Czech Technical University in Prague

Faculty of Electrical Engineering

Doctoral Thesis

February, 2015 Václav Vencovský

Czech Technical University in Prague

Faculty of Electrical Engineering
Department of Radioelectronics

Assessment of Sound by means of an Auditory Model - Prediction of Roughness

Doctoral Thesis

Václav Vencovský

Prague, February 2015

Ph.D. Programme: Electrical Engineering and Information Technology
Branch of study: Acoustics

Supervisor: Zdeněk Otčenášek

to Olga

This doctoral thesis was supported in part by the Ministry of Education, Youth and Sports of the Czech Republic in the Long Term Conceptual Development of Research Institutes grant of the Academy of Performing Arts in Prague: The "Sound quality" project; by Hlávkova Nadace; by the Czech Technical University in Prague by the grant No. CTU0909613 "Objective quality assessment of coded audio by means of an auditory model"; and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/204/OHK3/3T/13.

Acknowledgements

First of all I would like to thank to my supervisor Zdeněk Otčenášek for his support, ideas and criticism which made this work possible. For the same reasons, I thank also to my previous supervisor František Kadlec.

I thank to Marek Frič with whom I discussed issues related to voice and listening tests. He and also other colleagues at the Musical Acoustics Research Center, Academy of Performing Arts in Prague and Department of Radioelectronics, Czech Technical University in Prague kindly participated in the listening tests. I thank to František Rund, Libor Husník and professor Petr Maršálek from Czech Technical University in Prague for their help and organizing acoustic seminars and to professor Peter Svensson and others from the Acoustics group at Norwegian University of Science and Technology in Trondheim for their help and companionship during my study stay.

My thanks belong also to Alexandra Slobodníková for her help.

My greatest gratitude goes to my parents for their encouregment and support during my studies and to my wife Olga for her love and patience.

Václav Vencovský

Prague, February 2015

Abstract

The term roughness describes a specific – harsh, buzzy and rattling – sound sensation which may occur when listening to stimuli with fast temporal fluctuations, for example, amplitude- or frequency-modulated tones. Roughness is, as well as loudness or pitch, an important psychoacoustic parameter. The thesis describes a computational model predicting roughness of acoustic stimuli, composed of two successive stages: a peripheral and a central stage. The peripheral stage is composed of an auditory model which transforms the input acoustic stimulus into the simulated neural signal. The auditory model contains a set of algorithms simulating the function of the outer- and middle-ear, cochlear mechanics, the inner hair cells and auditory nerve synapse. The algorithms were adapted from the literature and composed into one model. The central stage – designed within the framework of this thesis – predicts roughness from the envelope of the simulated neural signal.

The peripheral stage of the roughness model employs a physical model of the basilar membrane (BM) response and cochlear hydrodynamics (the Nobili *et al.* cochlear model). Since this model simulates the limited frequency resolution of the peripheral ear, it is important for prediction of roughness. The author of this thesis adjusted the frequency selectivity of the model according to psychophysical masking data for harmonic complexes. The isointensity responses, input/output functions and impulse responses of the cochlear model are compared with data (reproduced from the literature) measured in the cochleae of live mammals. The model was as well verified using psychophysical masking data (reproduced from the literature) for pure tones and harmonic complexes. The results show that the Nobili *et al.* cochlear model can account for physiological and psychophysical phenomena which limits many cochlear models.

The roughness model is in this thesis used to predict the roughness of a large number of acoustic stimuli. The predicted roughness was compared with results of listening tests. The subjective data of roughness of the stimuli were reproduced from the literature or obtained by means of the listening tests conducted within the framework of the thesis. The roughness was reproduced from the literature for: sinusoidally amplitude-modulated tones, two tone stimuli composed of pure tones and harmonic complex tones, pseudo amplitude-modulated tones, stimuli with assymetrical temporal envelopes, sinusoidally frequency-modulated tones, unmodulated bandpass noise stimuli. The listening tests were conducted within the framework of this thesis for: sinusoidally amplitude-modulated harmonic complex tones, synthetic and real vowels /a/. The

predicted roughness agrees with the subjective data for most of the used stimuli. The largest discrepancies between the model predictions and the subjective data are for unmodulated bandpass noise stimuli and for real vowels. Both stimuli contain noise which complicates the roughness model performance. The roughness model covers both the effect of phase and the shape of the temporal envelope on roughness. This is its advantage in comparison to the roughness models known to the author.

Contents

List of Figures

List of Tables

Notation

Physical and signal processing notation[1]

A	amplitude
f	frequency
m	modulation index (e.g. amplitude-modulated tones)
p	pressure
t	time
ϕ	phase
Δf	frequency deviation (frequency-modulated tones)

Abbreviations

AM	amplitude modulated
AN	auditory nerve
BM	basilar membrane
CF	characteristic frequency
ERB	equivalent rectangular bandwidth
FM	frequency modulated
HI	hearing impaired
HL	hearing level
HSR	high spontaneous rate
IHC	inner hair cell

[1] In some parts of the text, an identical symbol is used with different meaning. This was motivated by the aim to follow the symbols given in the cited literature (for clarity). The used symbols are described in the text.

I/O	input/output
LSR	low spontaneous rate
MSR	medium spontaneous rate
NH	normal hearing
OHC	outer hair cell
RMS	root mean square
pAM	pseudo amplitude modulated
SAM	siunsoidally amplitude modulated
SFM	sinusoidally frequency modulated
SPL	sound pressure level

Chapter 1

Introduction

The term "roughness" was first used by von Helmholtz (1877) to describe a harsh, rattling and buzzy sound sensation which may occur when listening to stimuli with temporal fluctuations, for example, amplitude or frequency-modulated tones or unmodulated bandpass noises. The roughness is, as well as pitch or loudness, an important psychoacoustic parameter. It was given a ratio-scale unit called asper (Fastl & Zwicker, 2007).

Researchers observed that roughness together with loudness, fluctuation strength and sharpness contribute to the perception of annoyance (Fastl & Zwicker, 2007). Hence, roughness is an important issue in car industry in the field of vehicle noise control. Vehicle noise affects drivers in a range of ways, not only by potentially increasing the incidence of traffic accidents, but also through issues such as affecting decisions made by car buyers (Wang *et al.*, 2013). Roughness is also important for voice quality. Various pathologies of the human larynx lead to the voice with more roughness (dysphonic voices) (Patel *et al.*, 2012). For musical sounds, roughness is often related with dissonance (von Helmholtz, 1877).

The amount of perceived roughness of acoustic stimuli can be measured by means of listening tests. Since conducting listening tests is time consuming and expensive, objective methods which would give the same results are being sought. Although a number of various models predicting roughness have been developed (some of them are briefly reviewed below in Section 2.2), none have been nationally or internationally standardized yet (Wang *et al.*, 2013). This contrasts with models for other psychoacoustical parameters, for example, the international standard, ISO 532B, describes the

Zwicker model, which predicts loudness; and the German national standard, DIN 45692, describes a model predicting sharpness. The roughness models maybe, for example, employed for sound quality evaluation of vehicle noise (Wang *et al.*, 2013), and clinical assessment of voice and its rehabilitation (Howard *et al.*, 2012).

Some of the developed roughness models simulate the function of the peripheral ear. For example, the roughness model designed by Daniel & Weber (1997) takes into account the frequency selectivity of the peripheral ear; and the roughness model designed by Leman (2000) employs a computational model of the peripheral ear. This stems from a suggestion of von Helmholtz (1877) that the frequency selectivity of the human ear is limited and it thus cannot resolve spectral components of sound if they are too close in frequency. The spectral components then beat together in the ear which is assumed to cause the roughness sensation (von Helmholtz, 1877; Mathes & Miller, 1947; Vassilakis, 2001).

Aims of the thesis

The specific aim of this thesis is to design a roughness model which employs algorithms simulating the function of the peripheral ear. The model should predict roughness of various types of acoustic stimuli in agreement with results of listening tests. The roughness model developed in this thesis is composed of two successive stages: a peripheral stage and a central stage. The peripheral stage employs an auditory model (a model of the peripheral ear). The central stage then processes the outputs of the peripherals stage and predicts roughness.

Content of the thesis

The content of the thesis is as follows. Chapter 2 describes the term roughness; methods used in listening tests to measure roughness; the main outcomes of the listening tests and overviews of the developed roughness models given in the literature. Chapter 3 describes the roughness model. Firstly with its peripheral stage composed of individual parts of the hearing system (outer-/middle ear, cochlea, inner hair cells) adapted from the literature, and secondly the central stage developed within the framework of this thesis. Chapter 4 verifies a model of cochlear mechanics which is employed in the peripheral stage of the roughness model. Listening tests conducted within in the framework of this thesis are described in Chapter 5. Chapter 6 compares the model predictions of roughness with results of listening tests reproduced from the literature or conducted by the author (Chapter 5). Finally, the conclusion of the work is given in Chapter 7 together with an overview of the results and suggestions for the future research.

Chapter 2

Roughness: perception and models – state of the art

2.1 Perception of roughness

Herman von Helmholtz (von Helmholtz, 1877) conducted experiments with two simultaneously played tones presented to the same ear. He observed that when the tones slightly differ in frequency, the ear cannot distinguish them and the bats – fluctuations in loudness – are perceived. When the frequency difference between the tones is increased, the number of beats per second also increases. If it is higher than approximately 20 per second, the tones evoke a rattling, harsh and buzzy sound sensation which von Helmholtz described by the term "roughness". The amount of roughness increases with increasing frequency difference between the tones up to a certain point. If the frequency difference is further increased, the amount of roughness starts to decrease. It finally disappears for fluctuation rates higher than about 300 Hz. The similar bandpass characteristic of the perceived roughness was observed with sinusoidally amplitude-modulated (SAM) tones or noises and with sinusoidally frequency-modulated (SFM) tones (Fastl & Zwicker, 2007). Researchers proposed that the cause of the roughness perception maybe in the limited frequency resolution of the ear: the spectral components close in frequency are not resolved by the ear which causes fluctuations in the neural signal and, in turn, the sensation of roughness (von Helmholtz, 1877; Mathes & Miller, 1947; Terhardt, 1974; Vassilakis, 2001).

Researchers have conducted a number of perceptual experiments to measure the per-

ceived amount of roughness and its dependence on physical parameters of synthetic stimuli, for example, modulation depth or modulation frequency in case of amplitude-modulated tones, bandwidth in case of bandpass noises, frequency difference of two tone stimuli causing highest roughness, etc. (Mathes & Miller, 1947; Plomp & Steeneken, 1968; Terhardt, 1974; Kemp, 1982; Aures, 1985; Fastl & Zwicker, 2007; Pressnitzer & McAdams, 1999; Miśkiewicz *et al.*, 2006). Below is given a description of psychophysical methods and main outcomes of the perceptual studies.

2.1.1 Methods of roughness estimation

The methods used in the most of the perceptual studies estimating roughness can be divided into five groups.

Magnitude estimation tasks: The listeners are presented with a stimulus and asked to assign a number according to the amount of perceived roughness. Miśkiewicz *et al.* (2006) used this method to estimate the roughness of pure tone dyads – two simultaneously played pure tones. They presented listeners with the stimuli under test, no reference was used. Kemp (1982) employed a reference in his experiments. He presented listeners with a pair of stimuli – reference and test. The listeners were given a number describing the roughness of the reference, for example 10, and were asked to assign a number corresponding to the roughness of the test stimulus, for example, to assign 20 if the roughness of the test seemed to be twice the roughness of the reference.

Pair wise comparison tasks: Listeners are presented with a pair of stimuli and asked which is more rough. Pressnitzer & McAdams (1999) and Pressnitzer *et al.* (2000) used this method and then applied the Bradley-Terry-Luce (BTL) method (David, 1988) to transform the obtained responses into the interval scaled values. Terhardt (1974) employed different variants of the pair wise comparison tasks to measure roughness. For example, the scaling by estimation of sensation ratios.

Rating scale tasks: Listeners rate the roughness of presented stimuli on a given scale where one end of the scale corresponds to the lowest and the opposite end to the highest perceived roughness. The scales may be continuous or discrete. This method was, for example, used by Kreiman *et al.* (1994); Shrivastav *et al.* (2005) and Patel

et al. (2012) and in the experiments conducted within the framework of this thesis (see Chapter 5). The data obtained by means of this method are on an ordinal or an interval scale (Guski, 1997).

Rank ordering tasks: Listeners are presented with a set of stimuli and asked to order them according to their roughness. This method was, for example, used by Mathes & Miller (1947).

Adjustment tasks: Listeners are asked to adjust the parameters of the pointer which affect its roughness. For example, the task is to adjust the roughness of the pointer to be equal to the roughness of the examined stimulus. Amplitude-modulated (AM) tones are often used as pointers. Listeners may adjust the roughness of AM tones by adjusting its modulation depth (Fastl & Zwicker, 2007). This method or its slight alterations were, for example, employed by Aures (1984, 1985) and Patel *et al.* (2012).

2.1.2 Main outcomes of psychophyical studies

Unit of roughness: Roughness as a ratio scaled entity was given a unit called **asper**. Roughness of 1 asper is defined as the roughness of 100% amplitude-modulated (AM) tone with a frequency of 1 kHz, a sound pressure level (SPL) of 60 dB and a modulation frequency of 70 Hz (Fastl & Zwicker, 2007; Daniel & Weber, 1997).

Threshold and just noticeable difference: The threshold value of roughness is 0.07 asper for an AM tone with a level of 60 dB SPL and a modulation frequency of 70 Hz (Fastl & Zwicker, 2007; Daniel & Weber, 1997). Just noticeable difference (JND) of roughness is 17%. The relative variation of the modulation index, m, of AM tones, $\Delta m/m$, causing JND of the roughness is 10% (Fastl & Zwicker, 2007; Daniel & Weber, 1997).

The dependencies of roughness on parameters of acoustic stimuli: Below is given a brief overview of the measured dependencies of roughness on the parameters

of synthetic acoustic stimuli. The dependencies are shown in the figures in Chapter 6 which compares the behaviorally measured roughness with predictions of the roughness model described in Chapter 3.

- **Dependence on the modulation frequency:** The roughness of AM broad band noise measured as a function of the modulation frequency exhibits a bandpass characteristic with a maximum at the modulation frequency of 70 Hz. The similar characteristics were observed for sinusoidally amplitude-modulated (SAM) or sinusoidally frequency-modulated (SFM) tones with a frequency higher than 1 kHz. The maximum of roughness shifts toward low frequencies when the tone frequency is lower than 1 kHz (Fastl & Zwicker, 2007). The dependence of roughness of SAM tones and SFM tones on the modulation frequency is shown in Fig. 6.1 and 6.8, respectively.

- **Dependence on the modulation index:** It was observed that the dependence of roughness of SAM tones on the modulation index (depth), m, is given by a power law $R \sim m^p$. The value of exponent p varies among different perceptual studies: Terhardt (1968) obtained the value of 2 for a 1-kHz AM tone with modulation frequencies 40, 70 and 120 Hz and various levels; Fastl & Zwicker (2007) reported the value of 1.6; Guirao & Garavilla (1976) obtained the values ranging from 1.1 to 2.2; and Vogel (1975) estimated the value of 1.5. For the values of m higher than 1, the roughness do not follow the power law relation. It reaches maximum at $m \approx 1.2$ and then when m is further increased above this value, the roughness slowly decreases (Mathes & Miller, 1947; Terhardt, 1974). The dependence of roughness of SAM tone on the modulation index, m, given by the aforementioned power law relation for the three values of the estimated exponent, p, (1.6, 2, and 1.5) is shown in Fig. 6.2.

- **Dependence on the phase and the shape of the temporal envelope:** Roughness depends also on the relative phase between the individual spectral components of acoustic stimuli (Mathes & Miller, 1947; Pressnitzer & McAdams, 1999) and on the shape of the time envelope of acoustic stimuli (Pressnitzer & McAdams, 1999). The stimuli with abrupt rise and slow decay of the envelope were perceived with more roughness. These dependencies measured by Pressnitzer & McAdams (1999) are shown in Fig. 6.6 and 6.7, respectively.

