analytic description of the stimuli will be presented in section 2.2.2.

Current auditory models that are able to estimate perceived roughness are not able to distinguish between the roughness of ramped and damped signals.

1.3 Goal

The goal of this project is to learn more about the difference in roughness of ramped and damped signals and to create a model that to some extent can reproduce the subjective data by Pressnitzer & McAdams. The model will be based on knowledge about the auditory system and will contain the signal processing of the most important stages in the auditory periphery. As the auditory periphery contains multiple processing steps, the auditory model will contain multiple modules. For each module, a number of different implementations have been published. The impact on the difference in the roughness percept of ramped and damped signals for each implementation will be evaluated. The modules were either chosen because of their relation to physiology or their proven succesful contribution to auditory modeling.

The combinations of the module implementations will also be applied to a larger dataset of stimuli for which perceptual data are available, and compared with these corresponding subjective data. Finally, the model will be applied to high quality recordings of rolling balls to predict roughness trends for the size and speed of rolling balls.
Chapter 2

Methods

In this chapter, the basics of the project will be described. First of all, the construction of roughness model will be defined. Next, the stimuli used in the evaluation of the model are described analytically. An overview of available perceptual roughness data is presented and finally one of the modules of the roughness model is described in some detail.

2.1 Modeling

![Signal processing equivalent of the processing in the auditory system.](image)

Figure 2.1: Signal processing equivalent of the processing in the auditory system. The peripheral filter can be represented by a bandpass filter, the cochlear filter by a bank of bandpass filters, the adaptation process by half-wave rectification, automatic gain control and lowpass filtering and roughness extraction by a summation of partial roughnesses.

To model roughness perception, an auditory model was constructed in which four modules can be distinguished. These are represented by the dashed boxes in figure 2.1. The first component in the auditory model is the peripheral filter that approximates the signal processing of the outer ear and the middle ear. This can be implemented as a bandpass filter. The next two components, the auditory filterbank and the adaptation model, represent the signal processing of the cochlea and the auditory nerve. The auditory filterbank is a bank of bandpass filters that divides the wideband input signal into several smallband auditory channels. The adaptation model can be seen as an envelope extractor with an automatic gain control. Finally, the signals of the auditory nerve are processed by the brain. As far as roughness perception is concerned, this processing is approximated by a roughness extraction model that generates a roughness value \( R \). As roughness is related to modulation depth \( m \) according to equation (2.1):

\[
R \propto m^\alpha, \tag{2.1}
\]

the roughness extraction module estimates the modulation depth of the envelopes in the auditory channels. In the literature, constant value \( \alpha \) takes on values ranging from 1.2 to 2.
Table 2.1: Summary of used modules to form the auditory roughness extraction models. The reference column gives a reference to the literature that describes the implementation of the module in most detail.

<table>
<thead>
<tr>
<th>Module</th>
<th>Implementation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peripheral filtering</td>
<td>Pflueger et al.</td>
<td>[35]</td>
</tr>
<tr>
<td>Auditory filtering</td>
<td>Gammatone</td>
<td>[40]</td>
</tr>
<tr>
<td></td>
<td>Compressive gammachirp</td>
<td>[21]</td>
</tr>
<tr>
<td>Adaptation</td>
<td>Meddis</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>Van Immerseel and Martens</td>
<td>[17]</td>
</tr>
<tr>
<td></td>
<td>Dau et al.</td>
<td>[3]</td>
</tr>
<tr>
<td>Roughness extraction</td>
<td>Aures</td>
<td>[22]</td>
</tr>
<tr>
<td></td>
<td>Synchronisation Index Model (SIM)</td>
<td>[24]</td>
</tr>
</tbody>
</table>

Table 2.2: Available Matlab packages for the simulation of different stages in the hearing pathway.

The total roughness is constructed by the summation of all the partial roughnesses in the auditory channels. The exact roughness extraction process will be explained in chapter 5.

For each module, several implementations presented by different researchers were tested. The implementations are summarised in table 2.1. Some implementations were downloaded from the internet, verified and fitted to the general model. Table 2.2 is a reference for these sources.

2.2 Stimuli

To be able to understand the symbols and terminology in this report, the stimuli used in roughness perception are analytically described in this section. Subsequently, the sinusoidally amplitude modulated sinusoid, the frequency modulated sinusoid, the ramped and the damped signal will be described.

2.2.1 Amplitude and frequency modulation

Amplitude- and frequency-modulation techniques are often used when low-frequency information needs to be transported through a high-frequency channel. By changing a constant parameter of a sinusoid into a function, in this case the amplitude and frequency, respectively, the low-frequency information can be stored in the high-frequency sinusoid.

If the information is another sinusoid and the changed parameter is the amplitude, the resulting signal is called a sinusoidally amplitude-modulated (SAM) sinusoid. The high-frequency sinusoid is called the carrier and has a carrier or centre frequency $f_c$. The low-frequency sinusoid is called the modulator and has a modulation frequency $f_{\text{mod}}$. An SAM sinusoid has
the following general analytic description:

\[ s_{\text{SAM}}(t) = \{1 + m \cdot \sin (2\pi f_{\text{mod}}t + \phi_{\text{mod}})\} \cdot \sin (2\pi f_c t + \phi_c), \]  
\( (2.2) \)

where \( m \), \( 0 \leq m \leq 1 \) is the modulation depth, \( f_{\text{mod}} \) is the modulation frequency in Hz, and \( f_c \) is the centre frequency in Hz. Figure 2.2 shows an oscillogram of a SAM sinusoid with a centre frequency of 1 kHz, a modulation frequency of 70 Hz and a modulation depth of 1.

By omitting the phases and by applying the goniometric identity \( \sin(ax) \cdot \sin(bx) = \frac{1}{2} \cos((a-b)x) - \frac{1}{2} \cos((a+b)x) \), equation (2.2) can be expanded into

\[ s_{\text{SAM}}(t) = \sin (2\pi f_c t) + \frac{m}{2} \cos (2\pi (f_{\text{mod}} - f_c)t) - \frac{m}{2} \cos (2\pi (f_{\text{mod}} + f_c)t). \]  
\( (2.3) \)

Thus, in the frequency domain the SAM sinusoid has three components as shown in figure 2.3a.

A sinusoidally frequency-modulated (SFM) sinusoid can be described analytically as:

\[ s_{\text{SFM}}(t) = \sin \left(2\pi f_c t - \frac{\Delta f}{f_{\text{mod}}} \cdot \sin (2\pi f_{\text{mod}}t + \phi_{\text{mod}}) + \phi_c\right), \]  
\( (2.4) \)

where \( \Delta f \) is the frequency deviation in Hz. An oscillogram of an SFM sinusoid is given in figure 2.4. Note that the envelope of the signal is rather flat on this scale. The modulation index \( m_{\text{SFM}} \) of an SFM sinusoid is given by:

\[ m_{\text{SFM}} = \frac{\Delta f}{f_{\text{mod}}}. \]  
\( (2.5) \)

The derivation of the frequency characteristics of sinusoidally frequency-modulated sinusoids makes use of Bessel functions and is beyond the scope of this report. What can be said about the frequency spectrum of an SFM sinusoid is that it contains a infinitely many modulation components alongside the centre frequency, spaced by \( f_{\text{mod}} \) Hz. The number of relevant modulation components is dependent on the modulation index. Figure 2.3c shows an example of the amplitude spectrum of an SFM sinusoid where \( m_{\text{SFM}} < 1 \), and figure 2.3d shows an example of the amplitude spectrum of an SFM sinusoid where \( m_{\text{SFM}} > 1 \). The modulation components are not all in phase with the centre frequency component.
Figure 2.3: Amplitude characteristics of various signals in the frequency domain. a) an SAM sinusoid, \( m = 1 \); b) a ramped or damped signal (as their amplitude characteristics are identical), \( f_c = 5 \text{ kHz}, m = 1 \); c) an SFM sinusoid, \( m_{SFM} \leq 1 \); d) an SFM sinusoid, \( m_{SFM} > 1 \).

Figure 2.4: Time-domain representation of an SFM sinusoid, \( f_c = 1.6 \text{ kHz}, f_{mod} = 70 \text{ Hz}, \Delta f = 800 \text{ Hz} \).
Table 2.3: Parameters for the ramped and damped stimuli.

<table>
<thead>
<tr>
<th>centre frequency</th>
<th>( N_{\text{mod}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 kHz</td>
<td>2</td>
</tr>
<tr>
<td>5 kHz</td>
<td>4</td>
</tr>
<tr>
<td>10 kHz</td>
<td>7</td>
</tr>
</tbody>
</table>

2.2.2 Ramped and damped signals

As mentioned in the previous chapter, the ramped and damped signals as used in this report have been defined in Pressnitzer & McAdams [37]. A pure-tone carrier is amplitude modulated by a tone complex \( E_{\text{ramped}}(t) \):

\[
E_{\text{ramped}}(t) = \sum_{v=1}^{N_{\text{mod}}} -\frac{1}{N_{\text{mod}}} \sin(2\pi v f_{\text{mod}} t),
\]

in which the frequency components all fit within one Equivalent Rectangular Bandwidth around the centre frequency: \( N_{\text{mod}} \cdot f_{\text{mod}} \leq \frac{4}{3} \text{ERB}(f_c) \). The exact meaning of this condition will be clarified in chapter 3. Table 2.3 gives the values of the three used centre frequencies \( f_c \) and the corresponding number of components \( N_{\text{mod}} \). The ramped signal \( s_{\text{ramped}}(t) \) can now be constructed by amplitude modulating a carrier signal by the components of \( E_{\text{ramped}}(t) \). To obtain a maximum modulation depth of 1 for \( m = 1 \), the signal \( E_{\text{ramped}}(t) \) is normalised by dividing the total signal by its maximum.

\[
s_{\text{ramped}}(t) = -\left(1 + m \cdot \frac{E_{\text{ramped}}(t)}{\max(E_{\text{ramped}}(t))}\right) \cdot \sin(2\pi f_c t).
\]

A damped signal can be generated by time-reversing the ramped signal:

\[
s_{\text{damped}}(t) = s_{\text{ramped}}(-t).
\]

The damped signal can also be constructed by omitting the minus signs in equations (2.6) and (2.7). In figure 2.5 a ramped and damped signal with different centre frequencies are shown. It is clear to see that a higher centre frequency leads to a flatter ramp and damp because the signal consists of more modulation components.

2.2.3 Noise

White noise is generated by a function that delivers random values from a normal distribution. The frequency spectrum of this white noise signal is flat. Sinusoidally amplitude-modulated (SAM) noise is constructed in a similar fashion as the SAM sinusoid:

\[
s_{\text{SAM noise}}(t) = \{1 + m \cdot \sin(2\pi f_{\text{mod}} t)\} \cdot s_{\text{noise}}(t),
\]

where \( s_{\text{noise}}(t) \) is the white noise signal. All other parameters are the same as for the SAM sinusoid.
2.2.4 Windowing

Steep signal onsets can cause large transients in the model as well as in the hearing system. To dampen the effect of the signal onset, a squared sinusoid of 20 ms was applied as a fade in at the beginning of all stimuli:

\[ s_{\text{fade in}}(t) = \sin^2(\frac{1}{2} \pi \frac{1}{0.02} t). \]  

To dampen the effects of a too abrupt offset, the time-reversed version of the fade in was applied to the last 20 ms of the stimuli as a fade out.

2.3 Perceptual data

An overview of publications published over the years 1977–1999 is given in table 2.4. In these studies, the parameters of the stimuli as described in section 2.2 were varied systematically and the perceived roughness was determined accordingly.

The amount of perceived roughness can be fit to a perceptual scale. The roughness scale is mostly given in asper [48], a Latin word meaning rough. The roughness perceived for a sinusoidally amplitude modulated sinusoid with a centre frequency \( f_c \) of 1 kHz, a modulation frequency \( f_{\text{mod}} \) of 70 Hz, a modulation index \( m \) of 1 and a level \( L \) of 60 dB SPL is defined as 1 asper. Some experiments were conducted relative to a different reference signal, and plotted in percentage. Yet other experiments make use of the method of paired comparisons. Pairs of stimuli are presented and the observer has to judge which one has the higher roughness. These results can be scaled according to a statistical theory and for two experiments in table...
2.4 the Bradley-Terry-Luce (BTL) theory ([7], [8]) was used.

### 2.4 Peripheral filtering

Although the peripheral filter is an essential module in the total roughness model, it is not presented in a separate chapter. The peripheral filter describes the signal processing of the outer- and middle-ear chain. Due to the construction of the outer ear and the ear canal, resonance frequencies occur at which the ear is most sensitive. The middle-ear bones, hammer, anvil and stirrup, also limit the range of hearing due to mechanical limitations. For our purpose, the peripheral filter is a fixed bandpass filter which attenuates low frequencies and also high frequencies and has a resonance frequency at 4 kHz. An approximation of the transfer function of the peripheral filter is given by the combination of a high- and a lowpass filter in Pflueger et al. [35]. The highpass filter is described by:

\[
H_{\text{HP}}(z) = \frac{z^2 - 2z + 1}{z^2 - 1.9z + 0.9025},
\]  

(2.11)

The lowpass filter is defined by:

\[
H_{\text{LP}}(z) = \frac{0.109z^7 (z + 1)}{z^8 - 2.5359z^7 + 3.9295z^6 - 4.7532z^5 + 4.7251z^4 - 3.5548z^3 - 0.9876z + 0.2836}.
\]  

(2.12)

The cascade of the two filters is a linear time-invariant IIR filter that can be implemented in the discrete-time domain. A plot of the amplitude characteristic is given in figure 2.6.

This initial spectral weighting filter will be part of all model variations tested in chapter 6. In the following three chapters, the three remaining modules are treated in detail.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Parameters</th>
<th>Scale</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM sinusoids</td>
<td>(f_{\text{mod}}, f_c) (m, f_c) (L)</td>
<td>asper \asper\ relative in %</td>
<td>[2], [5]</td>
</tr>
<tr>
<td>pAM sinusoids</td>
<td>(\phi_c, f_c)</td>
<td>BTL</td>
<td>[37]</td>
</tr>
<tr>
<td>Pure tones</td>
<td>(f_c)</td>
<td>asper</td>
<td>[28]</td>
</tr>
<tr>
<td>SFM sinusoids</td>
<td>(f_{\text{mod}}) (\Delta f) (L)</td>
<td>relative in % \relative in % \relative in %</td>
<td>[2], [5]</td>
</tr>
<tr>
<td>White noise</td>
<td>bandwidth, (f_c)</td>
<td>asper</td>
<td>[2], [5]</td>
</tr>
<tr>
<td>SAM noise</td>
<td>(f_{\text{mod}}) (m) (L)</td>
<td>relative in % \relative in % \relative in %</td>
<td>[9]</td>
</tr>
<tr>
<td>Ramped and damped signals</td>
<td>(m, f_c)</td>
<td>BTL</td>
<td>[37]</td>
</tr>
</tbody>
</table>

Table 2.4: Available experimental data.
Figure 2.6: Amplitude characteristic of the transfer function of the peripheral filter by Pflueger et al. [35].
Chapter 3

Auditory filtering

The first model component related to the cochlea is the auditory filterbank. It represents the effective signal processing of the basilar membrane. The auditory filterbank can be seen as a bank of bandpass filters that divides the very often broadband input signal into multiple narrowband output signals. Figure 3.1 shows the principle of auditory filtering. As a first approximation, these filters can be considered as linear. Later, we will discuss some non-linear characteristics. Two properties of linear filters are relevant: the amplitude characteristic and the phase characteristic as a function of frequency.

The bandwidths and the shape of the auditory filters have been derived from perceptual masking experiments. The analytic description of the shape of the auditory filters has improved through the years, now incorporating more phenomena such as upward spread of masking. Two approximations of the auditory filter, the gammatone and gammachirp filter, will be described in this chapter. In section 3.3 the effect of the filterbank on asymmetric signals will be compared for the gammatone and the gammachirp filterbank.

Figure 3.1: The principle of the filterbank: a series of bandpass filters covering the audible range. Each approximately parabolic-shaped curve represents the transfer characteristic of a filter. On the linear frequency axis of this outline, the bandwidth of the filters is constant over frequency. In the model of the auditory filterbank this is not the case.
3.1 Filter bandwidth

3.1.1 Equivalent Rectangular Bandwidth (ERB)

On a linear scale, the bandwidth of the auditory filters increases with frequency. If these filters were rectangular, i.e., ideal bandpass filters, and if they were placed successively without overlap along the frequency axis, approximately 40 filters would cover the range of audible frequencies from 20 Hz to 20 kHz. These filter numbers are defined in the Equivalent Rectangular Bandwidth rate scale, in which the number of successive auditory filters is related to frequency. In fact, the human basilar membrane has about 3500 inner hair cells, each of which can be considered as an auditory filter. Hence, the basilar membrane can be viewed as a filterbank of a large number of strongly overlapping bandpass filters.

The ERB-rate scale belongs to the field of psychoacoustics and is derived using masking experiments with notched noise [29]. The principle is to apply white noise with a notch around a certain centre frequency. A pure tone is presented at the centre frequency of the notch and its masked threshold is measured for a range of notch widths. From the relation between threshold and notch width, the shape, and the ERB of the filter can be derived. Equation (3.1) shows the relation between centre frequency \( f_c \) (in Hz) and the corresponding filter bandwidth ERB in Hz:

\[
\text{ERB}(f_c) = 24.7 \cdot \left( \frac{4.37 \cdot f_c}{1000} + 1 \right).
\] (3.1)

The inverse function often comes in handy and is, therefore, presented in equation (3.2). The corresponding centre frequency in Hz for a given auditory filter bandwidth can be obtained by applying

\[
f_c = \frac{1000}{4.37} \cdot \left( \frac{\text{ERB}}{24.7} - 1 \right).
\] (3.2)

3.1.2 Tonotopic scale

As a scale for looking at the frequency representation the normal choice would be to either use a linear or a logarithmic representation. On both of these scales, however, the hair cells in the cochlea will not be spaced equidistantly and, in probable correspondence, the auditory filter bandwidth is not constant. For an understanding of the kind of signals that will arrive at the auditory nervous system, an idea of the auditory filter bandwidth characterising the distance between frequencies can help. This idea leads to a frequency scale, on which the inner hair cells are equidistantly spaced. This is also referred to as a tonotopic frequency scale.