- **Dependence on the level:** Loudness affects the roughness but the effect is

not strong. Terhardt (1968) measured approximately three fold increase of roughness when the level of AM tone was increased from 40 to 80 dB SPL. The experimental data measured by Terhardt (1974) showed even smaller increase of the roughness for the same level increment from 40 to 80 dB SPL. The data were obtained with SAM tones at a frequency of 1 and 4 kHz and showed that the roughness depends also on the measurement method. The data measured by Terhardt (1968) and Terhardt (1974) are shown in Fig. 6.3.

- **Summation of roughness:** Terhardt (1974) conducted experiments in order to investigate if the roughness sums up for the stimuli exciting different parts of the basilar membrane (BM). In other words, if the frequency difference between the stimuli is more than the critical bandwidth. The roughness of two SAM tones of frequencies which differed by more than the critical bandwidth was observed to be largest when the tones were modulated in phase and it never decreased under the roughness of any of the tones presented alone. It is assumed that the roughness caused by the BM fluctuations in individual critical bands, so called specific roughness, can be added across the critical bands to the resulting roughness of broadband stimuli (Terhardt, 1974; Daniel & Weber, 1997).

2.1.3 Roughness and dissonance

Simultaneously played tones may sound pleasant or unpleasant. An example of such tones is a musical interval of the octave or the fifth. These intervals are called *consonant*. Unpleasant intervals, such as the second or the augmented fourth (tritone) are called *dissonant*. It is referred to as *sensory consonance* and *sensory dissonance*: *sensory consonance* contributes to *musical consonance* which also includes other factors, for example, context and musical style (Oxenham, 2010). A historical review of the theories of *sensory* or *tonal consonance* was given by Plomp & Levelt (1965).

Dissonance was related to roughness by a theory given by von Helmholtz (1877). He observed that many partials of dissonant intervals are close in frequency and thus produce beats. Since the consonant intervals contain a lot of overlapping partials which do not produce beats, he suggested that the beats are the cause of dissonance. The suggestions of von Helmholtz (1877) were later supported and further improved by other researchers. For example, Plomp & Levelt (1965) asked listeners to rate consonance of intervals composed of two pure tones. Similar experiments with intervals composed of two pure tones were conducted by Kameoka & Kuriyagawa (1969).

See Section 2.2.1 below giving more details about the observed data.

The aforementioned studies and, for example, also Vassilakis (2005) showed a relationship between roughness and dissonance. However, despite this evidence, also the theories questioning this relationship exist (Plomp & Levelt, 1965; Oxenham, 2010). McDermott *et al.* (2010) conducted a thorough study with a large number of listeners (>200) in order to investigate the relationship between dissonance and roughness. They asked listeners to rate the pleasantness of nonmusical stimuli which independently varied in the beating and harmonic content. They subtracted the pleasantness ratings for diotically and dichotically presented stimuli to get the beating ratings. When the two pure tones are presented dichotically – separately, one tone to the left and the other one to the right ear – the perceived beats are attenuated. Similarly, they got the harmonicity ratings by subtracting the pleasantness ratings for harmonic and inharmonic stimuli with widely separated spectral components – without beats. Then, they estimated pleasantness for musical sounds including two-tone intervals (dyads) and three-tone chords (triads). The obtained pleasantness ratings correlated with the harmonicity ratings but not with the beating ratings. The results indicate that harmonicity underlies consonance.

Bidelman & Heinz (2011) published another study questioning the relationship between dissonance and roughness (see also Ebeling (2008)). They used an auditory model to analyze dyads and triads. They got a simulated neural signal and calculated the measure of beating (roughness) and pitch salience. The dissonance ratings correlated with the measure of pitch salience of the stimuli, not with the roughness.

2.1.4 Roughness and annoyance

Acoustic stimuli with roughness are often unpleasant (von Helmholtz, 1877; Terhardt, 1974; Daniel & Weber, 1997). Roughness together with loudness, sharpness and fluctuation strength contributes to psychoacoustic annoyance (PA). PA can quantitatively describe annoyance ratings obtained by psychoacoustic experiments (Fastl & Zwicker, 2007).

2.2 Roughness models

Various roughness models have been developed in the last decades to simulate the psychophysically measured roughness. A brief review was given by Vassilakis (2001) and Leman (2000). Leman (2000) divided the roughness models to *curve-mapping* models and models which take into account the function of the peripheral ear. Some of the roughness models are briefly reviewed below.

2.2.1 Curve-mapping models

The so called *curve-mapping* models estimate roughness from amplitude spectrum of the sound. The models detect frequency components in the spectrum of the sound and map it into a psychoacoustical curve of roughness (Leman, 2000; Vassilakis, 2001).

This approach goes back to von Helmholtz (1877) who observed that dyads of tones evoke the highest roughness when their frequency difference is constant, approximately 33 Hz. This observation was later adjusted by other researchers, for example, Plomp & Levelt (1965) who asked listeners to judge dissonance and by Miśkiewicz *et al.* (2006). Plomp & Levelt (1965) constructed dissonance curves showing the dependence of dissonance on the frequency difference between the tones in dyads. They observed the maximal dissonance for the frequency difference between the tones near one-quarter of the critical bandwidth. Different variants of these models were, for example, described by Kameoka & Kuriyagawa (1969) and Vassilakis (2001). Vassilakis (2001) argued that the models published in the aforementioned studies often predict roughness which disagree with results of listening tests. He improved one of the curve-mapping models and compared its performance with his experimental data on roughness of dyads composed of harmonic complexes (Vassilakis, 2001, 2005).

Potential applicability of the curve-mapping models is limited since they cannot handle stimuli with continuous spectrum, for example, noises (Vassilakis, 2001; Leman, 2000). Leman (2000) argued that the curve-mapping models cannot cover changes in amplitudes of the individual spectral components of the analyzed stimuli and the effect of the phase between spectral components which was observed by Mathes & Miller (1947) and Pressnitzer & McAdams (1999).

2.2.2 Models based on auditory processing

The second group of roughness models takes into account the function of the peripheral ear. Models described by Aures (1985); Daniel & Weber (1997); Leman (2000); Sethares (2005); Fastl & Zwicker (2007); Wang *et al.* (2013) represent this group.

Fastl & Zwicker (2007) estimated the amount of amplitude modulation in the auditory system by means of the temporal masking patterns. The temporal masking patterns are masking thresholds caused by temporally fluctuating maskers, for example, amplitude-modulated tones or noises. The thresholds are measured with test stimuli shorter than the period of the masker fluctuations. The masking period patterns are plotted as a function of the time position of the test stimuli within the masker. The difference between the maxima and minima of the temporal masking patterns reflects the temporal resolution of the auditory system. Fastl & Zwicker (2007) used the difference, ΔL, denoted temporal masking depth together with the modulation frequency, f_{mod}, to approximate roughness by this equation

$$R \sim f_{mod} \Delta L. \tag{2.1}$$

This equation was later elaborated to get more precise calculations for the frequency region across the entire basilar membrane (BM). For details see Fastl & Zwicker (2007).

Daniel & Weber (1997) improved a roughness model designed by Aures (1985). The model estimates roughness from the signal filtered by the Bark scale critical band filterbank. The signal in each critical band is then filtered by bandpass filters in order to account for the experimentally measured bandpass characteristics of roughness when it is shown as a function of the modulation frequency. The roughness is predicted from the modulation depth of the filtered signal and crosscorrelation coefficients between the filtered signals (Daniel & Weber, 1997). A very good agreement between the model predictions and subjective data was shown for SAM and SFM tones and also for unmodulated bandpass noises (Daniel & Weber, 1997). The Daniel and Weber roughness model was implemented into the sound analyses software PsySound3 (PsySound3, 2008). Wang *et al.* (2013) further slightly adjusted the Daniel and Weber roughness model and used it to predict the roughness of interior vehicle noise.

Leman (2000) introduced the *synchronization index* (SI) model. The SI model employs a model of the peripheral ear. It transforms the input acoustic stimulus into the

simulated neural signal in auditory nerve (AN) fibers. The simulated neural signal is then filtered by bandpass filters to account for the bandpass dependence of the roughness on the modulation frequency. A degree of phase locking of the filtered simulated neural signal to a particular frequency is then calculated in the spectral domain as a ratio between the short-term spectra and the direct current (DC) component of the whole signal. Leman (2000) developed two versions of the SI roughness model: one calculating the ratio between the degree of phase locking and the DC component of the whole signal for each model channel separately; and the other calculating the ratio for the sum of the degrees of phase locking over the all model channels. Leman (2000) used the SI roughness model to predict the dependence of roughness of SAM tones on the modulation frequency and roughness of intervals composed of two harmonic complex tones. The SI roughness model was implemented into the IPEM toolbox (IPEM, 2003). Wang (2009) used the SI roughness model to predict the roughness of vehicle noise.

Kohlrausch *et al.* (2005) used both, the Daniel and Weber and SI roughness models to predict the roughness (reproduced from Pressnitzer & McAdams (1999)) of stimuli with asymmetrical temporal envelopes. They altered the peripheral stages (auditory models) of the roughness algorithms by adding the inner hair cell (IHC) and auditory nerve (AN) synapse models simulating the adaptation of the neural signal in AN fibers. The IHC models emphasize onset of the stimuli. This adjustment made the Daniel and Weber roughness model sensitive to the shape of the waveform envelope. However, Kohlrausch *et al.* (2005) argued that the reached results were not satisfying.

2.3 Summary

Researchers have conducted a number of listening tests to measure the roughness sensation. They used various psychoacoustic measurement methods; gave the roughness a unit called asper; measured just noticeable difference of the roughness; and showed the dependency of roughness on parameters of acoustic stimuli, for example, on the modulation frequency of sinusoidally amplitude-modulated (SAM) tones (Fastl & Zwicker, 2007).

Results of the listening tests were used to design roughness models. Some of the roughness models calculates spectral components of the analyzed stimuli and then predict roughness from the frequency difference between the spectral components. These models are called curve mapping models (Leman, 2000; Vassilakis, 2001).

Some of the other roughness models take into account the function of the peripheral ear (Leman, 2000). Despite the vast number of the designed roughness models, non of them have been nationally or internationally standardized yet (Wang *et al.*, 2013).

Chapter 3

Roughness Model

This Chapter describes a roughness model designed by the author within the framework of this thesis. The model have been developed during the last two years and its previous versions have been published in Vencovský (2014a,b). The version described in this thesis will be sent for publication Vencovský (2015b). The roughness model is composed of two successive stages: a peripheral stage and a central stage (see Fig. 3.1). The peripheral stage simulates the individual parts of the peripheral ear (outer-, middle-, and inner-ear). Algorithms simulating the individual parts of the peripheral ear were adapted from different studies and composed into one model. The central stage processes the signal at the output of the peripheral stage. The stage was designed by the author.

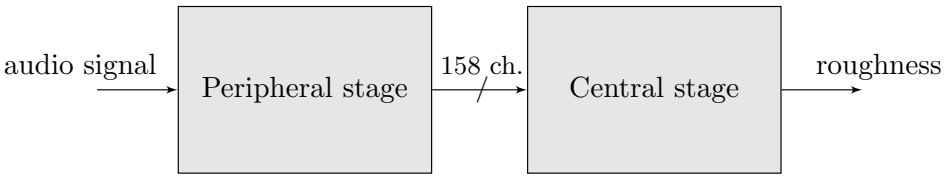

Figure 3.1: Roughness model diagram.

3.1 Peripheral stage

The peripheral stage is an auditory model simulating the function of the outer- and middle-ear, cochlea, inner hair cells (IHCs) and auditory nerve (AN) synapse. Fig. 3.2 shows a diagram of the peripheral stage. The input signal to the peripheral stage is the

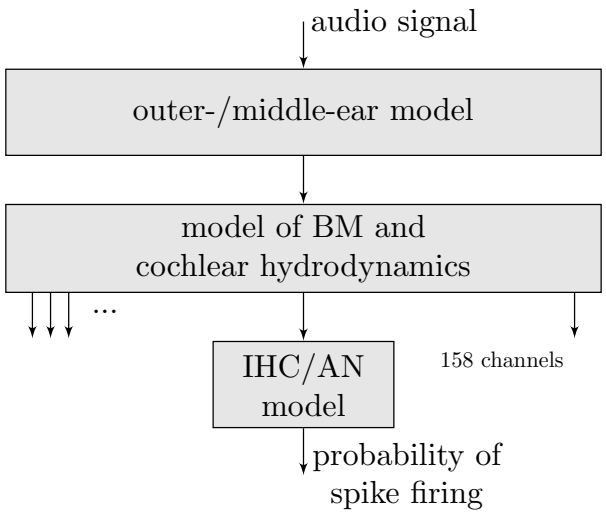

Figure 3.2: Peripheral stage diagram.

acoustic pressure at the entrance of the outer-ear; the output signal is the probability of neural discharge in the auditory nerve (AN) fibers.

3.1.1 Outer- and middle-ear model

Various one dimensional models have been developed to simulate the measured transfer characteristics of the auditory canal (outer-ear) (e.g. Hammershøi & Møller (1996); Hudde & Engel (1998a,b,c); Rund (2004)), and the middle-ear (e.g. Kringlebotn (1988); Goode *et al.* (1994); Hudde & Engel (1998a,b,c); O'Connor & Puria (2008)). One of the models simulating the transfer function of the outer- and middle-ear was developed in the Essex Hearing Research Laboratory at the University of Essex. The model was implemented into the system called Matlab Auditory Periphery (MAP) which is free to download from their website (MAP, 2014). Since this model is easy to implement and was used in the models of the peripheral ear employed in various studies (e.g. Meddis (2006)), it was incorporated into the peripheral stage of the roughness model.

The outer-ear model simulates the resonances observed in the transfer function of the outer-ear (e.g. Pickles (2008)) by two parallel 1-st order Butterworth bandpass filters. The gain, low and high cutoff frequency of one of the filters are 10 dB, 1 kHz and 4 kHz, respectively. The respective parameters of the second filter are 25 dB, 2.5 kHz and 7 kHz. The model filters the input acoustic wave by the bandpass filters. The filtered signal

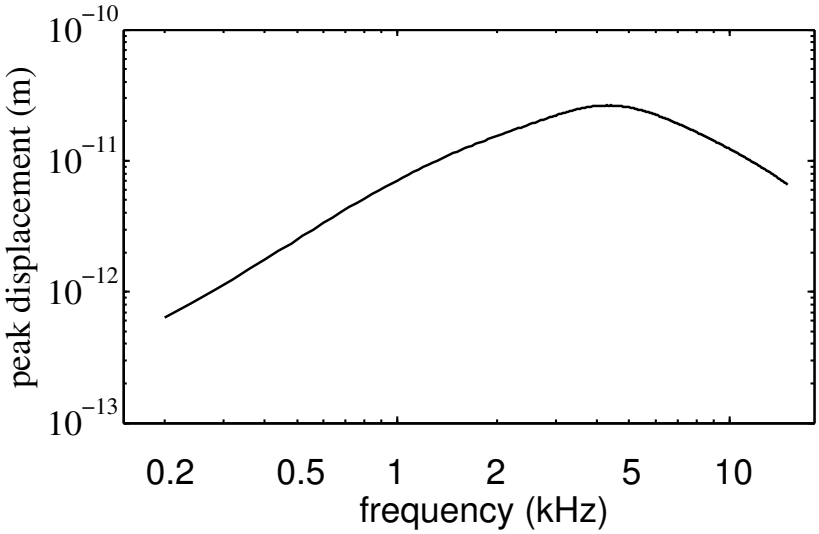

Figure 3.3: Amplitude transfer function of the outer- and middle-ear model. The function is shown as the simulated stapes peak displacement measured using tone bursts of various frequency and a level of 0 dB SPL.

at the output of the parallel bandpass filters is then added to the input acoustic wave. Such created signal is the output signal of the outer-ear model. The signal represents the acoustic pressure at the tympanic membrane. It is mentioned in the documentation of the MAP system (version MAP1.14) that the outer-ear model should receive more attention and its parameters maybe updated in the later versions.

The middle-ear model implemented in the MAP system was developed to simulate the human middle-ear transfer function measured on living species (Huber *et al.*, 2001). The model allows to simulate the acoustic reflex which attenuates the middle-ear transfer function for frequencies below 1 kHz and levels above 80 dB SPL (Pickles, 2008). Since the level of stimuli used in this thesis was not higher than 80 dB SPL, the middle-ear reflex was not incorporated into the middle-ear model employed in the peripheral stage of the roughness model.

The middle-ear model is composed of two successive filters: a 1st-order Butterworth low pass filter with a cutoff frequency of 50 Hz which simulates the transformation of the acoustic pressure to the displacement of the tympanic membrane, and a 1st-order Butterworth high pass filter with a cutoff frequency of 1 kHz which attenuates the displacement of the tympanic membrane at low frequencies. The filtered signal is then multiplied by a scalar of 45×10^{-9} transforming it into the stapes displacement in meters.

Fig. 3.3 shows the amplitude transfer function of the outer- and middle-ear model. The

transfer function was measured by tone bursts with a level of 0 dB SPL. It represents the peak value of the model output signal (stapes peak displacement).

3.1.2 Model of cochlear mechanics

The second part of the peripheral stage is a model simulating the basilar membrane (BM) response and cochlear hydrodynamics. Since this part of the peripheral stage models the limited frequency resolution of the peripheral ear, it is very important for roughness prediction. This Section thus first shortly reviews cochlear models and then describes a model designed by Mammano & Nobili (1993); Nobili & Mammano (1996); Nobili *et al.* (2003). This model was chosen for the peripheral stage of the roughness model. As shows Chapter 4, the cochlear model can account for various physiological and psychophysical phenomena which limits many of the cochlear models.

The author of this thesis adjusted parameters affecting frequency selectivity of the cochlear model. This was done in order to predict psychophysical masking data measured with harmonic complex maskers (see Section 4.2.4 in Chapter 4). The model with adjusted parameters is in this thesis called "the Nobili *et al.* cochlear model". Chapter 4 verifies the ability of the Nobili *et al.* cochlear to simulate physiological and psychophysical data: isointensity and impulse responses, input/output (I/O) functions, and masking thresholds for pure tone and harmonic complex tone maskers.

Researches have developed various types of cochlear models (for a review see Sieroka *et al.* (2006)). The models can be divided into two different groups (Epp *et al.*, 2010): models simulating the function of the cochlea – in this thesis called (according to Epp *et al.* (2010)) "functional" models – which are generally composed of filters; and models simulating the function and also the physiology of the cochlea – in this thesis called (according to Epp *et al.* (2010)) "physical" models.

Functional models: The models made of filters – usually composed to filterbanks simulating certain section of the BM along its length – are widely used in different applications (e.g. Huber & Kollmeier (2006); Harlander *et al.* (2014)). The models are easy to implement and computationally inexpensive (for a review see Lyon *et al.* (2010)). The functional models which lack a connection between the individual frequency channels cannot be used to investigate otoacoustic emissions. This limits the applicability of most of the functional models (Epp *et al.*, 2010).

- **Roex filter:** Results of listening tests led Patterson & Nimmo-Smith (1980) to suggest that the shape of the auditory filter can be approximated by a pair of exponential functions rounded at their top and bottom. This function is called rounded-exponential or "roex" function. Combination of roex filters of different parameters led to better approximation of the psychophysically measured shape of the magnitude characteristics of the human auditory filters (Patterson *et al.*, 1982; Unoki *et al.*, 2006). Since no time-domain implementation of the roex filter exists, they are not used in practical applications (Lyon *et al.*, 2010).

- **Gammatone filter:** Gammatone filters were designed to model the impulse response measured in the mammalian cochlea (de Boer, 1975; Aertsen & Johannesma, 1980; Patterson *et al.*, 1992). The envelope of such impulse responses can be approximated by the gamma distribution function. The transfer function of the gammatone filter is level-independent which disagree with the observations conducted in the mammalian cochlea. The observations show the highly nonlinear BM responses and the level-dependent shape of the cochlear filters (Robles & Ruggero, 2001). Some researchers have used gammatone filters to construct more complex nonlinear filters, for example, the dual resonance nonlinear (DRNL) filter (Meddis *et al.*, 2001) and the triple-path nonlinear (TRNL) filter (Lopez-Najera *et al.*, 2007). They have done it in order to simulate physiological observations. The DRNL filters were used as the peripheral stage of an auditory model which was proved to successfully simulate a variety of psychophysically observed phenomena (Jepsen *et al.*, 2008). Zhang *et al.* (2001) constructed a model simulating physiological observations; the model is composed of the gammatone filters with a time-varying gain and bandwidth (see also Zilany & Bruce (2006); Zilany *et al.* (2009)).

- **Gammachirp filter:** Irino & Patterson (1997) designed a gammachirp filter in order to address discrepancies between the physiology of the cochlea and gammatone filters. Irino & Patterson (2001) then adjusted the gammachirp filter to better represent physiological measurements. The filter was called the compressive gammachirp filter and beside physiological data, it accounts for level-dependent masking data (Patterson *et al.*, 2003) and the phase effects of harmonic complex tone maskers (Nishimura, 2005).

- **Other filterbank models:** Beside the roex, gammatone, and gammachirp filters, other types of filters have been designed: Jenison *et al.* (1991) developed a filter which takes into account level-dependent characteristics of cochlear filters and used

it to process speech stimuli, Pfeiffer (1970) designed the bandpass nonlinear (BPNL) filter, and Goldstein (1995) designed the multiple bandpass nonlinear (MBPNL) filter.

Physical models: The physical models simulate the function of the cochlea in more detail and can model a generation of otoacoustic emissions (Epp *et al.*, 2010). The BM can be modeled as a transmission line. Two main approximations exist: a longwave approximation (Peterson & Bogert, 1950; Zwislocki, 1950) assuming that the wavelength of the traveling wave on the BM is long in comparison to the height of the cochlear duct; and a shortwave approximation (Ranke, 1950) assuming that the wavelength is shorter than the height of the cochlear duct. Time-domain longwave transmission line models designed using the method of electroacoustical analogy were described, for example, by Schroeder (1978); Strube (1985); Giguére & Woodland (1994); Epp *et al.* (2010); Verhulst *et al.* (2012). Two and three dimensional models have been developed in order to model the cochlear mechanics more precisely. Siebert (1974) used a shortwave approximation and developed a two dimensional model. Allen (1977) used the Green's-function method to simulate the fluid coupling between different parts of the vibrating BM. He derived an integrodifferential equation and solved it numerically. Mammano & Nobili (1993); Nobili & Mammano (1996) and Nobili *et al.* (2003) used also the Green's function method and realized a cochlear model with realistic dimensions of the cochlea. Shera *et al.* (2004) showed that the model of Nobili *et al.* represents a different mathematical description of the same phenomenon as the transmission line models – both approaches are thus equivalent. Three dimensional models are very complex and it is usually necessary to do simplifications in order to obtain an analytical equation (Holmes, 1980; Steele & Taber, 1981; de Boer, 1995; Elliott *et al.*, 2011).

Nobili *et al.* cochlear model used in the peripheral stage

The peripheral stage of the roughness model employs a model of the BM response and cochlear hydrodynamics designed by Mammano and Nobili (Mammano & Nobili, 1993; Nobili & Mammano, 1996; Nobili *et al.*, 2003). The model was designed with the realistic dimensions and parameters of the guinea-pig cochlea (Mammano & Nobili, 1993; Nobili & Mammano, 1996) and then later modified to human

parameters (NOBILI *et al.*, 2003).

The BM is modeled as an array of $N = 300$ oscillators with a displacement, ξ_i, of the i-th oscillator given by

$$\begin{aligned} &m_i\ddot{\xi}_i(t) + h_i\dot{\xi}_i(t) + s_i[2\dot{\xi}_i(t) - \dot{\xi}_{i-1}(t) - \dot{\xi}_{i+1}(t)] + k_i\xi_i(t) \\ &= f_{H_i}(t) + f_{\mathrm{OHC}_i}[\eta_i(t)], \end{aligned} \tag{3.1}$$

where m_i, h_i, s_i and k_i are the mass, positional viscosity, sharing viscosity and stiffness of the BM, respectively. The oscillators are driven by the force $f_{H_i}(t)$ given by

$$f_{H_i}(t) = -G_{S_i}a_{S_i}(t) - \sum_{j=1}^{N} G_i^j\ddot{\xi}_j(t), \tag{3.2}$$

where $a_{S_i}(t)$ is the acceleration of the stapes and $\ddot{\xi}_i(t)$ is the acceleration of the individual oscillators. G_{S_i} and G_i^j are coefficients transforming the accelerations into the corresponding forces. The second force term, $f_{\mathrm{OHC}_i}[\eta_i(t)]$, is the nonlinear sigmoidal function transforming the OHC stereocilia displacement, $\eta_i(t)$, into the OHC force. The stereocilia displacement is given by

$$\bar{m}_i\ddot{\eta}_i(t) + \bar{h}_i\dot{\eta}_i(t) + \bar{k}_i\eta_i(t) = -g_i\ddot{\xi}_i(t), \tag{3.3}$$

where $\bar{m}_i$, $\bar{h}_i$ and $\bar{k}_i$ are the mass, viscosity, and stiffness, respectively, of a subsystem composed of the tectorical membrane (TM) and the reticular lamina (RL); $\eta_i(t)$ is the displacement of OHC stereocilia; and g_i is the gain factor coupling the BM and RL motion.

The author of this thesis multiplied the positional viscosity, h_i, by a constant of 0.31 in order to increase the frequency selectivity of the cochlear model. The frequency selectivity was increased in order to fit the psychophysical masking data obtained using harmonic complex tones as maskers (see Chapter 4, Section 4.2.4). Table 4.1 in Chapter 4 shows the equivalent rectangular bandwidths (ERBs) of the cochlear filters estimated in four discrete points along the length of the BM. Chapter 4 also shows the isointensity responses of the simulated cochlear filters (Fig. 4.3), input/output (I/O) functions (Fig. 4.6), impulse responses (Fig. 4.7), and compares the masking thresholds predicted by the model with results of listening tests (Fig. 4.8, Fig. 4.9 and Fig. 4.10, Fig. 4.11 and Fig. 4.12).

The model is composed of 300 oscillators with characteristic frequencies (CFs) ranging

between 20 Hz and 17 kHz. Only the signals in 158 channels of the 300 channels are fed into the subsequent IHC model. The author of this thesis picked the channels such that the frequency resolution of the model is 4 channels per the critical band with equivalent rectangular bandwidth (ERB) given by

$$B_{\mathrm{ERB}} = 24.7(4.37 f_{\mathrm{CF}} + 1), \tag{3.4}$$

where f_{CF} is the auditory filter CF in kHz (Moore & Glasberg, 1996).

The equations were implemented in the time domain by the implicit Euler method with 400 kHz sampling frequency. The implementation was done in Matlab (The MathWorks, Inc., Natick, MA) environment.

3.1.3 Model of inner hair cell and auditory nerve synapse

Organ of Corti located on the BM contains inner hair cells (IHCs) serving as receptors. Their name comes from the small hairs called stereocilia placed on their top. The vibrations of the BM are transferred into the vibrations of the stereocilia. This opens ion channels inside the stereocilia. Ions entering the IHC cause depolarization of the membrane potential. This leads to the opening of ion channels in the IHC membrane which finally results in the increased neural activity in the adjacent auditory nerve (AN) fibers (Pickles, 2008). Below is given a brief overview of the approaches used to model the IHC and AN synapse.

Models of inner hair cells: The stereocilia deflection and depolarization of the IHC membrane potential can be modeled by signal-processing models and biophysical models (for a review see Meddis & Lopez-Poveda (2010)). The signal-processing models often consist of a half-wave rectifier followed by a lowpass filter (e.g. Robert & Eriksson (1999); Dau *et al.* (1996); Zhang *et al.* (2001)). The models are easy to implement and require low number of parameters. However, they cannot account for all of the physiological phenomena observed in the IHC (Meddis & Lopez-Poveda, 2010). The biophysical models are composed of the algorithms simulating the individual processes in the IHC: bending of the stereocilia, opening of the ion channels, influx of the ions into the cell body, depolarization of the membrane potential, etc. (Shamma *et al.*, 1986; Zeddies & Siegel, 2004; Lopez-Poveda & Eustaquio-Martín, 2006).

Models of AN synapse: Depolarization of the IHC membrane potential may, through the subsequent processes in the IHC, lead to the spike firing into the AN fiber (Pickles, 2008). The neural activity of the AN fiber shows an adaptive characteristic: it is maximal after the signal onset and then slowly reduced. Püschel (1988) designed a signal-processing model composed of a cascade of feedback loops. This model was implemented into the auditory models (e.g. Dau *et al.* (1996); Jepsen *et al.* (2008)) and used in the practical applications (e.g. Harlander *et al.* (2014); Huber & Kollmeier (2006)). The biophysical models simulate these processes in two steps: calcium control of transmitter release (Sumner *et al.*, 2002; Meddis, 2006) and transmitter release (Meddis, 1986; Westerman & Smith, 1988; Meddis, 2006).

IHC and AN synapse model used in the peripheral stage

The peripheral stage of the roughness model employs a biophysical IHC model consisting of a set of successive algorithms described by Shamma *et al.* (1986); Sumner *et al.* (2002); Meddis (1986, 2006). These algorithms were implemented into the Matlab Auditory Periphery (MAP, 2014). The author of this thesis adapted the parameters of the implemented algorithms. The adapted parameters are given in Table A.1, A.2 and A.3 in the Appendix. The author also experimented with signal processing models of the IHC and AN synapse but the designed central stage worked best with the chosen algorithms described below.

The first of the algorithms transforms the BM displacement, $w(t)$, (output of the cochlear model) into the displacement of the IHC stereocilia, $u(t)$, (Shamma *et al.*, 1986). It is given by

$$\tau_c \frac{du(t)}{dt} + u(t) = \tau_c C_{\text{cilia}} \frac{dw(t)}{dt}, \tag{3.5}$$

where τ_c is the time constant and C_{cilia} is the gain factor. The algorithm is a highpass filter: stereocilia displacement thus at frequencies below a cutoff frequency of the highpass filter move in phase with the BM velocity and at frequencies higher than a cutoff frequency of the high pass filter move in phase with the BM displacement (Shamma *et al.*, 1986; Sumner *et al.*, 2002).

The bending of stereocilia opens the ion channels inside their body. This process is unilateral – bending of stereocilia toward the opposite direction closes the ion channels.

The opening of the ion channels affects its apical conductance, $G(u)$, which is modeled by a Boltzmann function given by Sumner *et al.* (2002)

$$G(u) = G_{\text{cilia}}^{\max}\left[1+\exp\left(-\frac{u(t)-u_0}{s_0}\right)\times\left[1+\exp\left(-\frac{u(t)-u_1}{s_1}\right)\right]\right]^{-1} + G_a, \tag{3.6}$$

where $G_{\text{cilia}}^{\max}$ is the maximal conductance with all the channels open; s_0, u_0, s_1 and u_1 are constants determining the shape of the nonlinear Boltzmann function; and G_a is the passive conductance. The passive conductance is given by

$$G_a = G_{\text{cilia}}^{\max}\left[1+\exp\left(\frac{u_0}{s_0}\right)\times\left[1+\exp\left(\frac{u_1}{s_1}\right)\right]\right]^{-1} + G_0, \tag{3.7}$$

where G_0 is the resting conductance.

Ions entering the IHC depolarize the membrane potential which is modeled by an analog circuit given by Shamma *et al.* (1986); Sumner *et al.* (2002)

$$C_m\frac{dV(t)}{dt} + G(u)(V(t)-E_t) + G_k(V(t)-E_k') = 0, \tag{3.8}$$

where C_m is the cell capacitance, $V(t)$ is the membrane potential, G_k is the voltage-invariant basolateral membrane conductance, E_t is the endocochlear potential, and E_k' is the reversal potential of the basal current corrected for the resistance of the supporting cells. The endocochlear potential is given by $E_k' = E_k + E_t R_p/(R_t+R_p)$, where E_k, E_t, R_p and R_t are the parameters (see Table A.1 in the Appendix).