It appears that this tonotopic scale is maintained in the higher centres of the auditory information processing throughout the nervous system. Literature provides figures of inner hair cell numbers \( N_{\text{hair cells}} \) in the human cochlea that range from about 3000 to 4000. In Dallos [4] a figure of approximately 3500 can be found which will be taken as a reference here. The length of the average human basilar membrane \( l_{\text{cochlea}} \) is approximately 34.8 mm. Equation (3.3), taken from Greenwood [12], relates distance from the apex \( x \) in mm to best frequency on the basilar membrane:

\[
f_c(x) = A_1 \cdot \left( 10^{\lambda x} - B_1 \right),
\] (3.3)
Figure 3.2: Relation between position on the basilar membrane, corresponding frequency and the width of the auditory filters at that place. The abscissa represents the position along the basilar membrane, with 0 mm corresponding to the apex. The continuous line relates place in mm and frequency in Hz, and the vertical lines represent the bandwidths of the auditory filters at that place, according to the Equivalent Rectangular Bandwidth values for the audible range.

where \(A, a\) and \(k\) are constants that can be fit to the basilar membrane of a certain species. For the human basilar membrane, \(A_1 = 165, \lambda = 0.06\) and \(B_1 = 1\) appear to be a good fit.

Assuming hair cells are placed homogeneously along the basilar membrane, the spatial distance between hair cells in the cochlea is

\[
\Delta x = \frac{l_{\text{cochlea}}}{N_{\text{hair cells}}} = \frac{34.8}{3500} \approx 0.01 \text{ mm}. \tag{3.4}
\]

An advantage of this kind of scale is that, as mentioned, also the ERB-bandwidth is approximately spaced equidistantly. Simply compare equation (3.3) with (3.5) and verify the similarity. Figure 3.2 shows a plot in which place, ERBs and frequency are combined.

Equation (3.5) demonstrates how the centre frequency of the corresponding auditory filter can be derived from the ERB-rate:

\[
f_c = \frac{1000}{4.37} \cdot \left(10^{\frac{\text{ERB}_{\text{rate}}}{21.46}} - 1\right). \tag{3.5}
\]

### 3.2 Filter shape

#### 3.2.1 Gammatone filter

Here we will present the gammatone filter as a first approximation of an auditory filter. In Hartmann [13] it is given that a gammatone filter has an impulse response of

\[
g_{\ell}(t) = a t^{n-1} e^{-2\pi b \text{ERB}(f) t} e^{j(2\pi f_c t + \phi)}. \tag{3.6}
\]
Chapter 3. Auditory filtering

Figure 3.3: Amplitude characteristic of the gammatone filter, \( f_c = 2 \text{ kHz} \), \( n = 4 \), \( b = 1.019 \). The continuous line shows the amplitude characteristic of the analytical gammatone filter and the dashed line shows the amplitude characteristic of the implementation of the gammatone filter used in this study [40].

This function can be seen as a pure tone with a gamma function as an envelope, hence the name gammatone. In equation (3.6), \( a \) is a normalisation factor, \( n \) is an integer mostly set to 4, representing the order of the filter and \( \phi \) is a phase factor.

The corresponding frequency-domain expression, the Fourier transform of the impulse response, gives the response characteristics of the filter in the frequency domain:

\[
G_t(f) = \frac{a \Gamma(n) e^{j\phi}}{(2\pi b \text{ERB}(f_c) + j2\pi(f - f_c))^n}.
\] (3.7)

Figure 3.3 shows the amplitude characteristic of the gammatone filter, where \( b \) is chosen to be 1.019. The phase characteristic of the gammatone filter is given in figure 3.4. A plot in which several filters are combined in a filterbank can be found in figure 3.5. The centre frequencies are spaced 2 ERBs apart to avoid visual cluttering.

Implementation

Gammatone filterbanks have been implemented by various researchers. A thorough description of the implementation of the gammatone filter by Slaney can be found in [40] and [22]. This implementation, used in our model, is taken from Slaney’s Auditory Toolbox (see table 2.2 for a reference). It is an FIR filter, constructed of 4 cascaded second-order filters.

The implementation was verified for its gain response which is plotted in figure 3.3. What catches the eye is the difference in response at the right hand side of the plot. The gain is less than -56 dB at that point, but this discrepancy should be kept in mind.
Figure 3.4: Phase characteristic of the gammatone filter, \( f_c = 2 \text{ kHz}, n = 4, b = 1.019 \).

Figure 3.5: Amplitude characteristic of a number of gammatone filters of a gammatone filterbank. The centre frequencies of the filters presented in this figure are spaced by 2 ERB.
3.2.2 Gammachirp filter

A closer approximation of the auditory filter is the gammachirp ([18]–[21], [34] and [45]). In contrast with the gammatone, in which the sinusoid has a constant frequency, the gammachirp contains a frequency modulation, hence the name chirp. The impulse response of the gammachirp is:

$$g_c(t) = at^{n-1}e^{-2\pi b\text{ERB}(f)t}e^{j(2\pi f_c t + c \ln t + \phi)}.$$  (3.8)

The only difference with the gammatone is the frequency modulation term $e^{jc\ln t}$. In the complete model the parameter $c$ depends on the input level of the input signal. Its operation is that of a compressive non-linearity. Consequently, a filterbank composed of gammachirp filters is indicated as compressive gammachirp filterbank. The parameter $c$ can generate level dependency and, corresponding with that, signal compression.

The frequency transformed version can be expressed as

$$G_c(f) = \frac{a \Gamma(n + j c)e^{j \phi}}{(2\pi b\text{ERB}(f_r) + j2\pi(f - f_r))^{n+jc}},$$  (3.9)

in which it is assumed the $c$ changes over time.

**Analysis**

To achieve an implementable version of the gammachirp filter and to be able to map the parameters of the filter to known experimental data, Irino and Unoki rewrote equation (3.9) in [20] as follows

$$G_c(f) = \frac{\tilde{a}}{(2\pi \sqrt{b^2 + (f - f_r)^2} \cdot e^{jn\theta})^{n+jc}} \cdot \frac{1}{\{2\pi \sqrt{b^2 + (f - f_r)^2})^{j c \cdot e^{-c \theta}}.$$  (3.10)

where

$$\tilde{a} = a \Gamma(n + j c)e^{j \phi},$$

$$\tilde{b} = b \text{ERB}(f_r),$$

$$\theta = \arctan \left( \frac{f - f_r}{b} \right).$$

If $\tilde{a}$ is chosen to be 1, equation (3.10) can be rewritten as

$$G_c(f) = G_t(f) \cdot H_A(f),$$  (3.11)

and the gain function is given by

$$|G_c(f)| = |G_t(f)| \cdot |H_A(f)| = \frac{1}{(2\pi \sqrt{b^2 + (f - f_r)^2}^{n}) \cdot e^{c \theta}}.$$  (3.12)
To arrive at the compressive gammachirp filter, the exponent of the asymmetric function \( |H_A(f)| = e^{c\theta} \) must be adjusted:

\[
c\theta = c_1\theta_1 + c_2\theta_2,
\]

where

\[
\theta_1 = \arctan \left( \frac{f - f_{r1}}{b_1 \text{ERB}(f_{r1})} \right) \quad \text{and} \quad \theta_2 = \arctan \left( \frac{f - f_{r2}}{b_2 \text{ERB}(f_{r2})} \right).
\]

The amplitude characteristic of the gammachirp filter can now be rewritten in terms of a level-independent asymmetric filter \( |G_{ca}(f)| \) instead of the symmetric gammatone filter \( |G_t(f)| \), multiplied with a level-dependent asymmetric function \( e^{c_2\theta_2} \):

\[
|G_{c}(f)| = |G_t(f)| \cdot e^{c_1\theta_1} \cdot e^{c_2\theta_2} = |G_{ca}(f)| \cdot e^{c_2\theta_2}. \tag{3.15}
\]

In equation (3.15) \( |G_{ca}| \) represents the passive gammachirp, i.e. the filter shape when the sound-induced wave travels passively along the basilar membrane. The application of \( e^{c_2\theta_2} \) makes it possible to model the influence of active processes in the cochlea on the filter shape. The level dependency is related to this asymmetric function. Because the gammachirp filterbank is compressive, in other words the filter gain decreases when signal levels increase, the filterbank is a nonlinear element. Figure 3.6 visualises the construction of this compressive gammachirp filter from the gammatone filter.

To achieve the desired filter shapes, \( f_{r1} \) must be set to the centre frequency of the filter and \( L \) corresponds to the level of the signal at the input of the filterbank. Irino suggested to use the parameters from table 3.1, as taken from [34]\(^1\). After applying these parameters to the equations below, all conditions have been satisfied and the gammachirp filterbank can be generated:

\[
f_{rat} = 0.573 + 0.0101 \cdot L, \tag{3.16}
\]

\[
f_{p1} = f_{r1} + \frac{c_1 b_1 \text{ERB}(f_{r1})}{n}, \tag{3.17}
\]

\[
f_{r2} = f_{rat} \cdot f_{p1}. \tag{3.18}
\]

The level dependency of the gammachirp filterbank is shown in the figures 3.7 and 3.8, the phase response is given in figure 3.9, and the asymmetrical shape of the filters within the

\(^1\)In the paper, the data can be found in Table I, the last row.
filterbank can be seen in figure 3.10. In the latter figure it becomes clear that, with respect to the gammatone filter, frequencies below the resonance frequency are attenuated less than frequencies above the resonance frequency. This feature can explain the upward spread of masking phenomenon. Figure 3.10 also shows that at 60 dB SPL, the gain of the resonance frequency increases towards low frequencies. This was already found in figure 3.7, where it becomes clear that a constant resonance gain for all auditory channels is only achieved for a level of approximately 42 dB SPL.

Implementation

The implementation of the gammachirp filterbank consists of a combination of the Slaney gammatone filterbank cascaded with two asymmetric compensation filters. It is important to note that parameter $b$ in the implementation of the gammatone filter must be replaced with parameter $b_1$ of the gammachirp filter.

The minimum-phase IIR asymmetric compensation filters $H_c(z)$ are constructed by a cascade of 4 second order filters $H_{cv}(z)$.

$$H_c(z) = \prod_{v=1}^{4} H_{cv}(z),$$  \hspace{1cm} (3.19)

where

$$H_{cv}(z) = \frac{(1 - r_v e^{j\varphi_v} z^{-1})(1 - r_v e^{-j\varphi_v} z^{-1})}{(1 - r_v e^{j\varphi_v} z^{-1})(1 - r_v e^{-j\varphi_v} z^{-1})},$$  \hspace{1cm} (3.20)

$$r_v = e^{-v_0 2 \pi \text{ERB}(f_r) f_s},$$  \hspace{1cm} (3.21)

$$\varphi_v = \frac{2\pi \{ f_r - p_0 v^{-1} p_2 c \text{ERB}(f_r) \}}{f_s},$$  \hspace{1cm} (3.22)

$$\phi_v = \frac{2\pi \{ f_r + p_0 v^{-1} p_2 c \text{ERB}(f_r) \}}{f_s}.$$  \hspace{1cm} (3.23)

To get an implementation with a small error with respect to the analytical asymmetric compensation filter, the following parameters were chosen ([45], p. 427): $p_0 = 2$, $p_1 = 1.5363 - 0.2531 \cdot |c|$, $p_2 = 0.3437 - 0.0099 \cdot |c|$, and $p_3 = 0.2342 - 0.0144 \cdot |c|$. The transfer function of the implemented version of the gammachirp filter can be seen in figure 3.11. When comparing this image with the analytical result, again the right hand side shows the greatest difference. This is due to the fact that the gammatone implementation deviates significantly from the real gammatone towards high frequencies.

If the level parameter $L$ is set to 80 dB SPL, an instability occurs in the first ERB channel. It appears that the asymmetric compensation filter becomes unstable. This problem was acknowledged but not solved during the project. Whenever this condition occurred, the unstable channel was excluded from the roughness calculation.

Important to note is that the implementation in the model, as used by us, does not estimate the input level. The input level of the signal is a parameter of the roughness model and
Figure 3.6: Visualisation of the construction of the compressive gammachirp filter, $f_c = 2$ kHz. The top panel shows the amplitude characteristic of the gammatone filter $|G_t|$ and the gains of the asymmetric compensation filters $|H_{a1}|$ and $|H_{a2}|$. The bottom panel shows the gain of the passive gammachirp filter $|G_{ca}| = |G_t| \cdot |H_{a1}|$ and the gain of the compressive gammachirp filter $|G_{cc}| = |G_{ca}| \cdot |H_{a2}|$ for a number of different input levels.
Figure 3.7: Gain at the resonance frequency of the compressive gammachirp filter in dB as a function of input level in dB for several centre frequencies. The plot was derived from the analytic description in equation (3.15).

Figure 3.8: Gain of the compressive gammachirp filter for several levels, $f_c = 2$ kHz, derived from the analytic description in equation (3.15). This plot is the same as $|G_{cc}|$ in the bottom panel of figure 3.6, but in this plot the levels are explicitly given. The gain of the filter decreases with increasing level, whereas the asymmetry increases with increasing level.
Figure 3.9: Phase characteristic of the compressive gammachirp filter, $f_c = 2$ kHz, $L = 70$ dB.

Figure 3.10: Amplitude characteristic of a number of gammachirp filters of a gammachirp filterbank. The centre frequencies of the filters presented in this figure are spaced by 2 ERB. The input level $L$ is 60 dB. The gain increase at the resonance frequency towards low frequencies can be explained by observing figure 3.7.
during the current simulation for a specific signal the filterbank parameters are fixed, including the parameter $L$. From psychoacoustics it is known that this is not correct. The filters instantaneously follow input level changes and change shape and bandwidth accordingly. The implementation of such an algorithm for the compressive gammachirp filter is currently not available, but should become available in the future. This would be a valuable addition to the described roughness model.

3.3 Auditory filtering and temporally asymmetric envelopes

To evaluate the effect of the two described filters on the asymmetry of the ramped and damped stimuli, the stimuli by Pressnitzer & McAdams are re-evaluated. The stimuli are applied to the filters corresponding to their centre frequencies. The following rules are used to estimate the modulation depth of the signal after filtering.

The envelope $\tilde{s}(t)$ of signal $s(t)$ is obtained by calculating the Hilbert transform of the signal $H\{s(t)\}$ and applying equation (3.24):

$$\tilde{s}(t) = |s(t) + j \cdot H\{s(t)\}| = \sqrt{s^2(t) + (H\{s(t)\})^2}.$$  \hspace{1cm} (3.24)

The modulation depth $m$ is estimated by dividing the root-mean-square of the envelope minus the mean of the envelope $\tilde{s}(t)$ by the mean of the envelope:

$$m = \sqrt{2} \cdot \frac{\sqrt{\tilde{s}(t) - \tilde{s}(t)}}{\tilde{s}(t)}.$$  \hspace{1cm} (3.25)
As shown, the fraction needs to be scaled by a factor $\sqrt{2}$. This estimation method is chosen because of the relation between roughness and modulation depth as depicted in equation (2.1) and because the method is closely related to one of the roughness extraction modules that will be presented in chapter 5.

Calculated are the differences in modulation-depth estimation between the damped and the ramped stimuli:

$$\Delta m = m_{\text{damped}} - m_{\text{ramped}}.$$  

(3.26)

Because the damped signal attributes to a larger roughness percept than the ramped signal, a positive value of $\Delta m$ would imply a positive influence to the roughness estimation for the filterbank. A negative value results in a smaller modulation depth for the damped stimulus than for the ramped stimulus which would conflict with the attribution to roughness.

Both the gammatone and gammachirp filter have a negligible effect on the difference between the roughness of the ramped and damped stimuli. The difference is less than 1 % with respect to the modulation depth of the ramped and damped signal.
Chapter 4

Adaptation

The second part of the modeling of the cochlea addresses the role of the hair cells and their connections with the auditory nerve fibres in the Organ of Corti. There are two types of hair cells, the inner and the outer hair cells. The task of the hair cells is: converting the motion of the basilar membrane into electrical signals in the auditory nerve with as good a dynamic range and response as possible. The conversion from mechanical movement into electrical signals is known as the transduction process.

A part of this process is adaptation. Hair cells generate faster responses to onsets and slower responses to offsets of sounds. Another part of the process is dynamic range compression: the human ear can handle sounds ranging from whispering to loud pop concerts with a good comprehensibility for both. The signal processing of the inner hair cells can be modeled by half-wave rectification followed by lowpass filtering and automatic gain control.

4.1 Meddis’ hair cell model

Inner hair cells follow the movement of the basilar membrane at their position and elicit electrical pulses in the auditory nerve accordingly. In rest, the nerve fibres fire spontaneously.

Figure 4.1: Schematic of an inner hair cell.
Figure 4.2: Conversion of the signal from the auditory filterbank $s_{af}(t)$ to the signal of the permeable membrane $s_{hwr}(t)$ in the Meddis hair cell model. $A_2$, $B_2$ and $\beta$ are constants.

with a rate of about 20–50 spikes per second. The probability of this firing increases proportionally with basilar membrane motion amplitude, but only for movements in one direction. This is a process that can be described as half-wave rectification superimposed on a constant term, representing the spontaneous rate.