The following set of algorithms adapted from Sumner *et al.* (2002); Meddis (2006) models the calcium control of the transmitter release. The calcium ions controlling the transmitter release enter the cell through the ion channels which are opened when the cell is depolarized. The calcium current, I_{Ca}, is given by

$$I_{\text{Ca}}(t) = G_{\text{Ca}}^{\max} m_{I_{\text{Ca}}}^3(t)(V(t)-E_{\text{Ca}}), \tag{3.9}$$

where $G_{\text{Ca}}^{\max}$ is the maximal calcium conductance with all the ion channels open and $m_{I_{\text{Ca}}}$ is the fraction of open calcium channels. Steady state value of the fraction of open ion channels $m_{I_{\text{Ca}},\infty}$ is simulated by a Boltzmann function given by

$$m_{I_{\text{Ca}},\infty} = [1+\beta_{\text{Ca}}^{-1}\exp(-\gamma_{\text{Ca}}V(t))]^{-1}, \tag{3.10}$$

where β_{Ca} and γ_{Ca} are constants. Fraction of open ion channels $m_{I_{\mathrm{Ca}}}$ is given by

$$\tau_{I_{\mathrm{Ca}}} \frac{dm_{I_{\mathrm{Ca}}}(t)}{dt} + m_{I_{\mathrm{Ca}}}(t) = m_{I_{\mathrm{Ca}},\infty}, \tag{3.11}$$

where $\tau_{I_{\mathrm{Ca}}}$ is a time constant. The concentration of calcium ions in the cell $[\mathrm{Ca}^{2+}]$ is calculated from the calcium current, I_{Ca}, by

$$\tau_{[\mathrm{Ca}]} \frac{d[\mathrm{Ca}^{2+}](t)}{dt} + [\mathrm{Ca}^{2+}](t) = I_{\mathrm{Ca}}(t), \tag{3.12}$$

where $\tau_{[\mathrm{Ca}]}$ is a time constant. The probability the release of neurotransmitter into the synaptic cleft, $k(t)$, is controlled by the calcium concentration, $[\mathrm{Ca}^{2+}]$, as is given by

$$k(t) = z[\mathrm{Ca}^{2+}]^3(t), \tag{3.13}$$

where z is the scalar constant converting calcium concentration $[\mathrm{Ca}^{2+}]$ into the release rate. The relation was adapted from Meddis (2006).

Meddis (1986) described the transmitter release and its circulation by means of the three equations:

$$\frac{dq(t)}{dt} = y(1 - q(t)) + xw(t) - k(t)q(t), \tag{3.14}$$

$$\frac{dc(t)}{dt} = k(t)q(t) - lc(t) - rc(t), \tag{3.15}$$

$$\frac{dw(t)}{dt} = rc(t) - xw(t). \tag{3.16}$$

The term q is the immediate store of neurotransmitter which is released at rate $k(t)$ into cleft c. The release rate is mediated by calcium concentration $[\mathrm{Ca}^{2+}]$. Some of the transmitter in the cleft is lost at rate l. The remaining transmitter is taken back into reprocessing store w at rate r and then into immediate store q at rate x.

The output of the peripheral stage is the content of the cleft, c, which is proportional to the probability of the spike firing into the AN fiber. The algorithms given by Sumner *et al.* (2002); Meddis (2006) use different sets of parameters to simulate the AN fibers of different spontaneous neural activity: high spontaneous rate (HSR), medium spontaneous rate (MSR), and low spontaneous rate (LSR). The author of this thesis adjusted the parameters of the IHC and AN model (see Tables A.1, A.2 and A.3 in the Appendix) in order to employ only one type of AN fibers which would cover as wide dynamic range as possible. Fig. 3.4 shows the input/output (I/O) functions of the peripheral stage measured using tone bursts with a frequency of 0.25, 1 and 4 kHz. The

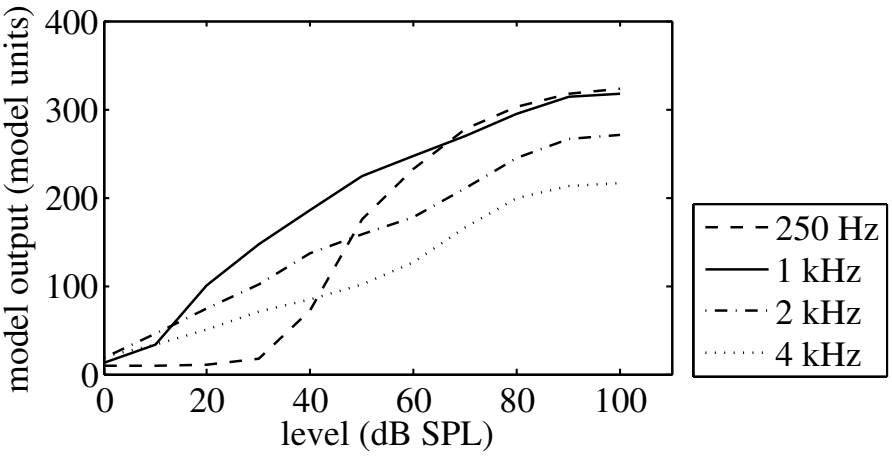

Figure 3.4: Input/output (I/O) functions of the peripheral stage measured using tone bursts with a frequency of 0.25, 1 and 4 kHz. The functions were measured in the channels of the peripheral stage with CF equal to the tone burst frequency.

responses were measured in discrete channels of the peripheral stage with CF equal to the frequency of the tone bursts.

3.2 Central stage

The central stage processes the output signal of the peripheral stage and gives the prediction of roughness of the processed stimuli. Fig. 3.5 shows a diagram of the central stage; it consists of a set of algorithms which were designed by the author within the framework of this thesis. The algorithms were designed and its parameters set in order to fit the subjective data of roughness (reproduced from Fig. 11.2 in Fastl & Zwicker (2007)) of 100% sinusoidally amplitude-modulated (SAM) tones. The aim was also to make the central stage sensitive to the shape of the temporal envelope of the signal at the output of the peripheral stage as was advised by Pressnitzer & McAdams (1999). This should allow to predict the effect of the phase of the individual spectral components on roughness and also the effect of the shape of the temporal waveform envelopes on roughness (Pressnitzer & McAdams, 1999).

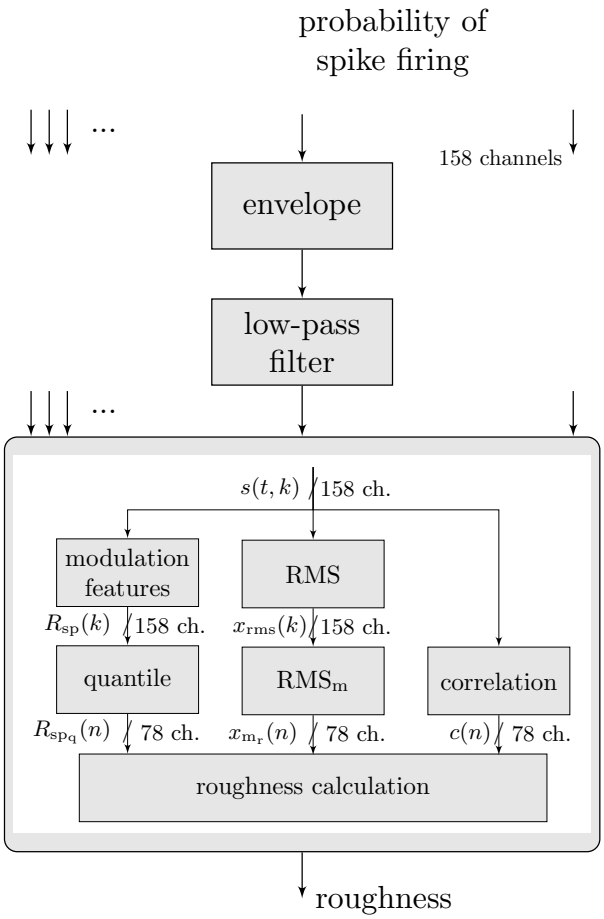

Figure 3.5: Central stage diagram.

3.2.1 Envelope and lowpass filter

The first two blocks of the central stage ("envelope" and "lowpass filter") process the signal in each output channel of the peripheral stage. The output of these two blocks is denoted the filtered envelope, $s(t, k)$, where t is the time and k is the channel number.

Fig. 3.6 (top panel, solid line) shows the signal in one channel of the peripheral stage obtained in response to a 100% SAM tone with a frequency of 1 kHz, a level of 60 dB SPL and a modulation frequency of 70 Hz. The block "envelope" detects peaks of each half wave of the signal fine structure and interpolates them by a cubic spline function. The top panel in Fig. 3.6 (dashed line) shows the calculated envelope.

The calculated envelope is then processed by a 1-st order Butterworth lowpass filter with a cutoff frequency of 70 Hz. The filter decreases the amplitude of temporal fluctuations

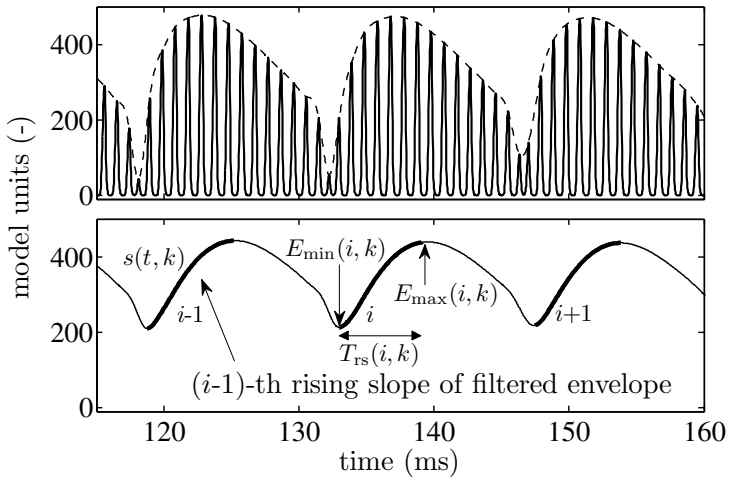

Figure 3.6: Top panel: Signal at one output channel (CF = 1 kHz) of the peripheral stage. The signal was obtained in response to a 100% sinusoidally amplitude-modulated (SAM) tone with a frequency of 1 kHz, a level of 60 dB SPL and a modulation frequency of 70 Hz. The dashed line represents the calculated signal envelope. Bottom panel: Envelope (dashed line in the top panel) smoothed by a 1st-order Butterworth filter with a cutoff frequency of 70 Hz. The tick lines represent the rising parts of the smoothed envelope.

with a frequency higher than about 70 Hz. In other words, it models the decrease of the roughness of stimuli with temporal fluctuations of a frequency higher than about 70 Hz (Fastl & Zwicker, 2007). The bottom panel of Fig. 3.6 shows the filtered envelope, $s(t,k)$. It was obtained by filtering the estimated envelope shown in the top panel of the same figure (dashed line).

3.2.2 Processing of filtered envelope

The filtered envelope, $s(t,k)$, is processed in successive time frames with duration $T_{\rm fr} = 30$ ms. The envelope in each time frame is fed into three parallel branches (see Fig. 3.5): the first branch detects features describing the temporal fluctuations of the filtered envelope, $s(t,k)$, (for one time frame); the second branch calculates root mean square (RMS) values of $s(t,k)$ (for one time frame); and the third branch calculates crosscorrelation coefficients between the individual channels of $s(t,k)$ (for one time frame). The parameters calculated by the three branches are then used to predict roughness of analyzed acoustic stimuli (for one time frame). The overall roughness of the analyzed acoustic stimuli is calculated as the median of the predicted roughness for

the individual successive time frames.

1-st branch: The first branch detects features describing the temporal fluctuations of the filtered envelope, $s(t,k)$, (for one time frame). Temporal fluctuations of the envelope of neural signal in AN fibers are believed to be a cause of the roughness sensation (Mathes & Miller, 1947; Terhardt, 1974). Pressnitzer & McAdams (1999) extended these observation and showed that not only the depth of the modulation but also the shape of the modulated envelope should be taken into account in order to predict roughness. Following these observations, the author of this thesis suggests that only the rising parts of the neural signal envelope may contribute to roughness – such that more abrupt rise of the envelope and a difference between the maximal and minimal value of the rising part of the envelope would result to more roughness. The central stage (1-st branch) thus detects the rising parts (slopes) of the filtered envelope, $s(t,k)$, and estimates features describing the duration of the rising part (slope) of the envelope and the amount of the envelope increase (rise).

The bottom panel in Fig. 3.6 (tick solid lines) shows the rising slopes of $s(t,k)$. The letter i denotes the rank of each rising slope within one time frame of $s(t,k)$. The minimum and maximum of i-th rising slope is in Fig. 3.6 denoted as $E_{\min}(i,k)$ and $E_{\max}(i,k)$, respectively. $E_{\min}(i,k)$ and $E_{\max}(i,k)$ and its temporal positions within the time frame are then used to calculated the features describing the temporal fluctuations (modulations) of the filtered envelope, $s(t,k)$, in one time frame: the modulation index (of each rising slope), $M(i,k)$, given by

$$M(i,k) = \frac{E_{\max}(i,k) - E_{\min}(i,k)}{E_{\max}(i,k) + E_{\min}(i,k)}; \tag{3.17}$$

and the duration (of each rising slope), $T_{\mathrm{rs}}(i,k)$, given as a difference between the temporal positions of $E_{\max}(i,k)$ and $E_{\min}(i,k)$.

How can these two features be used to predict roughness is explained by means of the results of listening tests (reproduced from Fastl & Zwicker (2007)) showing the dependence of roughness of 100% SAM tones on the modulation frequency. The dependence exhibits a bandpass characteristic: the perceived roughness is maximal for the modulation frequency of about 70 Hz, for the SAM tone frequency $\geq$1 kHz, and for the modulation frequencies lower than about 70 Hz, for the SAM tone frequency $<$1 kHz (Fastl & Zwicker, 2007) (see also Section 6.1.1, Fig. 6.1). Changes of the modulation frequency affect the duration, $T_{\mathrm{rs}}(i,k)$, and the modulation index, $M(i,k)$,

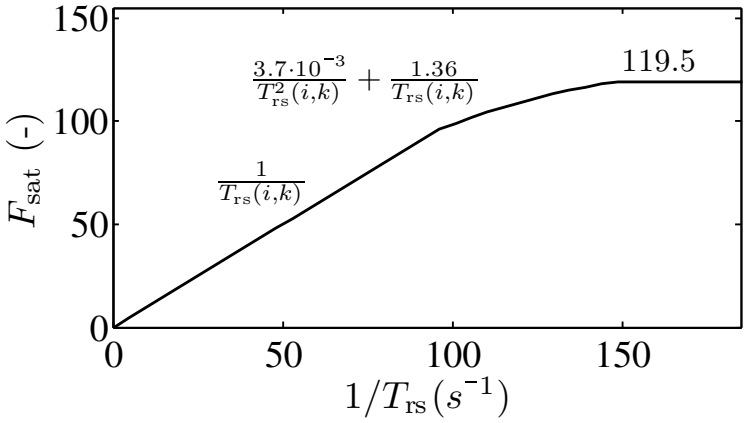

Figure 3.7: Transformation function given by Eq. (3.20). The function transforms the duration, T_{rs}, of the rising part of the filtered envelope of the signal at the output of each channel of the peripheral stage.

of the rising slopes in the following way. As the modulation frequency increases, the value of $T_{\mathrm{rs}}(i,k)$ decreases. Together with this, the lowpass filter in the second block of the central stage ensures that the value of the modulation index, $M(i,k)$, decreases as the modulation frequency increases above 70 Hz. The dependence of $M(i,k)$ on the modulation frequency is affected also by the cochlear model used in the peripheral stage. The function of the cochlear model resembles the function of bandpass filters. Bandwidth of the bandpass filters (cochlear filters) depends on the filter center frequency – increases with increasing center frequency as shows Eq. (3.4) (Moore & Glasberg, 1996). The spectral components of a SAM tone are at frequencies $f_c - f_m$, f_c, and $f_c + f_m$, where f_c is the tone frequency and f_m is the modulation frequency. The bandwidth of the cochlear filters thus determines the maximal value of the modulation frequency, f_m, for which all the spectral components (tone and sideband components) fall into the same cochlear filter. If f_c is low, the sideband components of the SAM tone will fall into the adjacent cochlear filters for lower f_m than if f_c is high. If the spectral components fall into the adjacent cochlear filters, they less interact (beat) together in the cochlear filter – lower values of $M(i,k)$. $M(i,k)$ will thus be affected at lower modulation frequencies than 70 Hz for tone frequencies below 1 kHz.

Features $T_{\mathrm{rs}}(i,k)$ and $M(i,k)$ can be combined together in order to predict the dependence of roughness on the modulation frequency for SAM tones (exhibiting a bandpass characteristic). If the modulation frequency increases, $T_{\mathrm{rs}}(i,k)$ decreases and $M(i,k)$ is approximately constant until the point of maximum roughness. If the modulation frequency is further increased, $M(i,k)$ starts to decrease. The dependence of roughness on the modulation frequency thus could be predicted by calculating the product of both

modulation features[1] (after some adjustments given below by Eq. (3.19) and (3.20)).

The product of both modulation features giving the predicted roughness, $R_{\mathrm{sp}}(k)$, in k-th channel of the central stage is given by[2]

$$R_{\mathrm{sp}}(k) = \max_i \{M_{\mathrm{sat}}(i,k) \cdot F_{\mathrm{sat}}^{1.5}(i,k)\}, \tag{3.18}$$

where $M_{\mathrm{sat}}(i,k)$ is the modulation index of the rising slope saturated at 0.5 as is given by

$$M_{\mathrm{sat}}(i,k) = \begin{cases} M(i,k) & \text{if } M(i,k) \leq 0.5 \\ 0.5 & \text{if } M(i,k) > 0.5, \end{cases} \tag{3.19}$$

and $F_{\mathrm{sat}}(i,k)$ is the parameter calculated from the duration of the rising slopes, $T_{\mathrm{rs}}(i,k)$, by

$$F_{\mathrm{sat}}(i,k) = \begin{cases} \frac{1}{T_{\mathrm{rs}}(i,k)} & \text{if } \frac{1}{T_{\mathrm{rs}}(i,k)} \leq 96s^{-1} \\ 119.5 & \text{if } \frac{1}{T_{\mathrm{rs}}(i,k)} \geq 149s^{-1} \\ \frac{3.7 \cdot 10^{-3}}{T_{\mathrm{rs}}^2(i,k)} + \frac{1.36}{T_{\mathrm{rs}}(i,k)} & \text{else.} \end{cases} \tag{3.20}$$

Eq. (3.19) and Eq. (3.20) transforms features $M(i,k)$ and $T_{\mathrm{rs}}(i,k)$.

The modulation index, $M(i,k)$ is bounded by Eq. (3.19) not to exceed the value of 0.5. The value was found experimentally in order to fit the bandpass characteristics (the dependence of roughness on the modulation frequency for 100% SAM tones, see Fig. 6.1). $M(i,k)$ is mostly in a range between 0 and 0.5 – in the model channels with CF close to the frequency of the spectral components of the SAM tones – and higher than 0.5 – in the adjacent channels. Eq. (3.20) thus affects the calculated roughness, $R_{\mathrm{sp}}(k)$, in the adjacent channels. This, as is more shown below, helps to reach the bandpass characteristic of the dependency of roughness for SAM tones – especially for the tones with a frequency <1 kHz.

The duration of the rising slope, $T_{\mathrm{rs}}(i,k)$, affects mainly the increasing portion of the bandpass characteristic showing the dependence of roughness of a SAM tone on the modulation frequency. Eq. (3.20) calculates the reciprocal of $T_{\mathrm{rs}}(i,k)$ and also shapes and limits its values (see Fig. 3.7). The upper limit ensures that $R_{\mathrm{sp}}(k)$ do not increase for the modulation frequencies above 70 Hz. The limitation was necessary since the decrease of $M(i,k)$ caused by the lowpass filter (the second block) and the cochlear

[1] This resembles the Fastl and Zwicker roughness model (Fastl & Zwicker, 2007) (briefly described also in Chapter 2).

[2] There is a mistake in Vencovský (2014a,b). The exponent 1.5 is not placed above the term corresponding to the duration of the rising slopes of the filtered envelope.

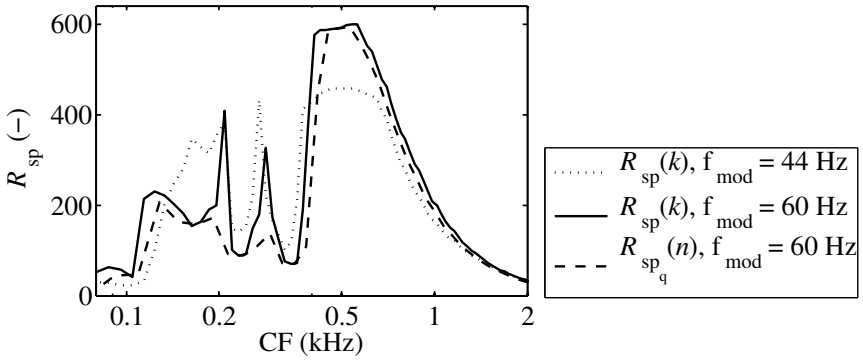

Figure 3.8: Predicted roughness in the individual channels of the central stage (k or n). The processed stimulus was a 100% sinusoidally amplitude-modulated (SAM) tone with a frequency of 250 Hz, a level of 60 dB SPL, and a modulation frequency 50 Hz. The solid line shows values of $R_{\mathrm{sp}}(k)$ given by Eq. (3.18). The dashed line shows values of $R_{\mathrm{sp_q}}(n)$ given by Eq. (3.21)

model was not enough to predict the decreasing portion of the bandpass characteristic (the dependence of roughness on the modulation frequency for SAM tones).

The last block in the 1st-branch calculates quantiles of $R_{\mathrm{sp}}(k)$ over four adjacent channels, k, as is given by

$$R_{\mathrm{sp_q}}(n) = \underset{20\%}{\mathrm{quantile}}\{R_{\mathrm{sp}}(k), R_{\mathrm{sp}}(k+1), R_{\mathrm{sp}}(k+2), R_{\mathrm{sp}}(k+3)\}, \quad \forall k \in 1, 3, 5, ..., 158-3. \tag{3.21}$$

This transforms the values of $R_{\mathrm{sp}}(k)$ in 158 channels into the values of $R_{\mathrm{sp_q}}(n)$ in 78 channels. The result of this transformation is shown in Fig. 3.8: the solid line shows $R_{\mathrm{sp}}(k)$, and the dashed line $R_{\mathrm{sp_q}}(n)$. The values were calculated in one time frame of the processed 100% SAM tone with a frequency of 250 Hz, a level of 60 dB SPL and a modulation frequency of 60 Hz. The dotted line in the same figure shows $R_{\mathrm{sp}}(k)$ for the same SAM tone with a modulation frequency of 44 Hz. This tone is perceived with more roughness than the SAM tone with a modulation frequency of 60 Hz (see Fig. 6.1 below). $R_{\mathrm{sp}}(k)$ at 44-Hz modulation frequency is higher than $R_{\mathrm{sp}}(k)$ at 60-Hz modulation frequency mainly for the channels with CF around the frequency of the spectral components of the SAM tones (in the places of dips in $R_{\mathrm{sp}}(k)$ between about 180 Hz and 350 Hz). Otherwise, $R_{\mathrm{sp}}(k)$ even increases (at frequencies above about 350 Hz). Eq. (3.21) thus helps to predict the dependence of roughness of SAM tones on the modulation frequency – especially for the SAM tones with a frequency <1 kHz. Quantiles are calculated across four adjacent bands with CF laying within the same critical band: as was mentioned above in Section 3.1.2, frequency resolution of

the peripheral stage is 4 channels per critical bandwidth given in ERBs (Moore & Glasberg, 1996).

2-nd branch: The 2-nd branch calculates the RMS values of the filtered envelope, $s(t,k)$ (for one time frame). The RMS values are given by

$$x_{\mathrm{rms}}(k) = \sqrt{\frac{1}{T_{\mathrm{fr}}} \int_{T_{\mathrm{fr}}} s^2(t,k)dt}, \tag{3.22}$$

where T_{fr} is the time frame duration: it is equal to 30 ms. The RMS values are then averaged as is given by

$$\begin{aligned} &x_{\mathrm{m_r}}(n) = \mathrm{mean}\{x_{\mathrm{rms}}(k), x_{\mathrm{rms}}(k+1), x_{\mathrm{rms}}(k+2), x_{\mathrm{rms}}(k+3)\}, \\ &\forall k \in 1,3,5,...,158-3. \end{aligned} \tag{3.23}$$

This processing decreases the number of channels to 78 which is equal to the number of channels of $R_{\mathrm{spq}}(n)$. The channels for which the RMS of the envelope is lower than the specific threshold value are set to 0 as is given by

$$\begin{aligned} x_{\mathrm{m_r}}(n) &= \begin{cases} x_{\mathrm{m_r}}(n) & \text{if } x_{\mathrm{m_r}} \geq 0.1 x_{\max} \\ 0 & \text{else.} \end{cases} \\ x_{\max} &= \max_n \{x_{\mathrm{m_r}}(n)\}. \end{aligned} \tag{3.24}$$

This excludes the channels with low excitation from the process of roughness prediction (see Eq. (3.26) below).

3-rd branch: The 3-rd branch calculates the crosscorrelation coefficients, $c(n)$, between the filtered envelopes, $s(t,k)$ (in one time frame): the coefficients are calculated between the first and the last channel in four adjacent channels of $s(t,k)$ as is given by

$$\begin{aligned} &c(n) = \mathrm{corr}\{s(t,k)s(t,k+3)\} \\ &\forall k \in 1,3,5,...,158-3, \\ &c(n) = \begin{cases} c(n) & \text{if } c(n) \geq 0 \\ 0 & \text{else.} \end{cases} \end{aligned} \tag{3.25}$$

Since the crosscorrelation is low especially for unmodulated noise stimuli, taking them into account during the roughness prediction (Eq. (3.26) below) decreases the predicted

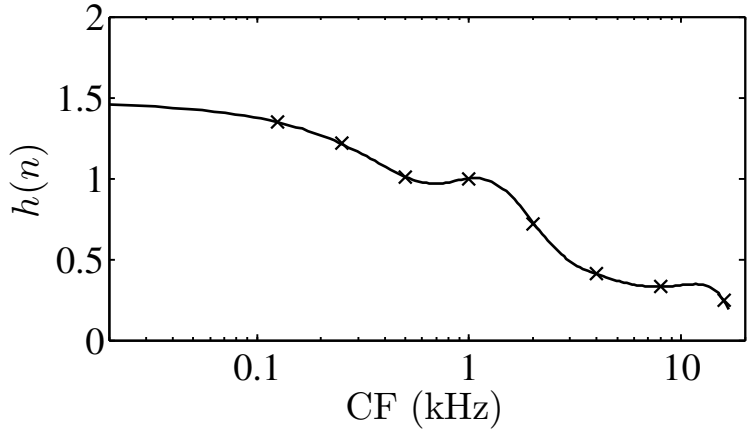

Figure 3.9: Weighting function applied in Eq. (3.26) to predict quantitatively similar data of roughness in aspers as show the experimental subjective data for 100% SAM tones (Fastl & Zwicker, 2007) (Fig. 11.2).

roughness for this type of stimuli. This approach was inspired by the roughness models designed by Aures (1985); Daniel & Weber (1997).

Roughness calculation: The parameters calculated by the three parallel branches are used to predict roughness from the filtered envelope, $s(t,k)$ (for one time frame). The predicted roughness for one time frame, R, is given by

$$R = \sqrt{b} \sum_{n=1}^{78} \frac{h(n) R_{\mathrm{sp_q}}(n) c(n) x_{\mathrm{m_r}}(n)}{\sum_{n=1}^{78} x_{\mathrm{m_r}}(n)}, \tag{3.26}$$

where $R_{\mathrm{sp_q}}(n)$, $x_{\mathrm{m_r}}(n)$ and $c(n)$ are the aforementioned parameters, $h(n)$ is the weighting function and b is the number of channels for which $x_{\mathrm{m_r}}(n) > 0$. Eq. (3.26) emphasizes the predicted roughness in the channels of higher RMS values and excludes the channels of very low RMS values. The term $\sqrt{b}$ and the weighting function, $h(n)$, ensures that the predicted roughness quantitatively fits the subjective roughness of SAM tones in aspers (Fastl & Zwicker, 2007). Fig. 3.9 shows the weighting function, $h(n)$, obtained by a cubic spline interpolation of the values shown by crosses: the values are 1.35, 1.22, 1.01, 1, 0.72, 0.41, 0.33 and 0.25, respectively, for the model channels with CF of 0.125, 0.25, 0.5, 1, 2, 4, 8 and 16 kHz.

Since the processed stimuli are generally longer than the 30 ms long time frames in which is the roughness predicted, the overall predicted roughness for a given stimulus is calculated as the median across the roughness values predicted in the individual

successive time frames. Chapter 6 compares the roughness predicted by this roughness model with subjective data.

3.3 Summary

This Chapter described a roughness model composed of a peripheral stage and a central stage. The peripheral stage simulates the function of the peripheral ear: it transforms the analyzed acoustic signal into the simulated neural signal in auditory nerve (AN) fibers. It is composed of algorithms simulating the individual parts of the peripheral ear – outer- and middle-ear, cochlear mechanics, inner hair cells (IHCs) and AN synapse. The algorithms were adapted from the various literature and composed into one model. The central stage is composed of algorithms designed within the framework of this thesis. The roughness model predicts roughness in 30-ms time frames of acoustic stimuli. The resulting roughness is calculated as the median across the predicted roughness in the individual successive time frames.

Chapter 4

Verification of the Nobili *et al.* cochlear model

This Chapter shows the evidence that the chosen model of cochlear mechanics – the Nobili *et al.* cochlear model adapted from Nobili *et al.* (2003) – employed in the peripheral stage of the roughness model may adequately simulate the function of the human cochlea. Responses of the Nobili *et al.* cochlear model are compared with experimental physiological data measured in the cochlea of live mammals. The experimental physiological data were reproduced from the literature (Sellick *et al.*, 1983; Ruggero *et al.*, 1990; Cooper & Rhode, 1992; Rhode & Cooper, 1996; Ruggero *et al.*, 1997; Rhode & Recio, 2000; Robles & Ruggero, 2001). The Nobili *et al.* cochlear model – extended by a simple signal processing inner hair cells (IHCs) model, auditory nerve (AN) model and an optimal detector – is then used to predict masking thresholds caused by the pure tone and harmonic complex tone (Schroeder phase) maskers. The predicted masking thresholds are compared with the human subjective data reproduced from the literature (Oxenham & Dau, 2001a,b, 2004; Shen & Lentz, 2009). These maskers were used since many of the cochlear models cannot account for them. The maskers thus place a strong constraint on models of cochlear mechanics (Oxenham & Dau, 2001a).

4.1 Responses of the cochlear model

The experimental physiological isointensity responses, input/output (I/O) functions and impulse responses reproduced from the literature are compared with similar data obtained using the Nobili *et al.* cochlear model. Since the Nobili *et al.* cochlear model was designed with real parameters and dimensions of the human cochlea (Nobili *et al.*, 2003), it was not the aim to achieve a quantitative agreement with the experimental responses. Instead of that, the experimental and model responses are compared in order to show that the Nobili *et al.* model simulates the phenomena observed in the mammalian cochlea. These phenomena are believed to occur also in the human cochlea where it can be measured only indirectly using psychophysical experiments (e.g. Lopez-Poveda *et al.* (2007)).