The inner hair cell model by Meddis as described in [25], [26] and [27] is based on physiological principles. A schematic representation of an inner hair cell can be found in figure 4.1. This is a simplification of a real hair cell which has more afferent fibres connected to the bottom. The model consists of three neurotransmitter reservoirs that are assumed to exist in the hair cell, and describes the amount of neurotransmitter in the reservoirs and in the synaptic cleft as a function of time. Two block diagrams of this model are shown in figures 4.2 and 4.3.

The first figure describes how the motion of the basilar membrane $s_{af}(t)$ at the hair cell location is transferred into a permeability $s_{hwr}(t)$ of the permeable membrane. The permeable membrane is situated at the base of the hair cell, facing the synaptic cleft. In the model, the conversion is described by increasing $s_{af}(t)$ with an offset value, scaling, and half-wave rectification. Equation (4.1) gives the mathematical representation of this process:

$$s_{hwr}(t) = \begin{cases} 
\beta \cdot (s_{af}(t) + A_2) / (s_{af}(t) + A_2 + B_2), & \text{for } s_{af}(t) + A_2 > 0 \\
0, & \text{for } s_{af}(t) + A_2 \leq 0 
\end{cases},$$  \hspace{1cm} (4.1)

where $A_2$ and $B_2$ are positive constants, $B_2 \gg A_2$, and $\beta$ is a constant needed to scale the transduction process.

Figure 4.3 shows the block diagram of the rest of the transduction process. The side to the left of the dashed line shows the three reservoirs within the hair cell, and the synaptic cleft is present at the right hand side of the dashed line. The dashed line represents the permeable membrane. The output of the hair cell model is the amount of neurotransmitter in the cleft and corresponds to a probability of generating an action potential in the auditory nerve ending.

The amount of neurotransmitter within the reservoirs and cleft over time is described by means of first-order differential equations, giving a positive value to the vectors pointing inwards the block and a negative value to the vectors leaving the block.
The amount of neurotransmitter leaving the free transmitter pool is given by the product of the value of the reservoir times the permeability of the permeable membrane, $s_{\text{hwr}}(t) \cdot q_{\text{ftp}}(t)$. Furthermore, the factory provides new neurotransmitter to the transmitter pool at rate $\epsilon_1 (1 - q_{\text{ftp}}(t))$ and the reprocessing store also provides neurotransmitter at rate $\epsilon_2 q_{\text{rs}}(t)$.

The differential equation for the free transmitter pool is given by

$$\frac{dq_{\text{ftp}}}{dt} = \epsilon_1 (1 - q_{\text{ftp}}(t)) + \epsilon_2 q_{\text{rs}}(t) - s_{\text{hwr}}(t) q_{\text{ftp}}(t).$$

(4.2)

For the cleft the amount of neurotransmitter can be described by a portion that is lost through diffusion $\epsilon_4 q_c(t)$, a portion that is taken back into the hair cell $\epsilon_3 q_c(t)$ and a portion that is released through the permeable membrane, $s_{\text{hwr}}(t) \cdot q_{\text{ftp}}(t)$:

$$\frac{dq_c}{dt} = s_{\text{hwr}}(t) q_{\text{ftp}}(t) - \epsilon_4 q_c(t) - \epsilon_3 q_c(t).$$

(4.3)

The amount of neurotransmitter in the reprocessing store $q_{\text{rs}}(t)$ consists of the part taken back from the cleft into the hair cell $\epsilon_3 q_c(t)$ minus the amount delivered to the free transmitter pool $\epsilon_2 q_{\text{rs}}(t)$:

$$\frac{dq_{\text{rs}}}{dt} = \epsilon_3 q_c(t) - \epsilon_2 q_{\text{rs}}(t).$$

(4.4)

The output of the inner hair cell model is the probability of firing or the discharge rate in spikes/s. It is the amount of neurotransmitter in the cleft $q_c(t)$ scaled by a factor $\epsilon_5$ to make it correspond to physiological data:

$$s_{\text{adapt}} = \epsilon_5 q_c(t).$$

(4.5)

A description of a revised version of this model can be found in Sumner et al. [42], [43]. Improvements in the revised version are the simulation of a larger range of phenomena, such as the responses of also medium spontaneous rate (MSR) and low spontaneous rate (LSR) fibre types, whereas the Meddis model only accounts for high spontaneous rate (HSR) fibres. The revised version also claims to have a closer agreement with recent findings in hair cell physiology.
Table 4.1: Meddis hair cell model parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_2$</td>
<td>5</td>
</tr>
<tr>
<td>$B_2$</td>
<td>300</td>
</tr>
<tr>
<td>$\beta$</td>
<td>2000</td>
</tr>
<tr>
<td>$\epsilon_1$</td>
<td>5.05</td>
</tr>
<tr>
<td>$\epsilon_2$</td>
<td>66.3</td>
</tr>
<tr>
<td>$\epsilon_3$</td>
<td>6580</td>
</tr>
<tr>
<td>$\epsilon_4$</td>
<td>2500</td>
</tr>
<tr>
<td>$\epsilon_5$</td>
<td>50000</td>
</tr>
</tbody>
</table>

4.1.1 Implementation

The input for the hair cell model is the output of an auditory filter. The differential equations in the model are solved numerically by using a timestep $dt$, which is equal to $1/f_s$, in which $f_s$ represents the sample frequency of the total system. The actual implementation was taken from the Auditory Toolbox by Slaney (see table 2.2), and was based on the earlier papers by Meddis. Because the response is level-dependent, the input signal needs to be scaled to the convention defined by Meddis. A level of 30 dB SPL corresponds to a root-mean-square value of 1 for the digital signal.

The dynamic behaviour and its dependence on signal level are tested by applying 500-ms pure tone bursts of 1 kHz, followed by 250-ms silence at the input. The level of the bursts is increased from 0 to 80 dB in steps of 10 dB. The signal is not processed by the peripheral filter nor by an auditory filter to observe the effect of the hair cell model only. Furthermore, no window is applied to the bursts. The output of the hair cell model is shown in figure 4.4. The input levels of the bursts, the onset peaks and the adapted values are given in table 4.2. The fast response to the signal onset is visible, as it takes about 3 ms to decrease by a factor $e^{-1}$ from the peak level. Furthermore, the dynamic range compression is clearly visible: where the input level increases with a factor $10^4$, the adapted level only increases with a factor of 3.5. Also, the ratio between onset peak and adapted level ranges from a factor 1 to 7.6.

In the further processing of the signals of the auditory nerves, the spontaneous rate is considered as a reference level. In our model this is achieved by shifting the output of the hair cell model down by its spontaneous rate, so that the spontaneous rate takes on the value of 0. Because the further processing cannot cope with negative values, the resulting signal is half-wave rectified. This process appears to play an important role in explaining the significant difference in roughness between the ramped and damped signal. The elimination of the spontaneous rate is also applied to the Van Immerseel & Martens hair cell model as described in the next section.
Table 4.2: Adaptation by the Meddis hair cell model for input signals at several levels.

<table>
<thead>
<tr>
<th>Input level (dB SPL)</th>
<th>Onset peak (spikes/s)</th>
<th>Adapted (spikes/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>65.2</td>
<td>65.15</td>
</tr>
<tr>
<td>10</td>
<td>66.3</td>
<td>65.96</td>
</tr>
<tr>
<td>20</td>
<td>69.6</td>
<td>68.47</td>
</tr>
<tr>
<td>30</td>
<td>80.1</td>
<td>76</td>
</tr>
<tr>
<td>40</td>
<td>112.6</td>
<td>96.4</td>
</tr>
<tr>
<td>50</td>
<td>210</td>
<td>138.5</td>
</tr>
<tr>
<td>60</td>
<td>472.4</td>
<td>191</td>
</tr>
<tr>
<td>70</td>
<td>1017.7</td>
<td>223</td>
</tr>
<tr>
<td>80</td>
<td>1715.4</td>
<td>225.5</td>
</tr>
</tbody>
</table>

4.2 Van Immerseel & Martens hair cell model

In the auditory model of Van Immerseel & Martens [16], a slightly simpler hair cell model is used. Their model was intended to be used for the extraction of pitch for speech analysis. The model is based on the model by Meddis and contains the important properties of the hair cell. The difference is that the physiological structure has been removed but the response is quite similar.

First of all, the input signal $s_{af}(t)$ is increased with an offset value $s_{off}$ and then half-wave rectified to obtain signal $s_{hwr}(t)$:

$$s_{hwr}(t) = \max \{0, s_{af}(t) + s_{off}\}.$$  \hspace{1cm} (4.6)

The output $s_{adapt}(t)$ of the hair cell model is derived by applying an automatic gain controlled
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_{\text{spont}} )</td>
<td>0.05</td>
<td>spikes/ms</td>
</tr>
<tr>
<td>( \rho_{\text{sat}} )</td>
<td>0.15</td>
<td>spikes/ms</td>
</tr>
<tr>
<td>( s_{\text{off}} )</td>
<td>0.4472</td>
<td></td>
</tr>
<tr>
<td>( d )</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>( \tau_a )</td>
<td>8</td>
<td>ms</td>
</tr>
<tr>
<td>( \tau_b )</td>
<td>40</td>
<td>ms</td>
</tr>
</tbody>
</table>

Table 4.3: Van Immerseel & Martens hair cell model parameters.

amplifier and dynamic range compression:

\[
\psi(t) = \frac{\rho_{\text{sat}} \cdot \phi(t)}{\left(\sqrt{\frac{\rho_{\text{sat}}}{\rho_{\text{spont}}} - 1}\right) \sqrt{s_{\text{off}} + \sqrt{s_{\text{LP}}(t)}}},
\]

(4.7)

where \( \rho_{\text{sat}} \) is the saturation firing rate, \( \rho_{\text{spont}} \) is the spontaneous firing rate, and \( s_{\text{LP}}(t) \) is the lowpass filtered version of signal \( s_{\text{hwr}}(t) \):

\[
s_{\text{LP}}(t) = s_{\text{hwr}}(t) \ast h_{\text{LP}}(t),
\]

(4.8)

* denoting a convolution product.

The impulse response of the lowpass filter can be written as

\[
h_{\text{LP}}(t) = u(t) \left( \frac{d}{\tau_a} e^{-\frac{t}{\tau_a}} + \frac{1 - d}{\tau_b} e^{-\frac{t}{\tau_b}} \right),
\]

(4.9)

where \( u(t) \) is the unit step function. The values of the parameters are given in Table 4.3.

The signal \( \psi(t) \) is filtered with a third-order lowpass filter with a cut-off frequency of 600 Hz to simulate the loss of phase locking in the auditory nerve. Although phase locking can extend beyond 600 Hz, this value was chosen because the hair cell model was incorporated in a pitch extraction model. For the roughness model, envelope frequencies above 600 Hz are also irrelevant, so this value was kept.

### 4.2.1 Implementation

To arrive at an implementable version of this model, equation (4.7) is discretised with a timestep \( dt \) of \( 1/f_s \). What is also needed is a discrete time version of the lowpass filter. The Laplace transform of the impulse response in equation (4.9) is given by

\[
H_{\text{LP}}(s) = \frac{1 + s(\tau_a - d\tau_a + d\tau_b)}{(1 + s\tau_a)(1 + s\tau_b)},
\]

(4.10)

The digital implementation of this filter can be obtained by applying the bilinear transform

\[
s \rightarrow \frac{2}{T_s} \frac{z - 1}{z + 1}
\]

(4.11)
Table 4.4: Adaptation by the Van Immerseel & Martens hair cell model for input signals at several levels.

<table>
<thead>
<tr>
<th>Input level (dB SPL)</th>
<th>Onset peak (spikes/s)</th>
<th>Adapted (spikes/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>52</td>
<td>51</td>
</tr>
<tr>
<td>10</td>
<td>56.5</td>
<td>53.5</td>
</tr>
<tr>
<td>20</td>
<td>70.5</td>
<td>60.5</td>
</tr>
<tr>
<td>30</td>
<td>115</td>
<td>76.7</td>
</tr>
<tr>
<td>40</td>
<td>250.5</td>
<td>102.9</td>
</tr>
<tr>
<td>50</td>
<td>602.5</td>
<td>131.1</td>
</tr>
<tr>
<td>60</td>
<td>1336.5</td>
<td>154.6</td>
</tr>
<tr>
<td>70</td>
<td>2521.5</td>
<td>171.3</td>
</tr>
<tr>
<td>80</td>
<td>4079.5</td>
<td>182</td>
</tr>
</tbody>
</table>

The level convention for this hair cell model is fixed by the value of $s_0$. This value is defined as $0.001 \cdot \text{max (signal)}$, where 80 dB SPL is the maximum amplitude that is expected in their model ([17], p. 62). Adopting the convention of the Meddis hair cell model, $s_0$ takes on the value of 0.4472. The Van Immerseel & Martens hair cell model was subjected to the same short-term adaptation test as the Meddis model. The result can be found in figure 4.5 and table 4.4. As the input level increases with a factor $10^4$, the adapted output increases with a factor 3.57. The ratio between the onset peak and adapted level ranges from 1 to 22.4.

Figure 4.5: Adaptation by the Van Immerseel & Martens hair cell model.
4.3 Dau et al. adaptation model

The adaptation model by Dau et al. is described in [6]. It was constructed by an approach to the modeling of adaptation different from the Meddis model. Instead of basing the model on physiology, a black box approach was taken. Adaptation in the auditory system was observed from physiological measurements and could be regarded as having several time constants. Before entering the adaptation model, the signal $s_{\text{filt}}$ arriving from an auditory filter is half-wave rectified and filtered by a 5th-order lowpass filter with cut-off frequency of 770 Hz. The resulting signal $s_{\text{hwr}}$ is passed on to the adaptation model.

The model consists of a cascade of $n$ so-called adaptation loops as drawn in figure 4.6. For each loop, the input signal is divided by a lowpass-filtered version of the input signal. The lowpass filter has a time constant $\tau_v$, where $v$ ranges from 1 to $n$, and this time constant differs per loop. The lowpass filtered input signal will initially be small, but increases when a signal input is applied. This is the automatic gain-control function of the loop. When cascading five of these loops ($n = 5$), with time constants ranging from 5 ms to 500 ms, the adaptation model allows to predict perceptual data on temporal processing.

4.3.1 Implementation

The code of the implementation of the loops was taken from Breebaart [3]. The time constants in table 4.5 were derived from that implementation and differ from the ones mentioned in [6]. The level convention of the Dau et al. model relates a root-mean-square value of the digital signal of 1 to a sound pressure level of 0 dB. The output of the model is given in model units and not in spikes/s, as the model is not related to physiology. Subjecting the Dau et al. adaptation model to the short-term adaptation test delivers figure 4.7. When the input increases with a factor $10^4$, the adapted output increases with a factor 394. The ratio between onset peak and adapted level ranges from 3.8 to 115.

As shown, the Dau et al. adaptation model has no spontaneous rate. In the same fashion as for the hair cell models, the output of the adaptation model is half-wave rectified to prevent negative values from arriving at the roughness extraction module.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_1$</td>
<td>5</td>
<td>ms</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>128.75</td>
<td>ms</td>
</tr>
<tr>
<td>$\tau_3$</td>
<td>252.5</td>
<td>ms</td>
</tr>
<tr>
<td>$\tau_4$</td>
<td>376.25</td>
<td>ms</td>
</tr>
<tr>
<td>$\tau_5$</td>
<td>500</td>
<td>ms</td>
</tr>
<tr>
<td>$n$</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: Dau et al. adaptation model parameters.

<table>
<thead>
<tr>
<th>Input level (dB SPL)</th>
<th>Onset peak (model units)</th>
<th>Adapted (model units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>0.52</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>1.54</td>
</tr>
<tr>
<td>20</td>
<td>541.5</td>
<td>74</td>
</tr>
<tr>
<td>30</td>
<td>1419.7</td>
<td>110</td>
</tr>
<tr>
<td>40</td>
<td>3175.3</td>
<td>135</td>
</tr>
<tr>
<td>50</td>
<td>6125.7</td>
<td>150</td>
</tr>
<tr>
<td>60</td>
<td>10503</td>
<td>170</td>
</tr>
<tr>
<td>70</td>
<td>16394</td>
<td>185</td>
</tr>
<tr>
<td>80</td>
<td>23577</td>
<td>205</td>
</tr>
</tbody>
</table>

Table 4.6: Adaptation by the Dau et al. adaptation model for input signals at several levels.

### 4.4 Adaptation and temporally asymmetric envelopes

Figure 4.8 summarises the adapted and peak values of the three adaptation models. The dynamic range compression of the Meddis hair cell model is the largest. The peak response of the Dau et al. model is many times larger than the peak response of the other two models. Furthermore, the Dau et al. model has no spontaneous rate.

Because the reaction to sharp onsets is much stronger than for offsets, the adaptation models are expected to have a different effect on damped than on ramped signals. In this section, the effect of the different implementations of the hair cell models on modulation depth is investigated. The approach is the same as for the auditory filter models. However, measuring asymmetry for the described adaptation models is more difficult because the adaptation is immediate for strong onsets as well as smeared over several signal periods. Figure 4.9 shows this phenomenon.

To get an idea of the effect of the adaptation models on signal asymmetry, modulation depths were estimated at the beginning and the middle of the signal at the output of the adaptation models. The modulation depths were estimated as depicted in equation (3.25). To obtain a good estimate of the modulation depth at the output, the output waveform was windowed over 1 cycle of the modulation frequency, from minimum to minimum. The minimum at the beginning of the 1-s stimulus was located between 21 ms and 50 ms and in the middle of the
Figure 4.7: Adaptation by the Dau et al. hair cell model. Note the absence of spontaneous activity.