4.1.1 Isointensity responses

Isointensity responses of the basilar membrane (BM) or the tectorial membrane (TM) to acoustic stimuli can be measured in live animals using laser velocimetry (Robles & Ruggero, 2001). Fig. 4.1 (left panel) shows magnitudes of the responses measured at a site of the chinchilla BM with CF of 10 kHz (Robles & Ruggero, 2001). The responses were measured by tone bursts with a level of 20, 40, 60 and 80 dB SPL (given next to the curves in the left panel of Fig. 4.1), respectively. The magnitudes of the isointensity responses depends on the level – broadens and its maximum shifts toward low frequencies as the level increases.

The right panel of Fig. 4.1 shows the estimates of cochlear gain calculated from the magnitudes of the isointensity BM or TM responses relative to the magnitudes of the stapes responses. Solid and dashed lines show the gains measured at the apical site of the TM of the chinchilla cochlea (CF of 0.35–0.5 kHz). The responses were measured using tone bursts with a level between 40 and 60 dB SPL. The data were reproduced from Rhode & Cooper (1996). Circles in Fig. 4.1 (right panel) show the data measured at the base (CF of 8.5 kHz) of the chinchilla BM. The data were reproduced from Ruggero *et al.* (1990) – the study used tone bursts with a level of 16 dB SPL to measure the responses. Squares in the right panel show the data measured at the base (CF of 17 kHz) of the guinea-pig BM. The data were reproduced from Sellick *et al.* (1983). Diamonds in the right panel show the data measured at the base (CF of 30 kHz) of the cat BM. The data were reproduced from Cooper & Rhode

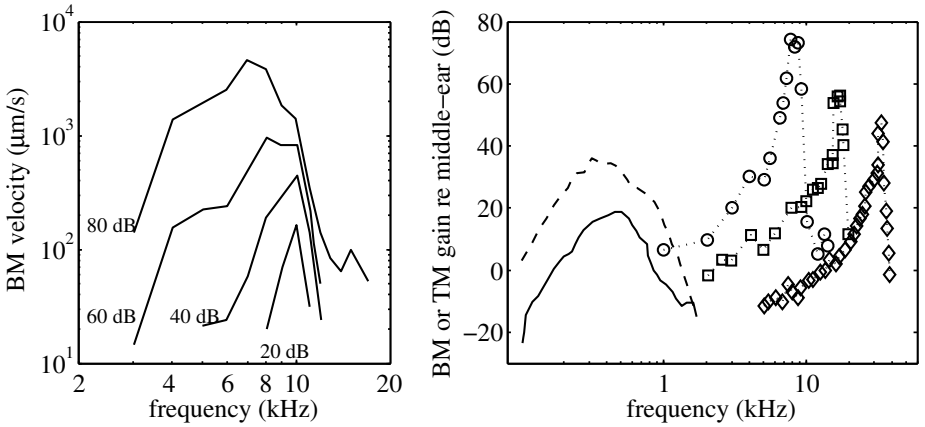

Figure 4.1: Left panel: Magnitudes of the isointensity responses – velocity of the basilar-membrane (BM) responses – to tone bursts of various levels as a function of frequency. The data were measured in the cochlea of live chinchillas at a site of the BM with CF of 10 kHz (Robles & Ruggero, 2001). Right panel: Gains estimated from the magnitudes of the isointensity responses of the tectorial membrane (TM) or the BM relative to the magnitudes of the stapes responses. The gains were measured at sites of various CF. Solid and dashed lines represent data measured in the TM of chinchillas (CF of 0.35–0.5 kHz,) (Rhode & Cooper, 1996); circles the data measured in the BM of chinchilla (CF of 8.4 kHz) (Ruggero *et al.*, 1990), squares the data measured in the BM of guinea-pig (CF of 17 kHz) (Sellick *et al.*, 1983); and diamonds the data measured in the BM of cat (CF of 30 kHz) (Cooper & Rhode, 1992). The data were reproduced from Robles & Ruggero (2001).

(1992).

The isointensity BM response phases measured in the live chinchilla cochleae are shown in Fig. 4.2. The data were reproduced from Ruggero *et al.* (1997). The left panel shows the data measured at a site of the BM with CF of 10 kHz using tone bursts with a level of 60 dB SPL. The phases were measured between the BM displacement toward scala tympani and condensation at the eardrum. The right panel shows the effect of level on the phases. The phases measured using tone bursts with a level of 40, 60 and 90 dB SPL are shown relative to the phase measured using 80-dB SPL tone bursts – positive values indicate phase lead relative to the 80-dB phase responses.

Fig. 4.3 shows the isointensity responses of the Nobili *et al.* cochlear model: the top panel shows magnitudes of the responses, and the bottom panel shows the response phases. The responses (peak BM displacements) were measured using pure tone bursts with a level of 20, 40, 60 and 80 dB SPL in three discrete outputs (along the length of the simulated BM) of the cochlear model – with CF of 0.25, 1 and 4 kHz. Table 4.1 shows equivalent rectangular bandwidth (ERB) of the measured response magnitudes at CF

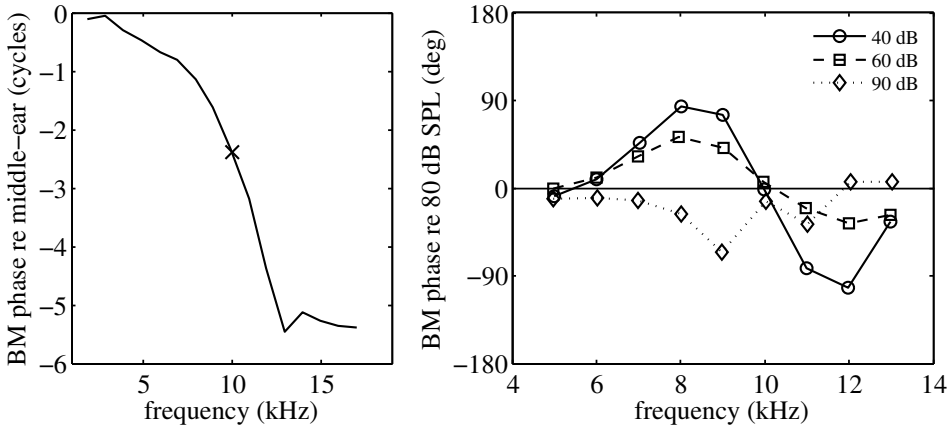

Figure 4.2: Isointensity BM response phases as a function of the frequency. The data were measured as the relative phases between the BM displacement toward scala tympani and condensation at the eardrum. The left panel shows the responses measured using tone bursts with a level of 60 dB SPL. The right panel shows the effects of level on the phases. The phases are expressed relative to the phases of responses at 80-dB tone bursts – positive phases indicate lead relative to the responses at 80 dB SPL. The data in both panels were reproduced from Ruggero *et al.* (1997). They were measured in the cochlea of live chinchillas at the site of the BM with CF of 10 kHz.

of 0.125, 0.25, 0.5, 1, 2 and 4 kHz. The bottom row shows ERB of the psychophysically estimated cochlear filters given by Eq. (3.4) (Moore & Glasberg, 1996).

The magnitudes of the isointensity responses of the Nobili *et al.* cochlear model – in agreement with the experimental data from the base of the mammalian cochlea (see Fig. 4.1) – broadens and its maximum shifts toward low frequencies as the level increases. The cochlear mechanics is still poor understood at low CFs (in the apical part of the cochlea) – the BM is accessible through scala vestibuli where the measurement techniques often require perforation of Reissner's membrane which may alter the physiological state of the preparation (Robles & Ruggero, 2001). The direction of the peak shift thus cannot be determined from the isointensity responses measured in the apical site of the mammalian cochlea (Robles & Ruggero, 2001). The isointensity responses measured in the inner hair cells (IHCs) and the auditory nerve (AN) fibers qualitatively agree with the BM responses at the base of the cochlea (CF $> \sim 1.5$ kHz). The data show similar broadening of the cochlear filters and the peak shift toward low frequencies as the level increases (Cheatham & Dallos, 2001). However, this contrasts with the IHC and AN responses measured in the intermediate ($0.75 <$ CF < 1.5 kHz) and apical (CF $< \sim 0.75$ kHz) part of the cochlea where the maximum magnitude does not shift or shifts toward high frequencies, respectively, as the level increases (Cheatham &

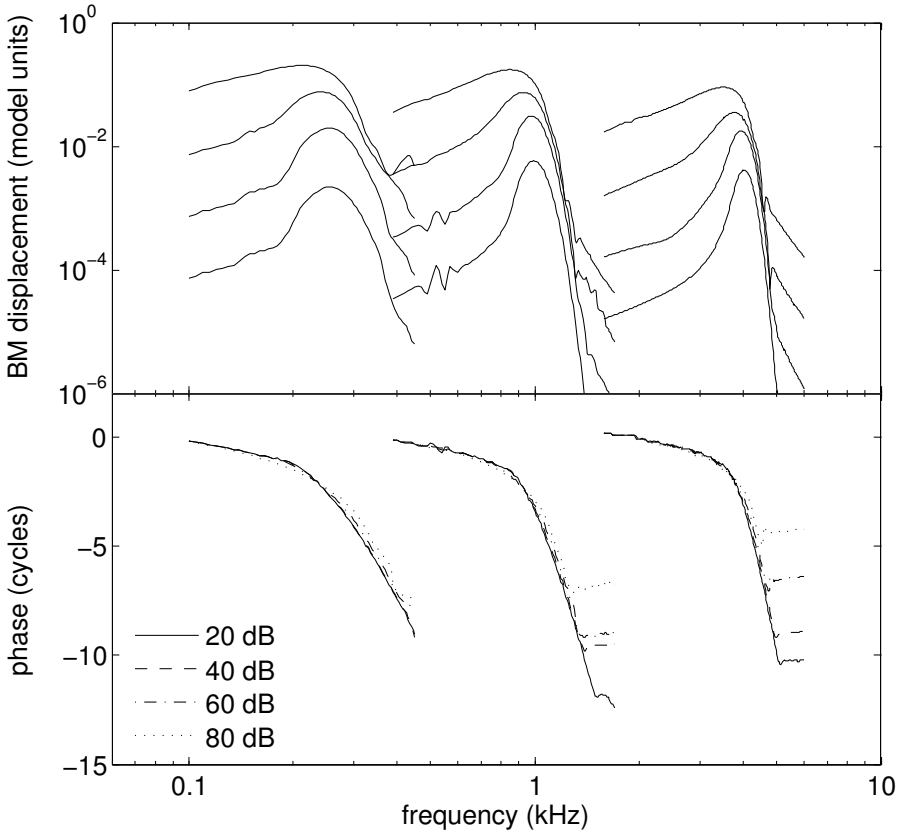

Figure 4.3: Isointensity responses of the Nobili *et al.* cochlear model. The top panel shows magnitudes of the responses. The bottom panel shows phases of the responses. The responses were measured at three discrete outputs of the cochlear model with CF of 0.25, 1 and 4 kHz using tone bursts with a level of 20, 40, 60 and 80 dB SPL.

Dallos, 2001; Anderson *et al.*, 1970; Zinn *et al.*, 2000; Carney *et al.*, 1999). This – under the assumption that the cochlear mechanics in small mammals is similar to the cochlear mechanics in humans – indicates that the Nobili *et al.* cochlear model does not adequately simulate the cochlear mechanics at low CF. However, the psychophysical tuning curves measured on human listeners indicate shifts of the filter's maxima in the direction toward low frequencies even in the apical part of the BM (Lopez-Poveda *et al.*, 2007). This would approve the applicability of the Nobili *et al.* cochlear model at low and intermediate CFs.

The isointensity response phases shown in the bottom panel of Fig. 4.3 have a similar shape as the experimental response phases shown in Fig. 4.2 (left panel). The effects of level on the isointensity response phases of the Nobili *et al.* cochlear model are shown in Fig. 4.4. The responses were measured at the output of the Nobili *et al.* cochlear model with CF of 1 kHz. The levels of the tone bursts were 40, 60, 80 and 90 dB SPL

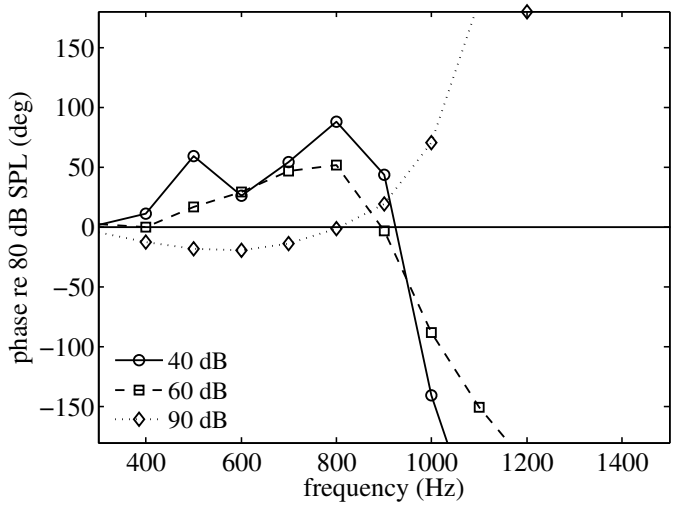

Figure 4.4: Isointensity phase responses of the Nobili *et al.* cochlear model measured using tone bursts of various levels. The responses were measured at the model output with CF of 1 kHz. The data are expressed relative to the phases of responses measured with 80-dB SPL tone bursts – positive values show the phase lead.

Table 4.1: Critical bandwidth of the Nobili *et al.* cochlear model

level	characteristic frequency (kHz)					
(dB SPL)	0.125	0.25	0.5	1	2	4
	equivalent rectangular bandwidth (Hz)					
20	43	62	89	141	225	390
40	43	62	90	148	245	521
60	43	70	122	201	337	818
80	54	98	168	307	528	1107
$\mathrm{ERB}_{\mathrm{Moore}}$	38	52	79	133	241	456

as is indicated in the legend. Positive values indicate phase lead relative to the phases measured using 80-dB SPL tone bursts. The model data measured with 40 and 60-dB tone bursts show, in agreement with the physiological data (Fig. 4.2), phase lead at frequencies below CF and phase lag at frequencies above CF. The phase lead and lag increases as the level decreases. The opposite show the responses measured using 90-dB tone bursts. The model (Fig. 4.4) and experimental mammalian (Fig. 4.2) data differ mainly at the highest frequencies where the experimental data seem to converge to values close to the 80-dB response phases whereas the model data not. This is caused by the plateaus – approximately constant phases independent on frequency – observed in the response phases (see Fig. 4.3). The plateaus in the Nobili *et al.* cochlear model

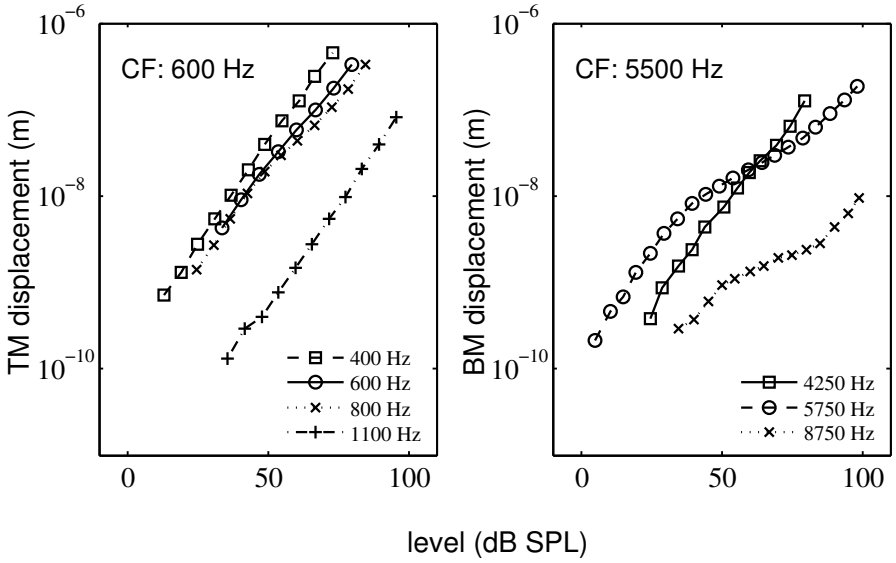

Figure 4.5: Input/output (I/O) functions of the responses measured in the cochlea of live chinchillas. The left panel shows the data measured at a site of the tectorial membrane (TM) with CF of 600 Hz using tone bursts with a frequency given in the legend. The data were reproduced from Rhode & Cooper (1996) (chinchilla CH16). The right panel shows I/O functions measured at a site of the BM with CF of 5.5 kHz using tone bursts with frequencies given in the legend. The data were reproduced from Rhode & Recio (2000) (chinchilla cb58).

response phases depend on the level.

4.1.2 Input/output functions

Input/output (I/O) functions of the BM or TM responses are compressively nonlinear if measured using stimuli with frequencies near CF. In contrast to this, responses to stimuli with higher or lower frequencies than CF exhibit linear growth (Robles & Ruggero, 2001). Fig. 4.5 (left panel) shows I/O functions measured in the TM of live chinchillas. The functions were measured at a site with CF of 600 Hz using tone bursts with a frequency of 400, 600, 800 and 1100 Hz. The 600-Hz and 800-Hz I/O functions are compressively nonlinear and the 1100-Hz I/O function is linear. The data were reproduced from Rhode & Cooper (1996) (chinchilla CH16). The compressive nonlinearity is stronger at the basal site of the BM (Robles & Ruggero, 2001). The right panel of Fig. 4.5 shows the I/O functions measured in the BM of chinchilla. The responses were measured at a site of the BM with CF of 5.5 kHz using tone bursts with a frequency of 4250, 5750 and 8750 Hz. The data were reproduced from Rhode & Recio (2000) (chinchilla cb58).

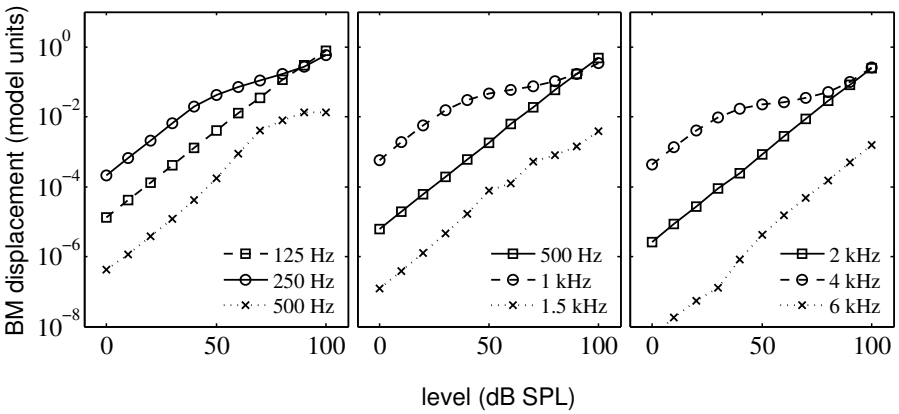

Figure 4.6: Input/output (I/O) functions of the Nobili *et al.* cochlear model. The functions shown in the left, center and right panel were measured in three discrete outputs of the cochlear model with CF of 0.25, 1 and 4 kHz, respectively. The frequencies of tone bursts used to measure the I/O functions are given in the legend of each panel.

The Nobili *et al.* cochlear model simulates the active function of the outer hair cells (OHCs) by force term $f_{\mathrm{OHC}_i}[\eta_i(t)]$ in Eq. (3.1). The force term is a sigmoid function which saturates for higher values of the OHC stereocilia displacement, $\eta(t)$. This allows to simulate the compressively nonlinear I/O functions. Fig. 4.6 shows the I/O functions of the Nobili *et al.* cochlear model responses (peak BM displacements). The left, center and right panel shows the functions measured at three discrete outputs of the cochlear model with CF of 0.25, 1 and 4 kHz, respectively. The frequencies of the tone bursts used to measure the functions are given in the legend of each panel. The data – in agreement with the experimental data shown in Fig. 4.5 – are compressively nonlinear if the frequency of the tone bursts is equal to CF.

4.1.3 Impulse responses

Responses of the BM and AN to acoustic clicks were shown to be near-invariant with stimuli level. This appears as near-invariant zero crossings of the fine time structure of the BM responses (e.g. Robles *et al.* (1976); de Boer & Nuttall (1997); Recio & Rhode (2000)) and level independent latencies of the peaks of histograms calculated from the responses measured in auditory nerve (AN) fibers (e.g. Goblick & Pfeiffer (1969); Lin & Guinan (2000)). Level near-invariance of impulse responses was observed using various measurement techniques: acoustic clicks (Robles *et al.*, 1976), indirectly by cross- or reverse-correlation using wideband noise stimuli (de Boer & Nuttall, 1997), and by applying inverse Fourier transform to the transfer functions measured

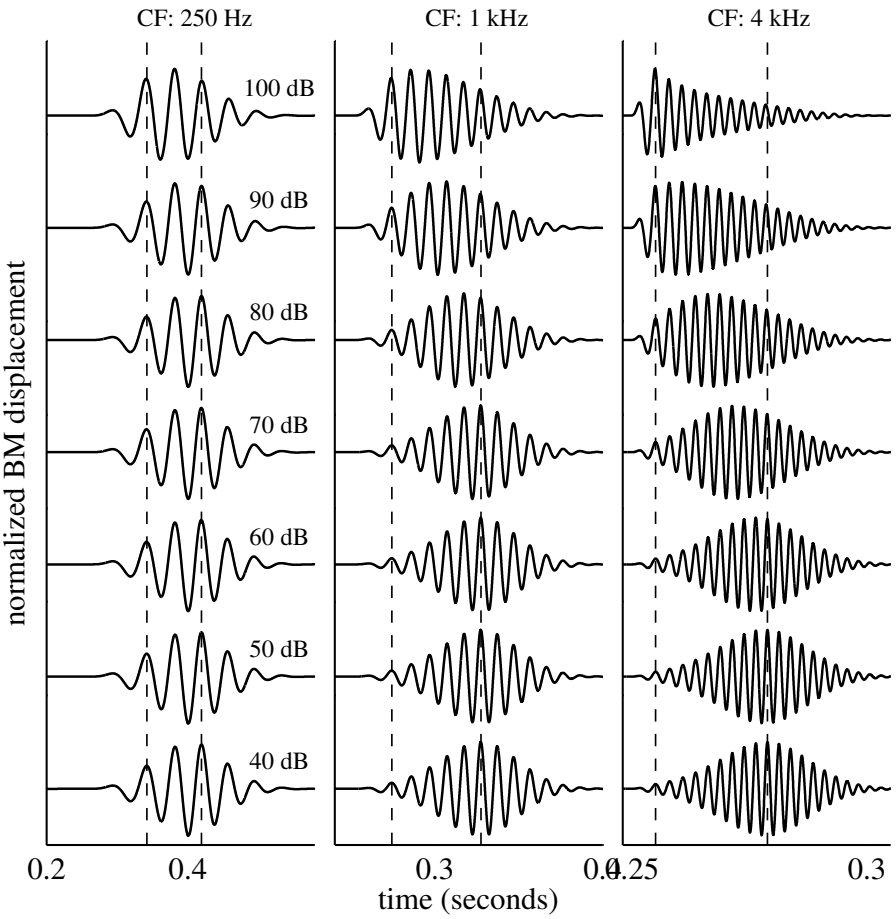

Figure 4.7: Impulse responses of the Nobili *et al.* cochlear model measured at the outputs with CF of 0.25, 1 and 4 kHz. The responses were measured using a unit impulse with a level between 40 and 100 dB SPL with a step of 10 dB as is given in the left panel.

using pure tones (Recio & Rhode, 2000).

The level near-invariance of the impulse responses has an implication for the cochlear mechanics, especially for the active feedback forces simulating the function of the OHCs (Shera, 2001). Shera (2001) studied this phenomenon, showed the implications for cochlear mechanics and designed a model in the frequency domain describing this phenomenon. Verhulst *et al.* (2012) adapted the concept of Shera (2001) and designed a time-domain cochlear model.

Fig. 4.7 shows the impulse responses of the Nobili *et al.* cochlear model. The responses were measured using a unit impulse of levels ranging from 40 to 100 dB SPL in a step of 10 dB. The left, center and right panel in Fig. 4.7 shows the impulse responses measured at the model output with CF of 0.25, 1 and 4 kHz, respectively. Dashed lines in each panel indicate peaks in the half-waves of the fine structure of the 40-dB impulse

response. The model impulse responses are, in agreement with the aforementioned physiological data, near-invariant (at least for levels between 40 and 80 dB SPL). A small shift of the peaks and zero crossings of the impulse responses is visible for the 90-dB and 100-dB responses measured at the model output with CF of 4 kHz.

Shera (2001) has argued that the level near-invariance of impulse responses places a strong constraint on the active function of the OHCs and contradicts many cochlear models. The Nobili *et al.* cochlear model was shown to fulfill this condition.

4.2 Prediction of masking experiments

The term "masking" describes a phenomenon when a stimulus is not perceived because of a presence of another stimulus called masker (Fastl & Zwicker, 2007). This phenomenon occurs in everyday life. We cannot understand what someone is saying in a loud environment, for example, on a loud street. If both stimuli – a masker and masked stimulus – are presented simultaneously, it is referred to as simultaneous masking. If the masker is presented before or after the masked stimulus, it is referred to as nonsimultaneous (forward or backward) masking (Fastl & Zwicker, 2007). Since cochlear mechanics affects the masking phenomenon, psychophysically measured masking thresholds can be used to verify cochlear models. This Section compares the predicted simultaneous masking thresholds measured using the Nobili *et al.* cochlear model with results of listening tests reproduced from the literature (Moore *et al.*, 1998; Oxenham & Dau, 2001a, 2004). Maskers used in the listening tests were pure tones and harmonic complex tones. The harmonic complex maskers were used since many of the cochlear models cannot account for them. The maskers thus place a strong constraint on models of cochlear mechanics (Oxenham & Dau, 2001a).

4.2.1 Method of the masking thresholds prediction

In order to predict the masking thresholds, the Nobili *et al.* cochlear model was extended by: (1) a signal processing model simulating the function of the inner hair cells (IHCs), adaptation in auditory nerve fibers and sensitivity of the human auditory system to temporal fluctuations; (2) and an optimal detector. These parts, which are briefly described below, were adapted from Dau *et al.* (1997). Models of the outer- and middle-ear were not incorporated into the overall auditory model. The author of

this thesis assumed that they should not significantly affect the masking thresholds for the used acoustic stimuli. They were omitted also in the auditory models used by Oxenham & Dau (2001a) to predict masking thresholds for the same stimuli as in this thesis.

The signal processing models of the IHC and AN synapse were used instead of the biophysical models implemented in the peripheral stage of the roughness model (see Section 3.1.3). This stems from the used optimal detector which was described with the signal processing models of the IHC and AN synapse (Dau *et al.*, 1997).

Auditory model: The first part of the auditory model is the Nobli *et al.* cochlear model. The output signal in each channel of the cochlear model, which represents the BM displacement, is first processed by a signal processing model of IHC (see Section 3.1.3). The IHC model is composed of a half-wave rectifier and a lowpass filter with a cutoff frequency of 1 kHz. This processing roughly simulates the transformation of the BM vibrations to the IHC membrane potential (Dau *et al.*, 1996, 1997; Jepsen *et al.*, 2008). The next stage simulates the adaptation of neural signal observed in the synapse of AN fibers. It is composed of five successive feedback loops (Püschel, 1988; Dau *et al.*, 1996, 1997; Jepsen *et al.*, 2008). The time constants of the successive feedback loops are 5, 50, 129, 253 and 500 ms (Dau *et al.*, 1996). The last stage of the auditory model is the modulation filterbank (Dau *et al.*, 1997). This stage simulates the sensitivity of the human auditory system to fluctuating sounds. The modulation filterbank is composed of twelve filters. The lowest modulation filter is a lowpass filter with a cutoff frequency of 2.5 Hz. The remaining are bandpass filters. The two bandpass filters between 0 and 10 Hz have a constant bandwidth of 5 Hz. The bandpass filters between 10 and 1000 Hz have a constant value of Q = 2 (logarithmic scaling). The amplitude transfer characteristics of the adjacent filters overlap at -3 dB points – spacing of the filters resembles the spacing of critical bands. Only the (Hilbert) envelopes are taken as the model output for the bandpass filters in the range between 10 and 1000 Hz (Dau *et al.*, 1997).

Optimal detector: In order to predict masking thresholds, the output signals of the auditory model – referred to as the "internal representation" (Dau *et al.*, 1997; Jepsen *et al.*, 2008) – obtained in response to the masker and the masker plus test tone must be compared. This is in this thesis done by means of an optimal detector described by Dau *et al.* (1996, 1997) and Jepsen *et al.* (2008). The same method was

used by Oxenham & Dau (2001a). The optimal detector first calculates a template which is the normalized difference between the internal representation of the masker plus suprathreshold (approximately 10 dB above threshold) test tone and the internal representation of the masker only. In the same way as the template, the optimal detector calculates a difference between the internal representation of the masker plus test tone of a specific level and the internal representation of the masker only. Whether or not is the test tone of the specific level detected is then calculated from the crosscorrelation between the template and the difference. The resolution of the optimal detector is limited by the internal noise. Variance of the internal noise is one of the parameters of the optimal detector. Its value is constant for all channels of the auditory model. The value of variance was set experimentally to the value for which all of the predicted thresholds for harmonic complex maskers predicted in Section 4.2.3 (see Fig. 4.9) are at or lower than the subjective thresholds.

The overall model was implemented and all the predictions were done in Matlab (The MathWorks, Inc., Natick, MA) environment.

4.2.2 Tone on tone masking: upward spread of masking threshold

If masker and masked stimuli are pure tones, it is referred to as tone on tone masking. Subjective masking thresholds in tone on tone masking conditions show the upward spread of masking thresholds toward high frequencies as the masker level increases (Moore *et al.*, 1998; Fastl & Zwicker, 2007). The upward spread of masking thresholds was observed also with other narrowband maskers, for example, narrowband noise (Moore *et al.*, 1998). The phenomenon is a consequence of the level dependent shape of the cochlear filters (see Fig. 4.1 and 4.3). This Section compares subjective masking thresholds for tone on tone maskers reproduced from Moore *et al.* (1998) with predicted masking thresholds measured using the auditory model with the Nobili *et al.* cochlear model.

Stimuli: The stimuli setting was same as the stimuli setting described by (Moore *et al.*, 1998). A pure tone with a frequency of 1 kHz was used as a masker. The level of the tone was either 65 or 85 dB SPL. A pure tone was used also as a test tone (masked tone). The starting phase of the test tone was 90° relative to the starting phase of the

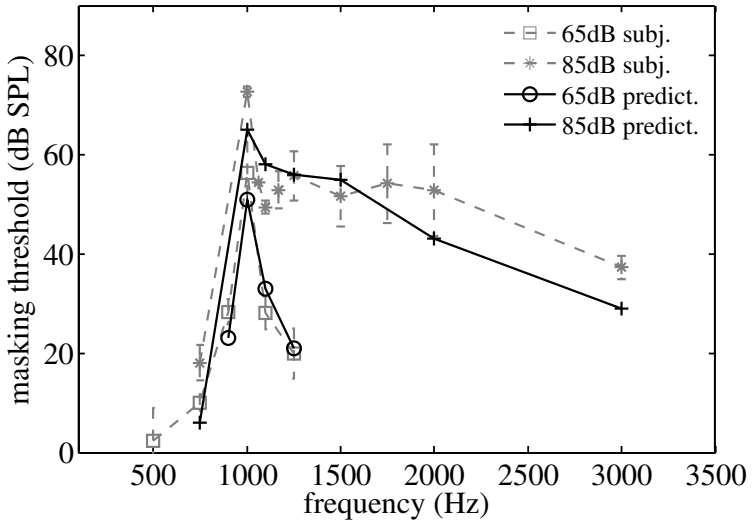

Figure 4.8: A test tone level masked by a pure tone masker as a function of the test tone frequency. Gray symbols and lines show mean values of the subjective masking thresholds reproduced from MOORE *et al.* (1998). Black symbols and lines show the predicted thresholds obtained by means of the Nobili *et al.* cochlear model. The level of the masker was 65 or 85 dB SPL.

masker – powers of the masker and test tone added when their frequencies were equal. For the 65-dB masker, the test tone frequency was 0.9, 1, 1.1 and 1.25 kHz; for the 85-dB masker, the test tone frequency was 0.75, 1, 1.1, 1.25, 1.5, 2 and 3 kHz. The duration of the masker and test tone was 200 ms and it was ramped on and of with 10-ms raised-cosine ramps.