Figure 4.8: Adapted (A) and peak (P) values for the three adaptation models as a function of level. Subscript ‘m’ corresponds to the Meddis hair cell model, subscript ‘i’ corresponds to the Van Immerseel & Martens hair cell model and subscript ‘d’ corresponds to the Dau et al. adaptation model. The latter model has no spontaneous rate and has the largest peak response. The output of the Dau et al. is not given in spikes/s but in model units.
stimulus between 486 ms and 515 ms. The applied ramped and damped signals did not pass the peripheral and auditory filtering modules to solemnly observe the effect of the adaptation models. Also, the spontaneous rate was subtracted as mentioned in the description of the implementations of the three models.

Figure 4.10 shows the modulation depth differences for the ramped and damped signals for the three adaptation models at the beginning, and figure 4.11 in the middle of the 1-s stimulus. The figure are presented in a similar manner as the perceptual data to be shown in figure 6.1, thus for three signal centre frequencies and three signal modulation depths.

Overall, the Dau et al. model seems to have the largest effect on asymmetry and the Meddis model seems to have the least effect on asymmetry. The effect is largest for the signals with a centre frequency of 10 kHz. This is probably due to the fact that these signals contain the largest number of spectral components, resulting in a sharper onset and corresponding adaptation module response for the damped signal.
Figure 4.10: Effect of adaptation models on asymmetry at the beginning of the 1-s stimulus, $f_{\text{mod}} = 70$ Hz, $L = 60$ dB. Modulation depth difference between ramped and damped output signals are shown as a function of the modulation depth of the input signal. Legend: modulation depth difference for the Meddis model $\Delta m_m$, the Van Immerseel & Martens model $\Delta m_i$, the Dau et al. model $\Delta m_d$.

Figure 4.11: Effect of adaptation models on asymmetry half way through the 1-s stimulus, $f_{\text{mod}} = 70$ Hz, $L = 60$ dB. Modulation depth difference between ramped and damped output signals are shown as a function of the modulation depth of the input signal. Legend: modulation depth difference for the Meddis model $\Delta m_m$, the Van Immerseel & Martens model $\Delta m_i$, the Dau et al. model $\Delta m_d$. 
Chapter 5

Roughness extraction module

To some sounds, the human brain appears to ascribe a positive amount of roughness. Grossly speaking, this roughness corresponds to temporal envelope fluctuations in the signal components with modulation frequencies ranging from 15 to 150-300 Hz. In this part of the report we will try to estimate the perceived roughness on the basis of the output of the previous stages of the model.

Two roughness extraction models will be presented in this chapter. The first one estimates the modulation depth per auditory filter channel and sums these values for all auditory filters to get a single roughness value. The second roughness extraction model is based on synchronisation of the auditory-nerve activity and operates in the frequency domain.

5.1 Aures’ roughness extraction model

To obtain a roughness value in asper, the Aures [2] model estimates the modulation depth per auditory channel. Similar to loudness summation, the partial roughnesses $r_i$ of all $N_{ch}$ auditory channels should be summed to achieve the total roughness $R$ of the input signal:

$$R = \sum_{i=1}^{N_{ch}} r_i.$$  \hspace{1cm} (5.1)

Figure 5.1 shows the structure of the model to calculate the partial roughness for a single auditory channel from an input signal $s_{in}(t)$ and, below the dashed line, the summation over all auditory channels. As can be seen from the figure, the processing steps before achieving the signals of the auditory channels are different from the ones described in the previous chapters. The Aures model makes use of the Bark scale and the channels are spaced by 1 Bark. The Bark scale (e.g. [48]) is a psychoacoustic scale similar to the ERB-rate scale, but with somewhat wider auditory filter bandwidths.

The first stages of the model consist of extracting the temporal envelopes of the various auditory channels. Since the low-frequency part ($< 15$ Hz) and the high-frequency part ($> 150$ Hz) do not contribute to roughness, these envelopes will be filtered accordingly in the next stage of the model. First the envelope extraction. The signals of the auditory channels
Figure 5.1: Structure of the Aures roughness extraction model. Thick lines represent time-dependent signals, thin lines represent constants. Between the two dashed lines, the signal of a single auditory channel is processed. Below the lower dashed line, the signals from all the auditory channels are combined.
are calculated in a filterbank as depicted in figure 5.2. The filter shape of the filters in the bank is defined a bit simpler than the shapes discussed in chapter 3. The components of the Fourier transform of the outer- and middle-ear filtered signal are calculated. Each component is amplitude weighted by two slopes according to the Bark scale. The value of the upward slope is $\zeta_1 = 27$ dB/Bark and the value of the downward slope $\zeta_2$ in dB/Bark is dependent on frequency $f$ and level $L$ in dB SPL:

$$\zeta_2(f, L) = -24 - \frac{230}{f} + 0.2 \cdot L. \quad (5.2)$$

For every Bark number, the values of the slopes according to a certain component are calculated. The filtered signals are transformed back into the time domain for each auditory filter, using the original phases of the frequency components. Signal $|e_i(t)|$ is the absolute value of the time-domain signal.

Now these signals are filtered in order to determine for each frequency component its contribution to the specific roughness. The signals are filtered by so-called modulation transfer functions $H_i(f_{\text{mod}})$, that represent the modulation sensitivity for channel $i$. The amplitude
characteristics of two filters are given in figure 5.3. The first, $H_1(f_{\text{mod}})$, represents the modulation transfer filter for the first Bark channel. The second, $H_9(f_{\text{mod}})$ represents the modulation transfer filter for the ninth Bark channel and all channels with a higher Bark number. The filters for Bark channel 2 to 9 are constructed by a linear interpolation between filters $H_1(f_{\text{mod}})$ and $H_9(f_{\text{mod}})$.

$$s_{\text{BP},i}(t) = F^{-1}\{F\{|e_i(t)|\} \cdot H_i(f_{\text{mod}})\}. \quad (5.3)$$

The amplitude characteristics of these filters are given in figure 5.3. The upward slope characterises the transition from fluctuation perception to rhythm perception to roughness perception. The downward slope may characterise the limit of phase locking in the auditory cortex [10].

Aures gives no general form to calculate the modulation depth $m_i^*$. It is dependent on the stimuli presented to the model. The modulation depth is calculated by dividing the amplitude range of the envelope of the bandpass filtered signal $\tilde{s}_{\text{BP},i}$ by the DC value $s_0$ of the signal:

$$m_i^* = \frac{\tilde{s}_{\text{BP},i}}{s_0}. \quad (5.4)$$

To model the influence of centre frequency on roughness, the general modulation depths are weighted according to their channel number $i$ by a function $g_i$. The values of the weighting function with respect to channel number are given in figure 5.4.

The partial roughness of channel $i$ in asper/Bark is now given by

$$r_i = \gamma g_i m_i^{*2}. \quad (5.5)$$

The constant $\gamma$ couples the value of 1 asper to the roughness estimate of an AM sinusoid with parameters $f_c = 1$ kHz, $f_{\text{mod}} = 70$ Hz, $m = 1$, and $L = 60$ dB. The squaring is due to the
assumed relation $R \propto m^2$ as depicted in equation (2.1).

Now it has appeared that wide-band noise signals, though showing considerable modulations in their temporal envelopes of their frequency channels, are not perceived as particularly rough. This is probably due to the lack of correlation between the envelopes of the channels. In SAM or SFM tones the temporal envelopes of the various frequency components fluctuate in synchrony which is supposed to contribute strongly to roughness. Hence, the contribution of the various channels to roughness must be weighted for their mutual correlation. To obtain a correct low roughness value for noise signals that have no coherent modulation, the bandpass filtered signals $s_{BP,i}(t)$ are correlated with neighbouring channels. The cross-correlation coefficient between the envelopes of channel $i - 1$ and $i$ is $\chi_{i-1}$, and the cross-correlation coefficient between channel $i$ and $i+1$ is $\chi_i$. Because of the implementation of the auditory filters in the Aures model, there is no phase delay between the auditory channels. The cross-correlation can be kept simple and is calculated by its standard form:

$$\chi(x, y) = \frac{\sum xy - \frac{1}{N} (\sum x)(\sum y)}{\sqrt{[\sum x^2 - \frac{1}{N}(\sum x)^2][\sum y^2 - \frac{1}{N}(\sum y)^2]}}.$$  \hfill (5.6)

where $\chi(x, y)$ is the cross-correlation between signals $x$ and $y$, and $N$ is the number of samples in signal $x$.

The total roughness $R$ is derived by summing the corrected specific roughnesses of all channels:

$$R = \gamma \sum_{i=1}^{N_{ch}} r_i \frac{\chi_{i-1} + \chi_i}{2}. \quad (5.7)$$
5.1.1 Modifications

Daniel & Weber [5] implemented the Aures model, filled in some missing details and made some modifications. The resolution of the filterbank was set to $\frac{1}{2}$ Bark and the $H_i$-filters were re-calculated. Daniel & Weber defined a general function for the calculation of the modulation depth $m_i^*$ for an arbitrary input signal. The DC value $s_{0,i}$ of the input signal $|e_i(t)|$ is determined by the mean of the absolute value:

$$s_{0,i} = |e_i(t)|. \quad (5.8)$$

The root-mean-square value $\tilde{s}_{BP,i}$ of the bandpass filtered signal is calculated by applying equation (5.9):

$$\tilde{s}_{BP,i} = \sqrt{\left( s_{BP,i}(t) - s_{0,i} \right)^2}. \quad (5.9)$$

The generalised modulation depth $m_i^*$ is now calculated by dividing the root-mean-square value by the DC value. The value of the generalised modulation depth was limited to one to prevent extremely high roughness values due to pulses in the envelope:

$$m_i^* = \begin{cases} \frac{\tilde{s}_{BP,i}}{s_{0,i}} & \text{if } \frac{\tilde{s}_{BP,i}}{s_{0,i}} \leq 1 \\ 1 & \text{if } \frac{\tilde{s}_{BP,i}}{s_{0,i}} > 1 \end{cases}. \quad (5.10)$$

Schrader [38] implemented the Daniel & Weber model in Matlab to allow easy modifications from modern insights to the model. Jourdes [22] again modified Schrader’s implementation and applied the ERB-rate scale and the gammatone filterbank. The model now contains 76 auditory channels that are spaced half an ERB. The modulation transfer filters in the model were re-evaluated and converted to the discrete-time domain, omitting the transformation to the frequency domain and the problems of windowing. The implementation of the modulation transfer functions does not allow a sample frequency $f_s$ of 44.1 kHz, because one of the IIR filters becomes unstable. A sample frequency of 48 kHz is used for this model. The modulation transfer functions are now given by

$$H_i = \begin{cases} H_1 & \text{for } 1 \leq i \leq 12 \\ H_{13} & \text{for } 13 \leq i \leq 28 \\ H_{29} & \text{for } 29 \leq i \leq 36 \\ H_{37} & \text{for } 37 \leq i \leq 65 \\ H_{66} & \text{for } 66 \leq i \leq 76 \end{cases}. \quad (5.11)$$

Figure 5.5 shows the modulation transfer functions as used by Jourdes.

The estimation of the modulation depth for channels with relatively low excitation, i.e. those channels that are far away from the modulated frequency components appears to be sensitive to large numerical errors. The skirts of those filters will attenuate the modulations, but still low level modulations will be present. These modulations are distorted by the slope of the skirt. Modulation depth estimation will attribute a value to these attenuated modulations and this value can be quite high. This was solved by introducing a calibration factor $\kappa_i$ that
Chapter 5. Roughness extraction module

scales the root-mean-square value of the channel to the maximum root-mean-square value of all channels:

$$\kappa_i = \frac{\sqrt{\left| e_i(t) \right|^2}}{\max \left( \sum_{i=1}^{76} \sqrt{\left| e_i(t) \right|^2} \right)}.$$  

(5.12)

The limitation on the modulation index estimate is removed and the generalised modulation depth is calculated by

$$m_i^* = \frac{s_{BP,i}}{s_{0,i}} \cdot \kappa_i.$$  

(5.13)

Because of the phase delay over channels introduced by the gammatone filters, the simple cross-correlation function from equation (5.6) cannot be used anymore. One of the signals in the cross-correlation is now shifted over 3 ms and for every shifted value the cross-correlation is calculated anew. The maximum value of all these cross-correlations is chosen. The value of 3 ms seems just to be enough to solve the problem of the phase delay and keep the roughness for white noise close to zero asper. Together with new values for the weights $g_i$ as presented in figure 5.6 this leads to the following definition of partial roughness:

$$r_i = \begin{cases} 
(g_i \cdot m_i^* \cdot \chi_i)^2 & \text{for } i = 1, \ldots, 7 \\
(g_i \cdot m_i^* \cdot \chi_{i-2} \cdot \chi_i)^2 & \text{for } i = 8, \ldots, 72 \\
(g_i \cdot m_i^* \cdot \chi_{i-2})^2 & \text{for } i = 73, \ldots, 76
\end{cases}$$  

(5.14)

Because of the new model components introduced by Jourdes, he had to modify the weights of the auditory channels $g_i$. In general, this is the only degree of freedom to optimise the complete roughness model. For each combination of the module implementations, the weights of the auditory channels will be optimised.
Chapter 5. Roughness extraction module

Figure 5.6: Weights of the auditory channels for the calculation of the partial roughnesses $r_i$ as used by Jourdes. Abscissa: channel number $i$ in ERBs, ordinate: filter weight $g_i$.

By summing the partial roughnesses over all 76 channels the total roughness $R$ in asper is obtained:

$$R = \gamma \cdot \sum_{i=1}^{76} r_i,$$

(5.15)

where calibration factor $\gamma$ has the same function but a different value as in equations (5.5) and (5.7). In the rest of the report, the implementation by Jourdes will be referred to as ‘the Aures roughness extraction model’.

The roughness value is calculated over a single period of the envelope of the periodic input signal. If the signal is non-periodic, a window of 200 ms is used. The observation window is applied after the adaptation module has adapted. In the evaluation, an adaptation period of 300 ms is used before the roughness is calculated.

5.2 Synchronisation index model (SIM)

As explained in the previous chapter, auditory nerve fibres fire electrical pulses in response to the movement of the hair cells. This firing can phase lock to frequency components present in the auditory filter corresponding to that hair cell. This phase-locking effect disappears at frequencies higher than about 4 kHz. Auditory nerves can thus also phase lock to a modulation frequency. This phase locking is also called synchronisation. The synchronisation index (e.g. [11]) is a measure that can describe the amount of phase locking in a certain auditory nerve. In physiological experiments firing patterns of auditory nerves are recorded and the synchronisation index was indeed developed in this field. The definition used here is obtained by applying the Fourier transform to a firing-pattern record and dividing the value of a certain phase-locked frequency by the DC value. The DC value corresponds to the average firing
rate. The resulting value is the synchronisation index of that frequency.

Leman [24] developed two models to estimate roughness by means of the synchronisation index. The models make use of the fact that roughness is defined as the energy of relevant modulation frequencies in the auditory channels with respect to the total energy. Both models by Leman make use of the ERB-rate scale containing 40 channels spaced 1 ERB. The first model (I) first calculates the synchronisation index for the modulation in a single channel and then sums these over all channels. The second model (II) evaluates the modulations over all channels before calculating the synchronisation index.

5.2.1 Model I

![Diagram of Model I](image)

**Figure 5.7:** Structure of the synchronisation index model I roughness extraction model.

First of all, the signals \( s_{\text{adapt},i}(t) \) after passing the adaptation stage are Fourier transformed into the signals \( S_{\text{adapt},i}(f) \) where \( i \) indicates the auditory channel. A Hamming window \( w_{\text{hamming}}(t) \) with a length of 200 ms is applied:

\[
S_{\text{adapt},i}(f) = F \{ s_{\text{adapt},i}(t) \cdot w_{\text{hamming}}(t) \}.
\]  

(5.16)

Same as with the Aures model, the observation window is applied after the adaptation module has taken on its adapted value. In this model the modulation transfer filters \( H_i(f_{\text{mod}}) \) are also applied, but now the output \( S_{\text{BP},i}(f) \) remains in the frequency domain:

\[
S_{\text{BP},i}(f) = g_i \cdot S_{\text{adapt},i}(f) \cdot H_i(f_{\text{mod}}).
\]  

(5.17)

The used modulation transfer functions differ from the ones in the Aures model. The filters are described by analytical equations and the result for 40 auditory channels is plotted in figure
5.8. Especially in the lower channels the filters are narrow and have a lower peak-frequency. Leman suggested that the analytical description needed some more fine-tuning. The maximum amplitude of the filters again is scaled by the parameter $g_i$. As mentioned in the description of the Aures model, this delivers a degree of freedom to compensate for frequency-dependent amplitude effects of the modules. Because of the bandpass filtering there is no information left above 300 Hz, and the rest of the calculations are performed over a frequency range from 0 to 300 Hz. This is done to increase calculation speed.

The DC component of the signal after the adaptation module $S_{DC,i}$ is simply the value for frequency 0 Hz. Because the frequency resolution of the discrete implementation of the model is limited, the first frequency bin will also contain some very low frequencies. The bandwidth of a single frequency bin $f_{bin}$ in Hz is given by

$$f_{bin} = \frac{f_s}{N_{\text{samples}}} = \frac{f_s}{f_s \cdot t_{\text{win}}} = \frac{1}{t_{\text{win}}}. \quad (5.18)$$

Because the synchronisation index model applies a window of $t_{\text{win}} = 200$ ms, the frequency resolution is 5 Hz. It is not recommended to use a smaller window, as the resolution of the modulation transfer function will decrease.

The DC value for a single channel is given by the next equation:

$$S_{DC,i} = S_{\text{adapt},i}(0). \quad (5.19)$$

The synchronisation index for the modulations in the channel is obtained by taking the absolute value of the quotient of the above calculated bandpass filtered signal and the DC value. This value is taken to the power $\alpha$, following the relation between modulation depth and roughness as depicted in equation (2.1). The value of $\alpha$ is chosen to be 1.6 in the
implementation by Leman. This leads to the following equation:

\[
    r_i(f) = \left| \frac{S_{BP,i}(f)}{S_{DC,i}} \right|^\alpha. \tag{5.20}
\]

The generalised roughness spectrum is summed over all channels:

\[
    r(f) = \sum_{i=1}^{N_{ch}} r_i(f) \tag{5.21}
\]

By integrating over frequency and scaling with a factor \( \gamma \), the roughness value \( R \) in asper is obtained:

\[
    R = \gamma \int r(f) df \tag{5.22}
\]

### 5.2.2 Model II

![Diagram of Model II](image_url)

Figure 5.9: Structure of the synchronisation index model II roughness extraction model.

Figure 5.9 shows the structure of the second roughness extraction model by Leman. This model will also take the phase of the frequency components in the synchronisation index with respect to other auditory channels into account. The conversion to the frequency domain and the bandpass filtering will be the same as equations (5.16) and (5.17) in model I.

The DC values of the spectra are summed over all channels:

\[
    S_{DC} = \sum_{i=1}^{N_{ch}} S_{adapt,i}(0), \tag{5.23}
\]
as well as the bandpass filtered signals:

\[ S_{BP}(f) = \sum_{i=1}^{N_{ch}} S_{BP,i}(f). \]  (5.24)

Since the frequency components of \( S_{BP}(f) \) are constructed as the vector sum of the non-zero frequency components of \( S_{BP,i}(f) \) in the channels, \( S_{BP}(f) \) will have an amplitude and a phase component. The roughness estimate for all frequency components is estimated by

\[ r(f) = \left| \frac{S_{BP}(f)}{S_{DC}} \right|^\alpha. \]  (5.25)

The total roughness \( R \) in asper is determined by integrating equation (5.25) over frequency and by scaling the integral by a factor \( \gamma \):

\[ R = \gamma \int r(f) df. \]  (5.26)

Model II is more promising than model I because it is able to account for coincidental modulations in noise signals by evaluating the phase of the Fourier components across channels. Model II is referred to as ‘the synchronisation index model’ in the rest of the report.

### 5.3 Roughness extraction module and temporally asymmetric envelopes

The Aures model estimates roughness by calculating modulation depths for each auditory channel. Without applying any module to the stimuli, the modulation depth is the same for the ramped and damped stimuli, so no supplementary effect on the difference in roughness between the two stimuli is expected from this model.

The synchronisation index model estimates roughness in the frequency domain. Without applying any module to the stimuli, only the phases of the frequency components are different for the ramped and damped stimuli. The frequency components of the stimuli by Pressnitzer & McAdams are contained within one auditory channel. Since no phase information is processed within a single channel by the synchronisation index model, no supplementary effect on the difference in roughness between the two stimuli is expected from this model.
Chapter 6

Evaluation of the roughness models

Our evaluated module implementations offer twelve possible combinations to form a roughness model (2 auditory filterbanks, 3 adaptation modules, 2 roughness extraction modules). Aside from the three panels in figure 6.1, that describe our reference to the roughness of ramped and damped signals, a large dataset with other stimuli is tested. These include SAM sinusoids, pAM sinusoids, low-frequency pure tones, SAM noise, white noise and SFM sinusoids for which various parameters are varied. Presenting all data in this chapter would take up hundreds of figures and would not lead to a great comprehension. Instead, for each stimulus, the combination that results in the best fit will be presented. The results for other combinations will be discussed in the text, in particular for those combinations which completely fail to model a specific data set.

In the described figures, the scales and axes are kept similar to the figures in the papers that describe the experimental data. All calculations were performed with a sample frequency $f_s$ of 48 kHz. First, the roughness values of reference signals were calculated. If the experiment yielded roughness values on an asper scale, the reference signal is a 1 asper SAM sinusoid which has a centre frequency of 1 kHz, a modulation frequency $f_{\text{mod}}$ of 70 Hz, a modulation depth $m$ of 1, and a level $L$ of 60 dB SPL.

6.1 Ramped and damped stimuli

Figure 6.1 shows the perceptual data of the roughness experiment by Pressnitzer & McAdams. The experiment was performed monaurally with headphones at a level of 60 dB SPL. The ramped and damped stimuli were presented for 1 s each. For each centre frequency, all possible combinations of pairs of the data points (ramped/damped, modulation depth) were presented. The listener had to judge which one of the two sounds was rougher. From these paired-comparison data, a roughness scale was computed based on the Bradley-Terry-Luce (BTL) theory ([7], [8]). The scale is essentially normalised and has zero mean. Hence both positive and negative values result. Because the experiments for the three centre frequencies were carried out separately, the three separate panels cannot be compared with each other with respect to the BTL-value.

The authors conclude the following: “The influence of modulation depth is visible for all centre frequencies; a higher modulation depth introduces more roughness. A significant effect
of the shape of the waveform envelope is also observed. For a given amplitude spectrum and a given modulation depth, the ‘reversed’ stimulus is systematically judged rougher that the ‘sawtooth’ stimulus. This effect is highly significant at all \( f_c \)’s for all modulation depths. It is, however, significantly smaller for stimuli centered on \( f_c = 2.5 \) kHz.” ([37], p. 2778–2779) As a reminder, the ‘sawtooth’ stimulus corresponds to the ramped stimulus in the terminology used in this report, whereas the ‘reversed’ corresponds to the damped stimulus.

Although the panels cannot directly be compared due to the applied normalised scale, each single panel provides relative information about the perceived roughness values. In figure 6.1 for example, it can be seen that, for a centre frequency of 2.5 kHz the ramped signal with a modulation depth of 0.8 is rougher than the damped signal with a modulation depth of 0.4. At 5 kHz, these two signals have about the same roughness.

Three characteristic plots out of the results of the twelve roughness models, as defined in table 6.1, are given in figures 6.2–6.4. The other plots resemble one of these plots and this is also summarised in table 6.1. The asper scales vary for the different plots, but as the BTL scale and the asper scale cannot be quantitatively related, only relative comparisons are legitimate.

Comparing the three plots in figures 6.2–6.4 to figure 6.1, there is not one that really has a perfect fit. Recalling the conclusion of Pressnitzer & McAdams, the influence of modulation depth is correct and the roughness of the ramped stimulus is systematically smaller than the roughness of the damped stimulus. But this difference is relatively much larger for the Dau et al. model, which overestimates the roughness difference, than for the other two hair-cell models which underestimate the roughness difference of ramped and damped signals. Figure 6.3 and the resembling figures as presented in table 6.1 agree best with Pressnitzer and McAdams’ conclusion.

In general, also demonstrated in chapter 4, the adaptation module is the relevant module for the roughness difference between the ramped and damped stimuli. The physiological hair cell models show too little an effect whereas the Dau et al. model shows too large an effect on the difference. Furthermore, the Aures roughness extraction model seems to have a positive
Chapter 6. Evaluation of the roughness models

Figure 6.2: Roughness estimate of the ramped and damped stimuli, $f_{\text{mod}} = 70$ Hz, $L = 60$ dB. Modules: gammatone filterbank, Dau et al. adaptation model and Aures roughness extraction model. Legend: roughness estimate for the damped stimulus $R_d$, roughness estimate for the ramped stimulus $R_r$.

Figure 6.3: Roughness estimate of the ramped and damped stimuli, $f_{\text{mod}} = 70$ Hz, $L = 60$ dB. Modules: gammatone filterbank, Meddis hair cell model and Aures roughness extraction model. Legend: roughness estimate for the damped stimulus $R_d$, roughness estimate for the ramped stimulus $R_r$.

effect on the overall shape of the figures.

6.2 Larger dataset

6.2.1 Sinusoidally amplitude modulated sinusoids

As roughness originates from modulations within auditory filters, the roughness of amplitude modulated stimuli has been studied thoroughly. Recalling equation 2.2:

$$s_{\text{SAM}}(t) = \{1 + m \cdot \sin(2\pi f_{\text{mod}} t + \phi_{\text{mod}})\} \cdot \sin(2\pi f_c t + \phi_c),$$

several parameters of the SAM sinusoid can be varied. Subsequently, the parameters modulation frequency $f_{\text{mod}}$, modulation depth $m$, level $L$ and phase of the carrier relative to the phase of the envelope have been evaluated.
Figure 6.4: Roughness estimate of the ramped and damped stimuli, $f_{\text{mod}} = 70$ Hz, $L = 60$ dB. Modules: gammatone filterbank, Meddis hair cell model and synchronisation index model. Legend: roughness estimate for the damped stimulus $R_d$, roughness estimate for the ramped stimulus $R_r$.

Table 6.1: The twelve combinations of the module implementations. The last column refers to the figure that mostly resembles the roughness estimate for the ramped and damped stimuli for that combination of modules.
Chapter 6. Evaluation of the roughness models

Figure 6.5: Roughness of SAM sinusoids. Calculated (closed circles (●) and solid and dashed lines) and perceptually measured roughness (dotted lines) in asper of SAM sinusoids as a function of modulation frequency $f_{\text{mod}}$ and centre frequency $f_c$, $m = 1$, $L = 60$ dB. Modules: gammachirp filterbank, Van Immerseel & Martens hair cell model, Aures roughness extraction model. Subjective data reproduced from [5].

Modulation frequency

The modulation frequency of the SAM sinusoid was varied for several centre frequencies. The best fit to the subjective data was obtained with a combination of the gammachirp filterbank, the Van Immerseel & Martens hair cell model and the Aures roughness extraction model. The results are given in figure 6.5. This figure was actually used to derive the weighting functions $g_i$ as described in chapter 5.

With the Aures roughness extraction model, an unexpected discrepancy occurs for centre frequencies 125 Hz, shown in the upper left panel of figure 6.5, and 250 Hz, shown in the lower right panel of figure 6.5. At zero modulation the calculated roughness is nonzero and significant. This is due to the fact that the half-wave rectified low-frequency pure tones are treated as envelopes by the hair cell model. The tone frequencies are simply within in the roughness frequency range. The hearing system also appears to have this feature. Recent experiments by Miśkiewicz [28] show a significant roughness percept for low-frequency pure tones. This roughness for unmodulated sinusoids is not contained in the experimental data shown in figure 6.5 This discrepancy does not occur for combinations with the synchronisation index model.

Modulation depth

The combination of the gammachirp filterbank, Meddis hair cell model and Aures roughness extraction model appears to have the best fit for the relation between roughness and modu-
Figure 6.6: Calculated roughness (closed circles (●) and solid line) in asper as a function of modulation depth \( m \) for an SAM sinusoid, \( f_c = 1 \text{ kHz}, f_{\text{mod}} = 70 \text{ Hz}, L = 70 \text{ dB} \). The dotted line with the open circles (○) is the power relation \( R = 1.36 \times m^{1.6} \), which is a good approximation for subjective data [5].

Modules: gammachirp filterbank, Meddis hair cell model, Aures roughness extraction model.

6.2.2 “Pseudo-AM” (pAM) sinusoids

Considering the SAM sinusoid as a complex of three cosines, the phase of the centre-frequency component \( \phi_c \) can be varied to obtain a so-called “pseudo-AM (pAM)” sinusoid [37]:

\[
s_{\text{pAM}}(t) = \frac{1}{2} \cos (2\pi (f_c - f_{\text{mod}}) t) + \cos (2\pi f_c t + \phi_c) + \frac{1}{2} \cos (2\pi (f_c + f_{\text{mod}}) t).
\]

Varying the phase of the centre frequency component affects the modulation depth of the total signal, whereas the amplitude characteristic remains the same. From figure 6.9 it is clear
Figure 6.7: Calculated roughness in asper as a function of modulation depth \( m \) for two SAM sinusoids, \( f_{\text{mod}} = 50 \, \text{Hz}, L = 70 \, \text{dB} \). The solid line with the closed circles (●) corresponds to an SAM sinusoid with a centre frequency \( f_c \) of 300 Hz and the dashed line with the asterisks (*) corresponds to an SAM sinusoid with a centre frequency \( f_c \) of 1 kHz. The dotted line with the open circles (○) is the function \( 1.12 \cdot m^{1.8} \) which corresponds well to the subjective data at 1 kHz [5]. Modules: gammachirp filterbank, Meddis hair cell model, Aures roughness extraction model.

Figure 6.8: Calculated (solid line, closed circles (●)) and subjective (open circles (○)) relative roughness of an SAM sinusoid as a function of level \( L \), \( f_c = 1 \, \text{kHz} \), \( f_{\text{mod}} = 70 \, \text{Hz} \), \( m = 1 \). Modules: gammatone filterbank, Van Immerseel & Martens hair cell model, Aures roughness extraction model. Subjective data reproduced from [48].
that if the centre frequency cosine-component is in phase with the two modulation cosine-components the roughness is largest. As for the ramped and damped stimuli, we cannot make any comments on the relation between separate panels as the results were obtained by paired-comparison experiments and scaled by the BTL theory.

The best qualitative fit is obtained with the combination of the gammatone filterbank, the Van Immerseel & Martens hair cell model and the synchronisation index model, as shown in figure 6.10.

### 6.2.3 Pure tones

Very recently, Miśkiewicz [28] demonstrated that low-frequency pure tones can also introduce a significant percept of roughness. This also occurs in our model, as the half-wave rectified low-frequency pure tones are treated as envelopes. The combination of the gammachirp filterbank, the Meddis hair cell model and the synchronisation index model is able to give a good qualitative agreement with the subjective data. This is shown in figure 6.11.

The Aures model seems not to be able to correctly reproduce the roughness of pure tones. This is, among other things, due to the modulation transfer function for the low-frequency channels. The modulation transfer function for the low-frequency channels of the synchronisation index model has a maximum at about 20 Hz (figure 5.8), where the value of the dataset by Miśkiewicz indeed shows the highest value. In the modulation transfer functions of
Figure 6.10: Calculated roughness of pAM sinusoids in asper as a function of phase $\phi_c$ of the centre frequency $f_c$ and centre frequency, $m = 1$, $L = 60$ dB. Modules: gammatone filterbank, Van Immerseel & Martens hair cell model, synchronisation index model.

Figure 6.11: Calculated (closed circles (●) and solid line) and subjective (geometric mean (○) and dotted line) roughness in asper of low-frequency pure tones as a function of frequency in Hz. Modules: gammachirp filterbank, Meddis hair cell model, synchronisation index model. Subjective data reproduced from [28].
Figure 6.12: Calculated (medians (●) and quartiles of 20 realisations) relative roughness of SAM noise as a function of modulation frequency $f_{\text{mod}}$, $m = 0.98$, $L = 60$ dB. The reference roughness is the roughness at a modulation frequency of 70 Hz. The subjective roughness (dotted lines) was measured relative to modulated noise with a modulation frequency of 70 Hz (medians (○) and quartiles) and relative to modulated noise with a modulation frequency of 300 Hz (medians (×) and quartiles) [9]. Modules: gammachirp filterbank, Van Immerseel & Martens hair cell model, Aures roughness extraction model.

the Aures model (figure 5.5), the maximum is at about 35 Hz.

6.2.4 Sinusoidally amplitude modulated noise

Fastl [9] evaluated the roughness of sinusoidally amplitude-modulated noise. The experiments were performed on a relative scale. Two reference sounds were used in two separate experiments. The first experiment had a reference sound that was rough and this sound was given a value of 100. The stimuli had to be scaled according to the reference sound. In the other experiment, a sound with faint roughness was applied as a reference. This sound was given a value of 10 and again the stimuli had to be scaled accordingly. This concept yielded two curves per parameter and in all cases the judgements relative to the rough sound differed from the judgements relative to the sound with faint roughness. Both results are included in the upcoming figures.

Modulation frequency

Similar to the SAM sinusoid, the dependency of roughness on modulation frequency for the SAM noise stimulus has a peak at 70 Hz. All combinations result in a reasonable fit except for a combination of the Dau et al. adaptation model and the synchronisation index model. Figure 6.12 shows the nice fit for the combination of the gammachirp filterbank, the Van Immerseel & Martens hair cell model and the Aures roughness extraction model.
Modulation depth

The roughness of SAM noise increases with increasing modulation depth, at least up to a modulation depth of 1. Plotted in figure 6.13 is the calculated roughness for a combination of the gammachirp filterbank, the Meddis hair cell model, Aures roughness extraction model. Combinations with the synchronisation index model generally have worse results than combinations with the Aures roughness extraction model.

Level

An increase in level generally relates to an increase in roughness, and this is also the case for SAM noise. However, none of the model combinations achieves a good fit to the data as presented in figure 6.14. All combinations underestimate the dependence on level. The best fit is obtained with the gammatone filterbank, the Meddis hair cell model and the Aures roughness extraction model. Combinations with the synchronisation index model again have the worst fit to the level-dependent data.

6.2.5 Unmodulated bandpass-filtered white noise

The roughness of unmodulated white noise is very low and approximates the threshold of roughness (0.07 asper). A dataset of bandpass filtered white noise was simulated. Because white noise has no periodicity, in the case of the Aures roughness extraction model, no evaluation window could be chosen with respect to the signal. In this case, the same 200-ms window as used by the synchronisation index model was used to obtain a roughness estimate.
Figure 6.14: Calculated (medians (●) and quartiles of 20 realisations) relative roughness of SAM noise as a function of level $L$, $f_{\text{mod}} = 80$ Hz, $m = 0.85$. The reference roughness is the roughness with a level of 80 dB. The subjective roughness (dotted lines) was measured relative to modulated noise with a level of 40 dB (medians (○) and quartiles) and relative to modulated noise with a level of 60 dB (medians (×) and quartiles) [9]. Modules: gammatone filterbank, Meddis hair cell model, Aures roughness extraction model.

White noise was bandpass filtered around a certain centre frequency and the bandwidth of the filter was varied. For the simulations twenty noise representations were applied per datapoint and the medians and quartiles are plotted. The combination of the gammatone filterbank, the Meddis hair cell model and the synchronisation index model results in the best fit. Generally the Aures roughness extraction model achieves too high roughness values for the lower three panels. The synchronisation index model performs best in the case of white noise signals.

6.2.6 Sinusoidally frequency-modulated sinusoids

The understanding of roughness of sinusoidally frequency-modulated (SFM) sinusoids is not as straightforward as the roughness of SAM sinusoids. Depending on the modulation index $m = \Delta f / f_{\text{mod}}$ the carrier and modulation components expand over a large frequency range and extend over several auditory filters. Three parameters of the SFM sinusoid have been varied respectively: modulation frequency $f_{\text{mod}}$, frequency deviation $\Delta f$ and level $L$.

Modulation frequency

In figure 6.16 the modulation frequency $f_{\text{mod}}$ is increased from 1 Hz to about 500 Hz. Because the frequency deviation $\Delta f$ is 800 Hz the modulation index $m$ remains smaller than 1. The spectrum is similar to the spectrum of the SAM sinusoid, and therefore it is not a surprise that there is a maximum at approximately 70 Hz. The shape of the figure is quite similar to that of the SAM sinusoid (figure 6.5), taking into account that the scale of the frequency axis is logarithmic in this figure.
The modules used in the figure are the gammachirp filterbank, the Van Immerseel & Martens hair cell model and the Aures roughness extraction model. Although the fit for the SAM sinusoids at a centre frequency of 2 kHz is good for all module combinations, the frequency range 8–70 Hz is not fitted well for all module combinations. The synchronisation index model consistently makes the rising flank shift much too high frequencies.

**Frequency deviation**

When varying the frequency deviation $\Delta f$, only the modulation index is varied. Beyond $\Delta f = 70$ Hz, the spectrum of the SFM sinusoid becomes broader and the carrier component changes in amplitude.

In general, all module combinations are able to generate a reasonable fit. The best fit is obtained by a combination of the gammachirp filterbank, the Dau et al. adaptation model and the synchronisation index model. Figure 6.17 shows the results for that combination.

**Level**

The last parameter to be varied for the SFM sinusoid is the level $L$. An increase in level results in an increase in roughness. A very good fit to the subjective data is demonstrated by the combination of the gammatone filterbank, the Van Immerseel & Martens hair cell model and the Aures roughness extraction model in figure 6.18. Combinations with the synchronisation index model generally result in a bad fit. The bad level-dependency of SIM also plays
Figure 6.16: Relative roughness of an SFM sinusoid as a function of modulation frequency, $f_c = 1.6 \text{ kHz}$, $\Delta f = 800 \text{ Hz}$, $L = 60 \text{ dB}$. The solid line with the closed circles (●) represents the calculated roughness and the dotted line represents the subjective roughness (medians (○) and quartiles) [5]. The reference is the roughness of the SFM sinusoid modulated with 70 Hz. Modules: gammachirp filterbank, Van Immerseel & Martens hair cell model, Aures roughness extraction model.

Figure 6.17: Relative roughness of an SFM sinusoid as a function of frequency deviation, $f_c = 1.6 \text{ kHz}$, $f_{\text{mod}} = 70 \text{ Hz}$, $L = 60 \text{ dB}$. The dashed line with the closed circles (●) represents the calculated roughness and the dotted line represents the subjective roughness (medians (○) and quartiles) [5]. The reference is the roughness of the SFM sinusoid with a frequency deviation of 800 Hz. Modules: gammachirp filterbank, Dau et al. adaptation model, synchronisation index model.
a role here.

6.3 Analysis

To summarise the large amount of data obtained by the simulations of the twelve models, two tables have been constructed. This analysis is performed informally; The ratings are derived subjectively by the author. In table 6.2, the twelve combinations of the modules are evaluated with respect to the separate stimuli. The numbers correspond to the numbers in table 6.1, in which the combinations are defined. The symbols indicate to which extend (‘{‘ = very bad, ‘‘ = bad, ‘’ = reasonable, ‘+’ = good, ‘++’ = very good) the simulated values correspond to the subjective data.

Two models appear to have the best overall results. They both contain the gammachirp filterbank and Aures roughness extraction model. The adaptation model can either be the Meddis or the Van Immerseel & Martens hair cell model. Their relation is again underlined here, although the ratings of the two combinations are not identical.

Table 6.3 describes, again informally, how each implementation of a module contributes to the fit to each stimulus with respect to the other implementation(s) within the module. If the implementations within the model repond the same to a certain stimulus, all implementations are rated with a circle ‘◦’. If one of the implementations produces a better result for a certain stimulus than the other implementation, the better implementation receives a plus sign ‘+’ and the other implementation a circle. Five symbols are possible: ‘{‘ = much worse, ‘‘} =
### Table 6.2: Schematic analysis of the twelve combinations of the modules with respect to the large dataset of stimuli. The twelve combinations are given in Table 6.1

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Parameters</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramped &amp; damped</td>
<td>$m, f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>SAM sinusoids</td>
<td>$f_{mod}, f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$m, f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$L$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>pAM sinusoids</td>
<td>$\phi_e, f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Pure tones</td>
<td>$f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>SAM noise</td>
<td>$f_{mod}$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$m$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$L$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>White noise</td>
<td>Bandwidth, $f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>FM sinusoids</td>
<td>$f_{mod}$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$\Delta f$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$L$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>6</td>
<td>-8</td>
<td>-3</td>
<td>-5</td>
<td>-13</td>
<td>-15</td>
<td>+1</td>
<td>-8</td>
<td>+1</td>
<td>-5</td>
<td>-11</td>
<td>7</td>
</tr>
</tbody>
</table>

### Table 6.3: Schematic analysis of the relative effect of the implementations of the modules with respect to the large dataset of stimuli.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Parameters</th>
<th>(\Gamma)-tone</th>
<th>(\Gamma)-chirp</th>
<th>Meddis</th>
<th>VI&amp;M</th>
<th>Dau</th>
<th>Aures</th>
<th>SIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramped &amp; damped</td>
<td>$m, f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>SAM sinusoids</td>
<td>$f_{mod}, f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$m, f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$L$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>pAM sinusoids</td>
<td>$\phi_e, f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Pure tones</td>
<td>$f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>SAM noise</td>
<td>$f_{mod}$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$m$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$L$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>White noise</td>
<td>Bandwidth, $f_c$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>FM sinusoids</td>
<td>$f_{mod}$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$\Delta f$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td></td>
<td>$L$</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>+2</td>
<td>+3</td>
<td>+1</td>
<td>+2</td>
<td>-3</td>
<td>+2</td>
<td>-2</td>
</tr>
</tbody>
</table>


The gammachirp filterbank achieves a slightly better result than the gammatone filterbank, and the Van Immerseel & Martens hair cell model slightly overclasses the Meddis model. The Dau et al. model generally does not achieve a good fit with the subjective data. With respect to the roughness extraction module implementations, the synchronisation index model scores bad with all level-dependent stimuli. Although the model has a good agreement with phase and white noise data, the Aures roughness extraction model is preferred generally. As a final application of the roughness models, the sounds of rolling balls are evaluated.
6.4 Rolling sounds

The roughness of rolling sounds was estimated by applying recordings of rolling wooden balls to our roughness model. Balls of several diameters were recorded at several speeds. The balls rolled over an MDF plate with a thickness of 8 mm. The plate was mounted on springs so that it could radiate without being intensively damped.

The sounds were analysed by a model consisting of the gammachirp filterbank, the Van Immerseel & Martens hair cell model and the Aures roughness extraction model. As seen in section 6.3, this combination of module implementations is recommended as it corresponds best to the large set of subjective data. The level of the sounds was scaled to 60 dB SPL. The velocity of the ball was not perfectly constant as friction was not excluded in the test. Hence, the average speed of the ball is taken in this analysis. As no periodicity within the envelope of the rolling sounds can be defined, successively half-overlapping 200-ms windows were used to determine the average roughness of the rolling sound.

The model predicts that the average roughness of the rolling sounds increases with speed and decreases with diameter, as shown in figure 6.19. We can thus conclude that auditory roughness may well be a temporal cue for the perception of size and speed of rolling objects. This suggestion was discarded in the earlier research by Houben [15]. The plotted variance is the bias-corrected sample variance $\hat{\sigma}^2$:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2.$$  \hfill (6.2)

The variance shows that it is difficult to calculate a roughness value for the two smaller balls. These balls start bouncing on the plate and this bouncing is mostly non-periodic. The periodicity of the envelope within a single observation window contributes to the roughness within the observed window and thus the roughness varies per observed window. The variance of the roughness for the large ball is small, because this ball sticks to the plate and does not bounce much.
Figure 6.19: Calculated roughness in asper (average (●, × and ○) and variance) of recorded rolling sounds as a function of speed in m/s for three balls with different diameters in mm. Modules: gammachirp filterbank, Van Immerseel & Martens hair cell model, Aures roughness extraction model.
Chapter 7

Conclusion

This report was concerned with the modeling of auditory roughness for signals with temporally asymmetric envelopes. Ramped and damped signals have temporally asymmetric envelopes, as well as many sounds in nature, including rolling sounds. The current roughness models seem not to be able to explain the perceptually found roughness difference between ramped and damped signals [37]. This sensitivity of the hearing system to temporal asymmetry needs to be modeled to be able to obtain correct roughness values for arbitrary sounds. In particular, human beings are to some extent able to discriminate between sizes and speeds of rolling objects. Although auditory roughness would be a good candidate for being a temporal cue in this ability to discriminate, this cue was found to be insignificant in previous research by Houben [15].

Four modules, three of which available in multiple selectable implementations, were combined into twelve roughness models. The models were tested on the difference they can ascribe to ramped signals on the one side and damped signals on the other side. The used modules are summarised in table 2.1. Most of the module implementation were available in downloadable packages, but the level-dependent compressive gammachirp filterbank, the Van Immerseel & Martens hair cell model and the synchronisation index model needed to be implemented from their description in the literature. The new roughness models should also produce good results for a large set of stimuli with temporally symmetric envelopes for which perceptual data is available. This was also evaluated in this report.

Concerning the roughness of signals with temporally asymmetric envelopes, the adaptation modules appear to be dominant. The auditory filterbank modules and the roughness extraction modules are of a secondary importance. The Dau et al. adaptation model gives the largest onset amplification, but for the difference in roughness for the ramped and damped stimuli the effect is too large. The physiological hair cell models by Meddis and by Van Immerseel & Martens, on the other hand, result in too little an effect on asymmetry.

For each set of perceptual data, there is a combination of module implementations that achieves a good fit. Unfortunately, this combination is not the same for all datasets. Each module implementation has its good and bad characteristics. The results of the simulations for the large set of data are summarised in the tables 6.2 and 6.3. Depending on the kind of stimulus applied to the roughness model, a preferable combination can be chosen from these
tables. A ‘safe’ combination appears to be the gammachirp filterbank with either the Meddis or the Van Immerseel & Martens hair cell model and the Aures roughness extraction model.

The roughness of rolling balls was scaled with respect to size and speed. With increasing diameter of the ball, the roughness decreases. With increasing speed of the ball, the roughness increases. Due to the non-periodic bouncing of the ball on the plate, there is a large variance of the roughness values for the balls with small diameters.

According to Miśkiewicz [28] low-frequency pure tones can introduce a significant roughness percept. This was also found in our results chapter. The half-wave rectified pure tones are treated as envelopes and thus attribute to roughness summation. More research could be conducted in this field, in modeling as well as in psychoacoustical experiments. The fact that low-frequency pure tones introduce roughness and thus can be annoying is interesting in the field of comfort engineering. Low-frequency cockpit sounds may for example be a culprit to the comfort of flying.

To be able to cascade the separate modules to form a single roughness model, some conventions needed to be obeyed. The adaptation models are dependent on the amplitude values of the digital input signal. A higher input level means more dynamic range compression and a higher onset response.

The gammachirp filter has level as an input parameter. In our model, this level is the level of the input signal. Each filter in the filterbank receives this level parameter, even though in most cases not all auditory channels are excited. Furthermore, the peripheral filter applies a level change to most of the spectral components within the input signal. An improvement to the current implementation of the gammachirp filterbank could thus be a level estimation algorithm per channel. This implies an algorithm that follows the time-signal of each channel and changes the level parameter accordingly. This will undoubtedly cost many extra calculations. For a follow up to this project we suggest to use a powerful computer and to rewrite some of the time-consuming modules into a compiled code to speed up the simulation process.

The Meddis model is an inner-hair-cell model, describing the response of high-spontaneous rate (HSR) fibres. The inner ear also contains low- and medium-spontaneous rate fibres. In the recent literature ([42], [43]), the effects of these fibres are combined into a revised version of the Meddis model. Although the HSR fibres contribute most to the onset amplification, implementing the revised version of the Meddis model will at least bring our model more close in line with physiological data.

The synchronisation index model (model II) can become promising if the level-dependency is improved. The model consumes significantly less calculation time than the Aures roughness extraction model as it contains less filters within the auditory filterbank and it does not contain a time-consuming cross-correlation calculation.

In conclusion, we can say that the adaptation module determines the difference in roughness between the ramped and damped signals. However, the three evaluated adaptation models cannot fully underlie the perceptual data. The twelve obtained roughness models can describe a large set of perceptual roughness data quite well, but unfortunately the complete
dataset cannot be fit by a single model. As said before, in the earlier research [15], the author was not able to derive a meaningful temporal cue from rolling sounds. As depicted in figure 6.19, the advanced roughness models reveal a systematic change of roughness with the physical parameters of the rolling object. It will therefore be promising to use this roughness model in future studies in which the perceptual evaluation and the underlying perceptual attributes will be investigated.
Bibliography

[1] American National Standards Institute
   *American national psychoacoustical terminology.*

   *Ein Berechnungsverfahren der Rauhigkeit.*

[3] Breebaart, D.J.
   *Modeling binaural signal detection.*

   *The cochlea.*
   New York: Springer, 1996

   *Psychoacoustical roughness: implementation of an optimized model.*
   Acustica, vol. 83, 1997, p. 113–123

   *A quantitative model of the “effective” signal processing in the auditory system. I. Model structure.*

   *The method of paired comparisons.*
   In: Griffin’s Statistical Monographs & Courses. Ed. by M.G. Kendall
   London: Charles Griffin & Company, 1963

[8] Ellermeier, W., M. Mader and P. Daniel
   *Scaling the unpleasantness of sounds according to the BTL model: ratio-scale representation and psychoacoustical analysis.*

[9] Fastl, H.
   *Roughness and temporal masking patterns of sinusoidally amplitude modulated broadband noise.*
   In: Psychophysics and physiology of hearing. Ed. by E.F. Evans and J.P. Wilson
Complex tone processing in the primary cortex of the awake monkey. I. Neural ensemble correlates of roughness.

Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localisation.

Critical bandwidth and consonance in relation to cochlear frequency-position coordinates.

[13] Hartmann, W.M.
Signals, sound and sensation.
Woodbury (New York): American Institute of Physics, 1997

[14] Helmholtz, H.L.F.
On the sensations of tone as a physiological basis for the theory of music.
Translated from the German: Lehre von den Tonempfindungen, 1863

The sound of rolling objects.

[16] Immerseel, L.M. van and J. Martens
Pitch of voiced/unvoiced determination with an auditory model.

[17] Immerseel, L.M. van
Een functioneel gehoormodel voor de analyse van spraak bij spraakherkenning.
Dissertation, Universiteit Gent, Gent, 1993

[18] Irino, T. and R.D. Patterson
Temporal asymmetry in the auditory system.

[19] Irino, T. and R.D. Patterson
A time-domain, level-dependent auditory filter: the gammachirp.

[20] Irino, T. and M. Unoki
An analysis/synthesis auditory filterbank based on an IIR implementation of the gammachirp.

[21] Irino, T. and R.D. Patterson
A compressive gammachirp auditory filter for both physiological and psychophysical data.
[22] Jourdes, V.J.P.
   *Estimation of perceived roughness.*
   Master’s project report, Technische Universiteit Eindhoven, Eindhoven, August 2004

[23] Kemp, S.
   *Roughness of frequency-modulated tones.*
   Acustica, vol. 50, 1982, p. 126–133

[24] Leman, M.
   *Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model.*
   In: Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, December 7–9, 2000

[25] Meddis, R.
   *Simulation of mechanical to neural transduction in the auditory receptor.*

[26] Meddis, R.
   *Simulation of auditory-neural transduction: further studies.*

[27] Meddis, R., M.J. Hewitt and T.M. Shackleton
   *Implementation of a computation model of the inner hair-cell/auditory-nerve synapse.*

[28] Miśkiewicz, A.
   *Roughness of low-frequency pure tones.*
   Submitted for publication to Acta Acustica

[29] Patterson, R.D.
   *Auditory filter shapes derived with noise stimuli.*

[30] Patterson, R.D.
   *The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold.*

[31] Patterson, R.D.
   *The sound of a sinusoid: spectral models.*

[32] Patterson, R.D.
   *The sound of a sinusoid: time-interval models.*

[33] Patterson, R.D. and T. Irino
   *Modeling temporal asymmetry in the auditory system.*
Bibliography


[34] Patterson, R.D., M. Unoki and T. Irino
  Extending the domain of center frequencies for the compressive gammachirp auditory filter.

[35] Pflueger, M., R. Hoeldrich and W. Riedler
  A nonlinear model of the peripheral auditory system.
  IEM-report 02/98, Institute of Electronic Music and Acoustics, Graz

[36] Phillips, D.P., S.E. Hall and S.E. Boehnke
  Central auditory onset responses, and temporal asymmetries in auditory perception.

[37] Pressnitzer, D. and S. McAdams
  Two phase effects in roughness perception.

[38] Schrader, J.E.
  A Matlab implementation of a model of auditory roughness.

[39] Seneff, S.
  A joint synchrony/mean-rate model of auditory speech processing.

[40] Slaney, M.
  An efficient implementation of the Patterson-Holdsworth auditory filterbank.
  Apple Computer Group, 1993
  Apple Computer Technical Report #35

[41] Stewart, R.W.
  DSPedia.
  Loch Lomond (Scotland): Version 3.2/14/01, April 2002
  Course material

  A revised model of the inner-hair cell and auditory-nerve complex.

  Adaptation in a revised inner-hair cell model.

[44] Terhardt, E.
  On the perception of periodic sound fluctuations (roughness).
[45] Unoki, M., T. Irino and R.D. Patterson
Improvement of an IIR asymmetric compensation gammachirp filter.

[46] Vogel, A.
Über den Zusammenhang zwischen Rauhigkeit und Modulationsgrad.

[47] Vogel, A.
Roughness and its relation to the time-pattern of psychoacoustical excitation.

[48] Zwicker, E. and H. Fastl
Psychoacoustics, facts and models.

[49] Zwicker, E. and R. Feldtkeller
Das Ohr als Nachrichtempfänger.
Appendix A

Matlab code

A.1 Main program

%% Auditory roughness estimation using several modules
%%
%% Author: Ronnie Duisters
%% Group: Human Technology Interaction
%% Department of Technology Management
%% Eindhoven University of Technology

% R = roughcalc(signal, L, fs, winlength, dN, alpha, switches)
%
% inputs:
% signal input signal
% L preferred input level in dB SPL (default = 60)
% fs sample frequency in Hz (default = 48000)
% winlength length of analysis window in s (default = 0.2)
% dN sample interval to be processed by the roughness extraction model
% (default = [round(0.3*fs) round(0.5*fs)])
% alpha roughness index (default = 1.6)
% switches array selecting the preferred models
% (default = [2 2 1 3])
%
% outputs:
% R roughness value
%
% switches [a b c d]
% a = middle ear filter:
% 0 = no filter
% 1 = Van Immerseel and Martens
% 2 = Pflueger et al.
% b = filterbank:
% 1 = Gammatone filterbank
% 2 = Gammachirp filterbank
% c = hair cell model:
% 0 = no hair cell model
% 1 = Van Immerseel and Martens hair cell model
% 2 = Dau et al. adaptation model
% 3 = Meddis hair cell model
% d = Roughness model
% 1 = SIM model I
% 2 = SIM model II
% 3 = modified Aures model (Jourdes)

function R = roughcalc(signal, L, fs, winlength, dN, alpha, switches)

global numH7 denH7 numH14 denH14 numH30 denH30 numH36 denH36 numH66 denH66

if nargin < 7
    switches = [2 2 1 3];
end
if nargin < 6
    alpha = 1.6;
end
if nargin < 5
    dN = [round(0.3*fs) round(0.5*fs)];
end
if nargin < 4
    winlength = 200e-3;
end
if nargin < 3
    fs = 48000;
end
if nargin < 2
    L = 60;
end
if nargin < 1
    help roughcalc;
    error('no input signal defined!');
end

winlength = winlength * fs;
wsize = round(min(length(signal), winlength));

signal = signal(:); % make signal a row-vector
signal = signal(1:wsize);

disp('adapting signal level...');
signal = AdaptLevel(signal, L, switches);

% outer and middle-ear filtering
if switches(1) == 1
    % outer and middle ear filter as used by Van Immerseel and Martens
    disp('filtering outer and middle ear...');
    wr = 2 * pi * 4e3;
    num = wr^2;
\[
\text{den} = [1 0.33*wr \ wr^2]; \\
[num2, den2] = \text{bilinear}(\text{num, den, fs}); \\
\text{signal} = \text{filter}(\text{num2, den2, signal}); \\
\]

\text{elseif} \ \text{switches}(1) == 2 \\
\quad \text{disp}('\text{filtering outer and middle ear...'}); \\
\quad \% \text{Outer and middle ear combined bandpass filter} \\
\quad \% \text{(Pflueger, Hoeldrich, Riedler, Sep 1997)} \\
\quad \% \text{Highpass component} \\
\quad \text{b} = 0.109*[1 1]; \\
\quad \text{a} = [1 -2.5359 \ 3.9295 \ -4.7532 \ 4.7251 \ -3.5548 \ 2.139 \ -0.9879 \ 0.2836]; \\
\quad \% \text{Lowpass component} \\
\quad \text{d} = [1 -2 1]; \\
\quad \text{c} = [1 -2*0.95 \ 0.95^2]; \\
\quad \text{signal} = \text{filter}([\text{b, d}, \text{conv(a, c)}], \text{signal}); \\
\text{else} \\
\quad \text{disp}('\text{no outer- and middle-ear filtering'}); \\
\end{\text{elseif}}

\% \text{auditory filterbanks} \\
\text{if} \ \text{switches}(4) == 3 \\
\quad \text{erbres} = 0.5; \\
\quad \text{erbmax} = 38; \\
\text{else} \\
\quad \text{erbres} = 1; \\
\quad \text{erbmax} = 40; \\
\\text{end} \\
\text{erbr} = \text{erbres:erbres:erbmax}; \\
\text{fc} = 229 * (10.^(\text{erbr} / 21.4) - 1); \\
\text{Nch} = \text{length(\text{erbr});} \\
\text{if} \ \text{switches}(2) == 1 \\
\quad \% \text{gammatone filterbank} \\
\quad \text{disp}('\text{performing gammatone filtering...'}); \\
\quad \text{fcoefs} = \text{MakeERBFilters(\text{fs, fc(length(fc):-1:1), 26, 1.019});} \\
\quad \text{fcoefs} = \text{fcoefs(Nch:-1:1,:}); \\
\quad \text{exc} = \text{ERBFilt} \text{erBank(\text{signal, fcoefs});} \\
\text{elseif} \ \text{switches}(2) == 2 \\
\quad \% \text{compressive gammachirp filterbank} \\
\quad \text{disp}('\text{performing gammachirp filtering...'}); \\
\quad \text{n} = 4; \\
\quad \text{b1} = 1.81; \\
\quad \text{c1} = -2.96; \\
\quad \text{b2} = 2.17; \\
\quad \text{c2} = 2.20; \\
\quad \text{fr1} = \text{fc}; \\
\quad \text{fr1} = \text{fr1} + \text{c1} * \text{b1} * (24.7 + 0.107939 \ * \text{fr1}) / \text{n}; \\
\quad \text{frat} = 0.573 + 0.0101 \ * \text{L}; \\
\quad \text{fr2} = \text{frat} \ * \text{fr1}; \\
\quad \text{fcoefs} = \text{MakeERBFilt} \text{ers(\text{fs, fc(length(fc):-1:1), 26, b1});} \\
\quad \text{fcoefs} = \text{fcoefs(Nch:-1:1,:}); \\
\quad \text{exc} = \text{ERBFilt} \text{erBank(\text{signal, fcoefs});}
for i = 1:length(exc(:,1)),
    [bgc1, agc1] = MakeAsymCmpFilters(fs, fr1(i), n, b1, c1);
    exc(i,:) = filter(bgc1, agc1, exc(i,:));
    [bgc2, agc2] = MakeAsymCmpFilters(fs, fr2(i), n, b2, c2);
    exc(i,:) = filter(bgc2, agc2, exc(i,:));
end

if level == 80, % low-frequency channels become unstable for the 80 dB condition
    for v = 1:2,
        exc(v,:) = 0 * exc(v,:);
    end
end

% adaptation models

if switches(3) == 1
    % hair cell model by Van Immerseel and Martens
    disp('applying vI&M hair cell model...');
    y0 = 0.4472;
    exc = max(exc + y0, 0);
    fspon = 0.05e3;
    fsat = 0.15e3;
    r = 0.86;
    tau1 = 8e-3;
    tau2 = 40e-3;
    num = [tau1-r*tau1+r*tau2 1];
    den = [tau1*tau2 tau1+tau2 1];
    [num2, den2] = bilinear(num, den, fs);
    for i = 1:length(exc(:,1)),
        q(i,:) = filter(num2, den2, exc(i,:));
        f(i,:) = (fsat * exc(i,:)) ./ (((sqrt(fsat/fspon) - 1) * ...
            sqrt(y0)) + sqrt(q(i,:))).^2;
    end
    [bn, an] = butter(3, 600 / (fs / 2));
    exc = f;
    exc = filter(bn, an, exc);
    exc = max(exc-50, 0);
elseif switches(3) == 2
    % adaptation model by Dau et al.
    disp('applying Dau et al. adaptation model...');
    exc = max(exc, 0);
    [bn, an] = butter(5, 770 / (fs / 2));
    exc = filter(bn, an, exc, [], 2);
    for i = 1:Nch,
        f(i,:) = fadapt(exc(i,:), fs);
    end
    exc = f;
    exc = max(exc, 0);
elseif switches(3) == 3
    % hair cell model by Meddis
    disp('applying Meddis hair cell model...');
    exc = max(exc, 0);
exc = MeddisHairCell(exc, fs);
exc = max(exc-65, 0)
else
    disp('no hair cell model');
end

% roughness extraction models

if switches(7) == 1
    % Leman roughness extraction model I
    disp('estimating roughness...');
    w = kron(ones(Nch, 1), hamming(wsize)');
    wexc = w .* exc(:,1:wsize);
    N = nextpow2(wsize);
    wsize = 2^N;
    fbin = fs / wsize;
    F = FilterWeights(erbmax/erbres, round(300/fbin), fc, 1, round(300/fbin), 1);
    if switches(2) == 0
        if switches(3) == 1 % gammatone,
            Rmax = [0 1.1 1.4 1.5 1 0.71 0.56 0.41 0.32]; % Van Immerseel & Martens
        elseif switches(3) == 2
            Rmax = [0 1.1 1.4 1.6 1 0.65 0.54 0.38 0.26]; % gammatone, Dau et al.
        elseif switches(3) == 3
            Rmax = [0 0.9 1 1.2 1 0.65 0.5 0.38 0.26]; % gammatone, Meddis
        end
    elseif switches(2) == 2
        if switches(3) == 1 % gammachirp,
            Rmax = [0 0.3 1.3 1.4 1 0.74 0.73 0.42 0.38]; % Van Immerseel & Martens
        elseif switches(3) == 2
            Rmax = [0 0.3 0.94 1.4 1 0.7 0.6 0.38 0.31]; % gammachirp, Dau et al.
        elseif switches(3) == 3
            Rmax = [0 0.84 0.94 1 1 0.82 0.64 0.44 0.35]; % gammachirp, Meddis
        end
    end
    fc = [0 125 250 500 1e3 2e3 4e3 8e3 16e3];
    ERBrate = 21.4 .* log10(4.37 * fc / 1000 + 1);
    gzi = interp1(ERBrate, Rmax, erbr, 'cubic');
    F = diag(gzi, 0) * F;
    l = length(F(1,:));
    sexc = fft(wexc, wsize, 2);
    sexc = sexc(:,1:l);
    DC = kron(sexc(:,1) / 2, ones(1, l));
    Bd = F .* sexc ./ DC;
    Rf = sum(abs(Bd) .^ alpha, 1);
    R = sum(Rf) / Nch;
    R = abs(R);
    Rf = abs(Rf);
elseif switches(7) == 2
    % Leman roughness extraction model II
    disp('estimating roughness...');
    w = kron(ones(Nch, 1), hamming(wsize)');
    wexc = w .* exc(:,1:wsize);
N = nextpow2(wsize);
wsiz = 2^N;
fbin = fs / wsize;
F = FilterWeights(erbmax/erbres, round(300/fbin), fc, 1, round(300/fbin), 1);
if switches(2) == 0
    if switches(3) == 1 % gammatone,
        Rmax = [0 1.1 1.4 1.5 1 0.71 0.56 0.41 0.32]; % Van Immerseel & Martens
    elseif switches(3) == 2
        Rmax = [0 1.1 1.4 1.6 1 0.65 0.54 0.38 0.26]; % gammatone, Dau et al.
    elseif switches(3) == 3
        Rmax = [0 0.9 1 1.2 1 0.65 0.5 0.38 0.26]; % gammatone, Meddis
    end
elseif switches(2) == 2
    if switches(3) == 1 % gammachirp,
        Rmax = [0 0.3 1.3 1.4 1 0.74 0.73 0.42 0.38]; % Van Immerseel & Martens
    elseif switches(3) == 2
        Rmax = [0 0.3 0.94 1.4 1 0.7 0.6 0.38 0.31]; % gammachirp, Dau et al.
    elseif switches(3) == 3
        Rmax = [0 0.84 0.94 1 1 0.82 0.64 0.44 0.35]; % gammachirp, Meddis
    end
end
fc = [0 125 250 500 1e3 2e3 4e3 8e3 16e3];
ERBrate = 21.4 .* log10(4.37 * fc / 1000 + 1);
gzi = interp1(ERBrate, Rmax, erbr, 'cubic');
F = diag(gzi, 0) * F;
l = length(F(1,:));
sexc = fft(wexc, wsize, 2);
sexc = sexc(:,1:l);
DC = sum(kron(sexc(:,1) / 2, ones(1, l)));
Bd = F .* sexc;
Bd2 = sum(Bd, 1) ./ DC;
Rf = abs(Bd2) .^ alpha;
R = sum(Rf) / Nch;
elseif switches(4) == 3
    % adapted Aures model
    exc = exc(:, dN(1):dN(2));
disp('extracting roughness...');
    % calculation of the DC values
    etmp = abs(exc);
s0 = kron(ones(1, length(exc)), mean(etmp, 2));
    etmp = abs(exc);
end
% Weighting filtering
sBP(1:11,:) = filter(numH7, denH7, excd(1:11,:), [], 2);
sBP(12:28,:) = filter(numH14, denH14, excd(12:28,:), [], 2);
sBP(29:36,:) = filter(numH30, denH30, excd(29:36,:), [], 2);
sBP(37:65,:) = filter(numH36, denH36, excd(37:65,:), [], 2);
sBP(66:76,:) = filter(numH66, denH66, excd(66:76,:), [], 2);
sBPrms = rms(sBP);
rexc = rms(exc);
maxi = max(rexc);
if maxi > 0
Appendix A. Matlab code

```matlab
calib = rexc / maxi;
else
    calib = 0;
end

% modulation depth estimation
for k = 1:Nch,
    % calibration factor
    if s0(k) > 0
        mdepth(k) = sBPrms(k) / s0(k);
        mdepth(k) = mdepth(k) * calib(k);
    else
        mdepth(k) = 0;
    end

    % calculation of the shifted cross correlation factor
    if k < Nch - 1,
        amount = 0.003 * fs;
        ki(k) = shiftcov(sBP(k,:), sBP(k+2,:), amount);
    end
end

% definition of gzi
if switches(2) == 1
    if switches(3) == 1
        % gammatone, Van Immerseel & Martens
        Rmax = [0 0.35 0.8 0.99 1 0.75 0.57 0.53 0.42];
    elseif switches(3) == 2
        % gammatone, Dau et al.
        Rmax = [0 0.3 1 1 1 0.64 0.49 0.51 0.45];
    elseif switches(3) == 3
        % gammatone, Meddis
        Rmax = [0 0.35 0.8 0.9 1 0.65 0.47 0.43 0.32];
    end
elseif switches(2) == 2
    if switches(3) == 1
        % gammachirp, Van Immerseel & Martens
        Rmax = [0 0.35 0.8 0.99 1 0.88 0.64 0.68 0.65];
    elseif switches(3) == 2
        % gammachirp, Dau et al.
        Rmax = [0 0.05 0.15 0.99 1 0.73 0.61 0.77 0.6];
    elseif switches(3) == 3
        % gammachirp, Meddis
        Rmax = [0 0.35 0.85 0.9 1 0.82 0.55 0.56 0.5];
    end
end

fc = [0 125 250 500 1e3 2e3 4e3 8e3 16e3];
ERBrate = 2 * 21.4 .* log10(4.37 * fc / 1000 + 1);
gzi = interp1(ERBrate, Rmax, erbr, 'cubic');

% calculate specific roughness ri
ri(1:7) = (gzi(1:7) .* mdepth(1:7) .* ki(1:7)).^2;
ri(8:Nch-3) = (gzi(8:Nch-3) .* mdepth(8:Nch-3) .* ki(6:Nch-5) .* ki(8:Nch-3)).^2;
ri(Nch-2) = (gzi(Nch-2) .* mdepth(Nch-2) .* ki(Nch-4)).^2;
ri(Nch-1) = (gzi(Nch-1) .* mdepth(Nch-1) .* ki(Nch-3)).^2;
ri(Nch) = (gzi(Nch) .* mdepth(Nch) .* ki(Nch-2)).^2;

R = sum(ri);
end```
Appendix A. Matlab code

A.2 Subfunctions

A.2.1 roughinit.m

% [], = roughinit(fs, switches);
% Initialises the roughness model by creating variables and saving them
% to the file roughinit.mat.
% Variables are created depending on sample frequency 'fs', and the
% selected module implementations defined in 'switches'..
%
function [], = roughinit(fs, switches);

global numH7 denH7 numH14 denH14 numH30 denH30 numH36 denH36 numH66 denH66

if nargin < 2
    switches = [2 2 1 3];
end
if nargin < 1
    fs = 48000;
end

t = 0:1/fs:0.6;

if switches(4) == 3
    disp('calculating filter weights...');
    [numH7, denH7, numH14, denH14, numH30, denH30, numH36, denH36, numH66, denH66] ...
        = fir_hweight(fs);
    twin = 0.5;
    dN = [round(0.3*fs) round(0.3*fs+fs/70)];
    ref1 = (1 + sin(2 * pi * 70 * t)) .* sin(2 * pi * 1e3 * t);
    win = sin(0.5 * pi * 50 .* (0:1/fs:20e-3)).^2;
    ref1 = [ref1(1:length(win)) .* win ref1(length(win)+1:length(ref1)-length(win)) ...
            ref1(length(ref1)-length(win)+1:length(ref1)) .* fliplr(win)];
    disp('calculating asper reference...');
    R01 = roughcalc(ref1, 60, fs, twin, dN, 1.6, switches);
    save roughinit fs numH7 denH7 numH14 denH14 numH30 denH30 numH36 denH36 numH66 denH66 switches R01
else
    twin = 0.6;
    dN = [floor(0.3*fs) floor(0.5*fs)];
    ref1 = (1 + sin(2 * pi * 70 * t)) .* sin(2 * pi * 1e3 * t);
    win = sin(0.5 * pi * 50 .* (0:1/fs:20e-3)).^2;
    ref1 = [ref1(1:length(win)) .* win ref1(length(win)+1:length(ref1)-length(win)) ...
            ref1(length(ref1)-length(win)+1:length(ref1)) .* fliplr(win)];
disp('calculating asper reference...');
R01 = roughcalc(ref1, 60, fs, twin, dN, 1.6, switches);

save roughinit fs switches R01
end

A.2.2 AdaptLevel.m

% out = Adaptlevel(in, level, switches);
% % Scales signal 'in' to the demanded digital 'level' in dB SPL % with respect to the selected module implementations defined % in array 'switches'.
%

function out = AdaptLevel(in, level, switches)

% Calculate current rms value
[rows cols] = size(in);
rmsSig = sqrt(sum(in.^2, 2) / cols);

% Calculate the factor to use for adapting the signal, % depending on the adaptation model
if switches(3) == 2
    factor = 10.^(level / 20 - log10(rmsSig));
else
    factor = 10.^(((level - 30) / 20 - log10(rmsSig));
end

% Adapt the signal
out = zeros(rows, cols);
for i = 1:rows,
    out(i,:) = in(i,:) * factor(i);
end

A.2.3 rms.m

% out = rms(in);
% % Calculates the rms value of array 'in'
%
function out = rms(in)

[m, n] = size(in);
out = sqrt(sum(x.^2, 2) / n);
A.2.4 shiftcov.m

% Author: V.J.P. Jourdes
% This function calculates the maximum of the cross correlation between two
% signal calculated after shifting one of them of a certain number of
% samples.
% y = shifcov(f1, f2, amount);
% f1: signal 1
% f2: signal 2
% amount: defines the maximum number of samples that the signal will be
% shifted.
% function y = shiftcov(f1, f2, amount)

L = length(f1);
ff2 = f2;
x = 1;

for i = 1:10:amount,
    cfac = cov(f1, f2);
den = diag(cfac);
den = sqrt(den * den')
    if den(2,1) > 0
        r(x) = cfac(2,1) / den(2,1);
    else
        r(x) = 0;
    end
    x = x + 1;
f2(1) = [];
f2(L) = 0;
end

for i = 1:10:amount,
f1 = [zeros(1,i) f1(1:L-i)];
cfac = cov(f1, ff2);
den = diag(cfac);
den = sqrt(den * den')
    if den(2,1) > 0
        r(x) = cfac(2,1) / den(2,1);
    else
        r(x) = 0;
    end
    x = x + 1;
end

y = max(r);
A.2.5 Other subfunctions

Some subfunctions were obtained from the packages mentioned in table 2.2. These subfunctions are summarised in table A.1. Note that the subfunctions MakeERBFilters.m, AsymCmpCoef.m and FilterWeights.m were modified in our model.

<table>
<thead>
<tr>
<th>Subfunction</th>
<th>Package</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>MakeERBFilters.m</td>
<td>Auditory Toolbox</td>
<td>Slaney</td>
</tr>
<tr>
<td>ERBFilterbank.m</td>
<td>Auditory Toolbox</td>
<td>Slaney</td>
</tr>
<tr>
<td>MakeAsymCmpFilters.m</td>
<td>GCFB</td>
<td>Irino &amp; Unoki</td>
</tr>
<tr>
<td>AsymCmpCoef.m</td>
<td>GCFB</td>
<td>Irino &amp; Unoki</td>
</tr>
<tr>
<td>fadapt.m</td>
<td>-</td>
<td>Breebaart</td>
</tr>
<tr>
<td>MeddisHairCell.m</td>
<td>Auditory Toolbox</td>
<td>Slaney</td>
</tr>
<tr>
<td>FilterWeights.m</td>
<td>IPEM Toolbox</td>
<td>Leman, Lesaffre &amp; Tanghe</td>
</tr>
</tbody>
</table>

Table A.1: Externally obtained subfunctions.

A.3 Implementation example

A.3.1 roughplotsasym.m

% This script generates roughness plots for ramped and damped signals
% as a function of modulation depth and centre frequency.
%
clear all

global numH7 denH7 numH14 denH14 numH30 denH30 numH36 denH36 numH66 denH66

load roughinit

t = 0:1/fs:0.6;
m = [0.4 0.6 0.8];
fc = [2500 5e3 10e3];
fmod = 70;
alpha = 1.6;

% define observation interval
if switches(4) == 3
    twin = 0.4;
    dN = [round(0.3*fs) round(0.3*fs+fs/fmod)];
else
    twin = 0.6;
    dN = [floor(0.3*fs) floor(0.5*fs)];
end

Esaw = zeros(3, length(t));
Esaw2 = zeros(3, length(t));

for i = 1:2,
    Esaw(1,:) = (1/i) * cos(2 * pi * i * fmod .* t - pi/2) + Esaw(1,:);
end
for i = 1:4,
    Esaw(2,:) = (1/i) * cos(2 * pi * i * fmod .* t - pi/2) + Esaw(2,:);
end
for i = 1:7,
    Esaw(3,:) = (1/i) * cos(2 * pi * i * fmod .* t - pi/2) + Esaw(3,:);
end
for i = 1:2,
    Esaw2(1,:) = (1/i) * cos(2 * pi * i * fmod .* t + pi/2) + Esaw2(1,:);
end
for i = 1:4,
    Esaw2(2,:) = (1/i) * cos(2 * pi * i * fmod .* t + pi/2) + Esaw2(2,:);
end
for i = 1:7,
    Esaw2(3,:) = (1/i) * cos(2 * pi * i * fmod .* t + pi/2) + Esaw2(3,:);
end
for i = 1:length(fc),
    fprintf(‘\nfc = %4.0i\n’, fc(i));
    for j = 1:length(m),
        % damped
        signal = (1 + m(j) * (Esaw(i,:) / max(Esaw(i,:)))) .* cos(2 * pi * fc(i) .* t - pi/2);
        win = sin(0.5 * pi * 50 .* (0:1/fs:20e-3)).^2;
        signal = [signal(1:length(win)) .* win ... % window signal
                  signal(length(win)+1:length(signal)-length(win)) ...
                  signal(length(signal)-length(win)+1:length(signal)) .* fliplr(win)];
        R1(i,j) = roughcalc(signal, 60, fs, twin, dN, alpha, switches);
        % ramped
        signal = (1 + m(j) * (Esaw2(i,:) / max(Esaw2(i,:)))) .* cos(2 * pi * fc(i) .* t + pi/2);
        win = sin(0.5 * pi * 50 .* (0:1/fs:20e-3)).^2;
        signal = [signal(1:length(win)) .* win ... % window signal
                  signal(length(win)+1:length(signal)-length(win)) ...
                  signal(length(signal)-length(win)+1:length(signal)) .* fliplr(win)];
        R2(i,j) = roughcalc(signal, 60, fs, twin, dN, alpha, switches);
    end
end
% convert scale to asper
R1 = R1 / R01;
R2 = R2 / R01;
% auto axis scaling
minax = min(min(min(R1)), min(min(R2)));
maxax = max(max(max(R1)), max(max(R2)));
minax = minax - 0.1 * max(abs(minax), abs(maxax));
maxax = maxax + 0.1 * max(abs(minax), abs(maxax));
legtop = maxax - 0.1 * maxax;
figure(11);
clear;
subplot(231), plot(m, R1(1,:), 'k.-', m, R2(1,:), 'ko--');
xlabel('modulation depth');
ylabel('roughness (asper)');
axis([0.3 0.9 minax maxax]);
text(0.35, legtop, 'f_c = 2.5 kHz');

subplot(232), plot(m, R1(2,:), 'k.-', m, R2(2,:), 'ko--');
xlabel('modulation depth');
axis([0.3 0.9 minax maxax]);
text(0.35, legtop, 'f_c = 5 kHz');

subplot(233), plot(m, R1(3,:), 'k.-', m, R2(3,:), 'ko--');
xlabel('modulation depth');
legend('R_d', 'R_r', 1);
axis([0.3 0.9 minax maxax]);
text(0.35, legtop, 'f_c = 10 kHz');
# Appendix B

## Listing of used symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a, b, b_1, b_2$</td>
<td>Gammatone and gammachirp parameters</td>
</tr>
<tr>
<td>$A_1, B_1, \lambda$</td>
<td>Greenwood parameters</td>
</tr>
<tr>
<td>$A_2, B_2, \beta$</td>
<td>Meddis hair cell model parameters</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Dependency of roughness on modulation depth</td>
</tr>
<tr>
<td>$c, c_1, c_2$</td>
<td>Chirp parameters</td>
</tr>
<tr>
<td>$\chi$</td>
<td>Cross-correlation factor</td>
</tr>
<tr>
<td>$d$</td>
<td>Van Immerseel &amp; Martens hair cell parameter</td>
</tr>
<tr>
<td>$E_{\text{ramped}}$</td>
<td>Modulation components of the ramped signal</td>
</tr>
<tr>
<td>$\text{ERB}$</td>
<td>Equivalent Rectangular Bandwidth</td>
</tr>
<tr>
<td>$\text{ERB}_{\text{rate}}$</td>
<td>Equivalent Rectangular Bandwidth rate</td>
</tr>
<tr>
<td>$\epsilon_1 \ldots \epsilon_5$</td>
<td>Meddis hair cell parameters</td>
</tr>
<tr>
<td>$f_c$</td>
<td>Centre frequency</td>
</tr>
<tr>
<td>$f_{\text{mod}}$</td>
<td>Modulation frequency</td>
</tr>
<tr>
<td>$f_r$</td>
<td>Resonance frequency</td>
</tr>
<tr>
<td>$f_{r_1}, f_{r_2}, f_{\text{rat}}, f_{p1}$</td>
<td>Gammachirp frequency parameters</td>
</tr>
<tr>
<td>$\Delta f$</td>
<td>Frequency deviation</td>
</tr>
<tr>
<td>$F{\ldots}$</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>$F^{-1}{\ldots}$</td>
<td>Inverse Fourier transform</td>
</tr>
<tr>
<td>$\phi, \varphi$</td>
<td>Phase</td>
</tr>
<tr>
<td>$G_c$</td>
<td>General gammachirp filter</td>
</tr>
<tr>
<td>$G_{ca}$</td>
<td>Passive gammachirp filter</td>
</tr>
<tr>
<td>$G_{cc}$</td>
<td>Compressive gammachirp filter</td>
</tr>
<tr>
<td>$G_t$</td>
<td>Gammatone filter</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Asper scale calibration factor</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>Gamma function</td>
</tr>
<tr>
<td>$h$</td>
<td>Impulse response</td>
</tr>
<tr>
<td>$h_{\text{LP}}$</td>
<td>Lowpass filter impulse response</td>
</tr>
</tbody>
</table>

Table B.1: Listing of used symbols.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>Transfer function</td>
</tr>
<tr>
<td>$H_{a1}, H_{a2}$</td>
<td>Asymmetric compensation filters</td>
</tr>
<tr>
<td>$H_{HP}$</td>
<td>Highpass filter</td>
</tr>
<tr>
<td>$H_{i}$</td>
<td>Modulation transfer function for channel $i$</td>
</tr>
<tr>
<td>$H_{LP}$</td>
<td>Lowpass filter</td>
</tr>
<tr>
<td>$\mathcal{H}{\ldots}$</td>
<td>Hilbert transform</td>
</tr>
<tr>
<td>$i$</td>
<td>Auditory channel number</td>
</tr>
<tr>
<td>$j$</td>
<td>Imaginary symbol; $j = \sqrt{-1}$</td>
</tr>
<tr>
<td>$k$</td>
<td>Sample number</td>
</tr>
<tr>
<td>$L$</td>
<td>Sound pressure level (dB SPL)</td>
</tr>
<tr>
<td>$l_{\text{cochlea}}$</td>
<td>Length of the cochlea</td>
</tr>
<tr>
<td>$m$</td>
<td>Modulation depth</td>
</tr>
<tr>
<td>$m_{\text{SFM}}$</td>
<td>Modulation index of an SFM sinusoid</td>
</tr>
<tr>
<td>$m_{\text{damped}}$</td>
<td>Modulation depth of the damped signal</td>
</tr>
<tr>
<td>$m_{\text{ramped}}$</td>
<td>Modulation depth of the ramped signal</td>
</tr>
<tr>
<td>$\Delta m$</td>
<td>Modulation-depth difference between the damped and ramped signal</td>
</tr>
<tr>
<td>$n$</td>
<td>Filter order</td>
</tr>
<tr>
<td>$N_{\text{ch}}$</td>
<td>Number of channels</td>
</tr>
<tr>
<td>$N_{\text{hair cells}}$</td>
<td>Number of hair cells</td>
</tr>
<tr>
<td>$N_{\text{mod}}$</td>
<td>Number of modulation components</td>
</tr>
<tr>
<td>$N_{\text{samples}}$</td>
<td>Number of samples</td>
</tr>
<tr>
<td>$p_0, p_1, p_2, p_3$</td>
<td>Gammachirp implementation parameters</td>
</tr>
<tr>
<td>$q_c$</td>
<td>Amount of neurotransmitter in the cleft</td>
</tr>
<tr>
<td>$q_{\text{ftp}}$</td>
<td>Amount of neurotransmitter in the free transmitter pool</td>
</tr>
<tr>
<td>$q_{\text{rs}}$</td>
<td>Amount of neurotransmitter in the reprocessing store</td>
</tr>
<tr>
<td>$r_i$</td>
<td>Specific roughness value</td>
</tr>
<tr>
<td>$R$</td>
<td>Roughness value</td>
</tr>
<tr>
<td>$\rho_{\text{sat}}$</td>
<td>Saturation rate</td>
</tr>
<tr>
<td>$\rho_{\text{spont}}$</td>
<td>Spontaneous rate</td>
</tr>
<tr>
<td>$s$</td>
<td>Signal in the time domain</td>
</tr>
<tr>
<td>$S$</td>
<td>Signal in the frequency domain</td>
</tr>
<tr>
<td>$\tilde{s}$</td>
<td>Envelope of a signal</td>
</tr>
<tr>
<td>$s[k]$</td>
<td>Signal, discrete time domain</td>
</tr>
<tr>
<td>$s(t)$</td>
<td>Signal, continuous time domain</td>
</tr>
<tr>
<td>$s(z)$</td>
<td>$z$-domain signal</td>
</tr>
<tr>
<td>$s_0$</td>
<td>Signal DC value</td>
</tr>
</tbody>
</table>

Table B.2: Listing of used symbols, continued.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{BP}$</td>
<td>Bandpass filtered signal</td>
</tr>
<tr>
<td>$s_{hwr}$</td>
<td>Half-wave rectified signal</td>
</tr>
<tr>
<td>$s_{af}$</td>
<td>Signal from the auditory filterbank</td>
</tr>
<tr>
<td>$s_{adapt}$</td>
<td>Signal from the adaptation model</td>
</tr>
<tr>
<td>$s_{damped}$</td>
<td>Ramped signal</td>
</tr>
<tr>
<td>$s_{LP}$</td>
<td>Lowpass filtered signal</td>
</tr>
<tr>
<td>$s_{noise}$</td>
<td>White noise signal</td>
</tr>
<tr>
<td>$s_{off}$</td>
<td>Signal offset value</td>
</tr>
<tr>
<td>$s_{ramped}$</td>
<td>Damped signal</td>
</tr>
<tr>
<td>$s_{SAM}$</td>
<td>Sinusoidally amplitude modulated signal</td>
</tr>
<tr>
<td>$s_{SAM,noise}$</td>
<td>Sinusoidally amplitude modulated noise signal</td>
</tr>
<tr>
<td>$s_{SFM}$</td>
<td>Sinusoidally frequency modulated signal</td>
</tr>
<tr>
<td>$s_{fade,in}$</td>
<td>Fade-in signal for windowing</td>
</tr>
<tr>
<td>$\bar{s}$</td>
<td>Mean of signal $s$</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>Bias-corrected sample variance</td>
</tr>
<tr>
<td>$\varsigma_1, \varsigma_2$</td>
<td>Slopes of the Aures auditory filter weighting</td>
</tr>
<tr>
<td>$t$</td>
<td>Continuous time</td>
</tr>
<tr>
<td>$t_{win}$</td>
<td>Time length of observation window</td>
</tr>
<tr>
<td>$\tau_a, \tau_b$</td>
<td>Time constants of the Van Immerseel &amp; Martens hair cell model</td>
</tr>
<tr>
<td>$\tau_1, \ldots, \tau_5$</td>
<td>Time constants of the Dau et al. adaptation model</td>
</tr>
<tr>
<td>$u$</td>
<td>Unit step function</td>
</tr>
<tr>
<td>$v$</td>
<td>Index</td>
</tr>
<tr>
<td>$w_{\text{hamming}}$</td>
<td>Hamming window</td>
</tr>
<tr>
<td>$x$</td>
<td>Distance from the apex</td>
</tr>
<tr>
<td>$z$</td>
<td>$z$-domain variable</td>
</tr>
</tbody>
</table>

Table B.3: *Listing of used symbols, continued.*