Procedure of the masking threshold prediction: The masking thresholds were predicted by means of the auditory model and the optimal detector. Only the channels of the auditory model with CF between $0.9f_s$ and $1.1f_s$ (f_s is a test tone frequency) were fed into the optimal detector. The remaining channels were not taken into account in order to avoid the detection of the test tone caused by combination products produced by the interaction between the masker and masked tone in the cochlea. During psychophysical experiments, the combination products maybe masked by noise in order to avoid its masking effects. This method was used to get the data shown below (reproduced from MOORE *et al.* (1998)). The optimal detector first calculated a template (see Section 4.2.1) from the output of the auditory model in response to the masker plus the suprathreshold (∼10 dB above threshold) test tone and the masker only. The masking threshold was then predicted using a tracking algorithm. First, the test tone of an arbitrary chosen level plus the masker only was fed into the auditory

model. The optimal detector then compared (see Section 4.2.1) the corresponding outputs of the auditory model (output signals in the corresponding channels of the auditory model) with the template (calculated template in the corresponding channels of the auditory models) and gave a positive or negative response – the test tone was detected or not. The level of the test tone was then decreased by a step of 5 dB if the test tone was detected, otherwise the level was increased by 5 dB. The measurement was then repeated. After one reversal – a point where the decreasing level of the test tone started to be increased, or vice versa – the step size was set to 1 dB and after next two reversals, a mean value of the test tone levels for the last two reversals, either from up to down or down to up, was taken as the predicted masking threshold.

Results: Fig. 4.8 shows the mean values of the test tone levels at the threshold of masking plotted as a function of the test tone frequency. The subjective data are shown as gray markers and connected by dashed lines: squares show the thresholds for the 65-dB masker, and stars show the thresholds for the 85-dB maskers. The data were reproduced from Moore *et al.* (1998): for the 85-dB masker from Fig. 4 (the data for a 75-dB SPL low-pass noise masking combination products), and for the 65-dB from Fig. 1, 2 and 3. The predicted masking thresholds are shown as black markers: circles show the thresholds for the 65-dB masker and plus signs show the thresholds for the 85-dB masker. The predicted thresholds are for most of the stimuli within a range of the standard deviations of the subjective data. The predicted data show the upward spread of masking thresholds as the masker level increases.

4.2.3 Schroeder phase maskers: masker phase effects

A number of studies showed the effects of relative phase between the spectral components of the masker on the masking threshold. The effects were observed with simultaneous maskers (e.g. Smith *et al.* (1986); Kohlrausch & Sander (1995); Lentz & Leek (2001); Oxenham & Dau (2001a)) and forward maskers (e.g. Wojtczak & Oxenham (2009)). Maskers used in these studies were harmonic complexes composed of a series of equal amplitude pure tones, with the starting phase, θ_n, of each n-th harmonics given by the equation introduced by Schroeder (1970) and later modified by Lentz & Leek (2001)

$$\theta_n = C\pi n(n-1)/N, \tag{4.1}$$

where N is the overall number of harmonics within the complexes and C is the parameter which sets the phase curvature of the complexes. So called "positive Schroeder phase" and "negative Schroeder phase" complexes are stimuli with the starting phases given by Eq. (4.1) for $C = +1$ and $C = -1$, respectively. These stimuli have similar temporal envelopes but may produce masking thresholds which differ as much as 20 dB (Smith *et al.*, 1986).

It is believed that the masker phase effects are connected with the shape of the temporal envelope of the complexes after the auditory filtering (Smith *et al.*, 1986; Kohlrausch & Sander, 1995). Smith *et al.* (1986) filtered the stimuli with a cochlear model and observed low peaks and high valleys in the temporal envelope of the less effective maskers. Recio & Rhode (2000) observed the same in the mammalian cochlea. Kohlrausch & Sander (1995) conducted psychophysical experiments with tone bursts shorter than the period of the Schroeder phase maskers. They adjusted the temporal position of the tone bursts within the maskers and observed larger variations of the masking thresholds expressed as a function of the temporal position of the tone bursts (modulation masking patterns) for less effective maskers. This indirectly supports the above mentioned physiological and model observations of low peaks and high valleys in the BM responses to stimulation by the less effective maskers. The observations indicate that the maskers are less effective because the masked signal is detected in the low valleys of the temporal envelope after auditory filtering (Smith *et al.*, 1986; Kohlrausch & Sander, 1995).

Masking thresholds for positive and negative Schroeder phase maskers varying in a number of harmonics and bandwidth are in this Section predicted by the Nobili *et al.* cochlear model. The predicted thresholds are compared with the subjective data obtained with normal hearing listeners (reproduced from Oxenham & Dau (2001a)).

Stimuli: The stimuli setting was same as described by Oxenham & Dau (2001a). Two different masker configurations were employed in the experiment. Both were harmonic complexes with fundamental frequency $f_0 = 100$ Hz. The first configuration, called "comparison", contained $N = 19$ equal amplitude harmonics ranging between 0.2 to 2 kHz. The second configuration, called "broadband", contained $N = 49$ equal amplitude harmonics ranging between 0.2 and 5 kHz. The starting phases of the individual harmonics of the maskers were given by Eq. (4.1) ($C = +1$ for positive Schroeder phase masker ($m+$) and $C = -1$ for negative Schroeder phase masker ($m-$)). In order to test the effects of number of harmonics and masker bandwidth on masking

thresholds, different variants of the "broadband" masker configuration were made. The variants differed just in the number of harmonics, its starting phases were not changed. Only the harmonics between 0.8 and 1.2 kHz were present for the "1 kHz narrowband" masker and the harmonics between 3.2 and 5 kHz for the "4 kHz narrowband" masker. The harmonics between 0.2 and 1.2 kHz were present for the "1 kHz lowpass" masker and the harmonics between 0.8 and 5 kHz for the "1 kHz highpass" masker. A level of each harmonic component was set to 60 dB SPL. The duration of all maskers was 320 ms and it was ramped on and off with 10-ms raised-cosine ramps. A pure tone with a duration of 260 ms was used as a test tone. It was temporally centered within the masker and ramped on and off with 30-ms raised-cosine ramps. The frequency of the test tone, f_s, was either 1 or 4 kHz and its starting phase was set to be equal to the starting phase of the masker harmonic component with the same frequency.

Procedure of the masking threshold prediction: The procedure used to predict masking thresholds was same as in Section 4.2.2 unless otherwise stated. The masking thresholds were predicted by means of the auditory model and optimal detector. Only the channels of the auditory model with CF between $0.7f_s$ and $1.3f_s$ were fed into the optimal detector in order to avoid detection of the test tone caused by combination products. The same range was used by Oxenham & Dau (2001a).

Results: The open markers in Fig. 4.9 show the mean values across listeners of the psychophysically measured masking thresholds published by Oxenham & Dau (2001a) (Experiment 1, Fig. 3). The filled markers show the thresholds predicted by the auditory model with the Nobili *et al.* cochlear model. Thresholds for the positive Schroeder phase maskers ($m+$) are plotted as downward-pointing triangles and those for the negative Schroeder phase maskers ($m-$) as upward-pointing triangles.

The best agreement between the subjective and predicted thresholds is for the "1 kHz comparison" masker. For the rest of the stimuli, the agreement is only qualitative. The model predicts the smallest difference between the positive and negative masker thresholds for the "1 kHz narrowband" masker. The predicted thresholds for the 4-kHz maskers are much lower than the subjective data. This may indicate too narrow cochlear filters around 4 kHz. However, it may also indicate a different way of test tone detection at higher frequencies caused by the lost phase locking of the neural signal to the test tone fine structure (Johnson, 1980). The used optimal detector would not be able to account for this effect.

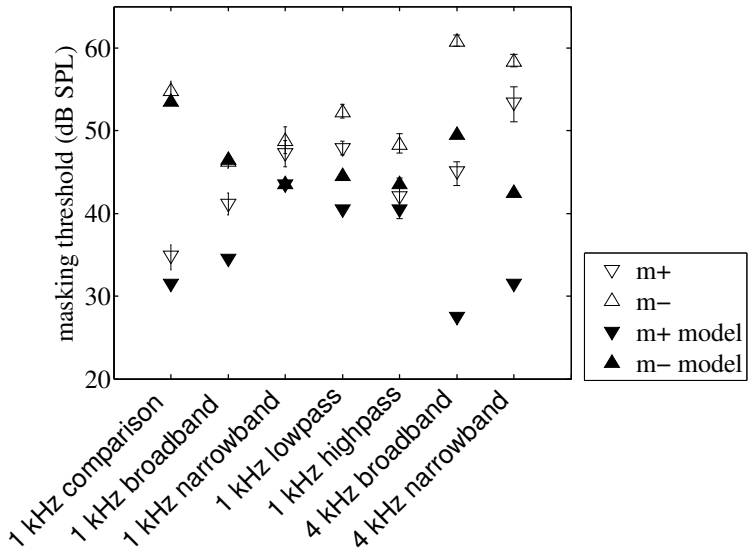

Figure 4.9: Subjective and predicted masking thresholds for Schroeder phase maskers. Open markers represent the mean values of the subjective masking thresholds reproduced from (Oxenham & Dau, 2001a). Filled markers represent the predicted masking thresholds. The data were obtained for different compositions of the maskers as is given in the abscissa. Triangles depict the thresholds for the positive Schroeder phase ($m+$) maskers, and upside down triangles for the negative Schroeder phase ($m-$) maskers.

Although, the agreement between the subjective and predicted thresholds is mostly only qualitative, the Nobili *et al.* cochlear model can account for the phase effects. These results were not reached by many of the cochlear models (see Oxenham & Dau (2001a)) and this type of stimuli thus places a strong constraint on models of cochlear mechanics.

4.2.4 Complex maskers: frequency selectivity

Oxenham & Dau (2001a) (Experiment 2) used harmonic complex tones to estimate frequency selectivity of the human auditory system. Their psychophysically measured data are below compared with the predicted data. The author used these psychophysical data to adjust the frequency selectivity of the Nobili *et al.* cochlear model.

Stimuli: The stimuli setting was same as described by Oxenham & Dau (2001a) (Experiment 2). Complex tones composed of $N = 40$ harmonics ranging from 0.1 to 4 kHz with fundamental frequency $f_0 = 100$ Hz were used as maskers. The starting

phases, θ_n, of the harmonics were given by Eq. (4.1) ($C = +1$ for the positive Schroeder phase ($m+$) and $C = -1$ for the negative Schroeder phase ($m-$) masker) and by the relation $\theta_n = \pi/2$ (cosine phase masker). The masker level was 60 dB SPL per harmonic component, its duration was 320 ms and it was ramped on and off with 10-ms raised-cosine ramps. A pure tone of a frequency $f_s = 2$ kHz was used as a test tone. The duration of the test tone was 260 ms, and it was temporally centered within the masker and ramped on and off with 30-ms raised-cosine ramps. Harmonics of the masker placed symmetrically around the test tone frequency (2 kHz) were removed in order to create a spectral notch within the masker. Notches of various widths were created such that the distances between the test tone frequency and the edge of the masker were 0.05, 0.1, 0.2 and $0.4f_s$ which corresponded to 1, 3, 7 and 11 removed harmonics, respectively. The starting phase of the test tone was set randomly in each trial of the threshold prediction.

Procedure of the masking threshold prediction: The test tone starting phase affects the predicted masking thresholds. Since the test tone starting phase is set randomly in each trial of the threshold prediction, the method comparable with the method used to measure the subjective masking thresholds – 2-down, 1-up tracking rule (see Oxenham & Dau (2001a)) – was used. The test tone level was decreased after two consecutive positive responses given by the optimal detector and increased after each negative response. The initial step size was 5 dB which was then after the first four reversals decreased to 2 dB. The threshold was then estimated as the mean of the remaining six reversals. The thresholds were predicted five times for each stimuli setting and the mean value was taken as the result. As well as in the experiment above (Section 4.2.3), only the model channels with CF between $0.7f_s$ and $1.3f_s$ were fed into the optimal detector to avoid the potential threshold detection caused by combination products.

Results: Fig. 4.10 shows the mean values of the subjective (open markers) and predicted (filled markers) masking thresholds measured with the positive Schroeder phase ($m+$, downward-pointing triangles), negative Schroeder phase ($m-$, upward-pointing triangles) and cosine phase (diamonds) maskers. The subjective data were reproduced from Oxenham & Dau (2001a) (Experiment 2, Fig. 5). The data represent the mean values of the masking thresholds estimated in normal hearing listeners. The abscissas of the graphs show the normalized distance between the edge of the notch and the test tone frequency (2 kHz) – 0 corresponds to the masker with all harmonics;

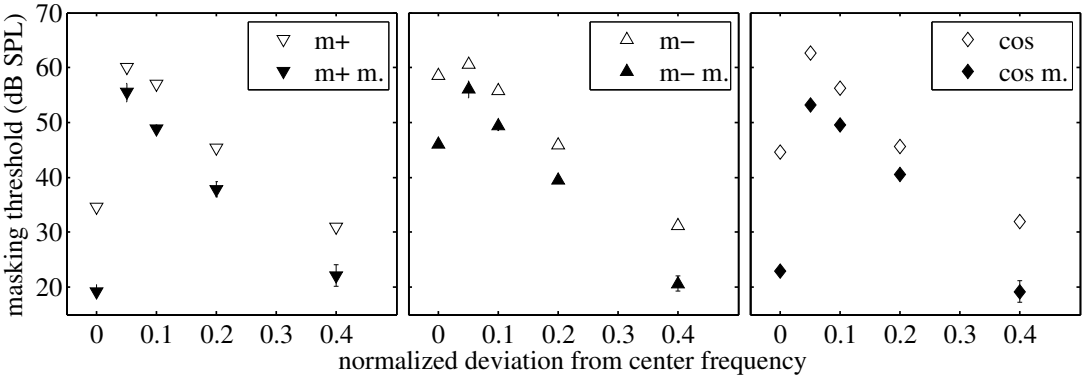

Figure 4.10: Mean values of the subjective and predicted masking thresholds for harmonic complex maskers. Open markers represent the mean values of the subjective masking thresholds reproduced from (Oxenham & Dau, 2001a). Filled markers represent mean values of the predicted masking thresholds obtained by means of the auditory model. The abscissa shows the normalized distance between the edge of the notch and the test tone frequency (2 kHz). The left, center and right panel shows the data for the positive Schroeder phase ($m+$), negative Schroeder phase ($m-$), and cosine phase maskers, respectively

0.05, 0.1, 0.2 and 0.4 correspond to the masker with 1, 3, 7 and 11 removed harmonics, respectively.

The predicted and subjective data agree only qualitatively. However, the model seems to reach the same frequency selectivity as show the subjective data. The highest discrepancy is for the cosine phase masker with all spectral components – for the normalized deviation from the center frequency of 0.

4.2.5 Schroeder phase maskers: effects of masker level

Oxenham & Dau (2001b) and Shen & Lentz (2009) observed that the psychophysically measured masking threshodls for Schroeder phase maskers depend also on the masker sound pressure level (SPL). This Section compares the subjective data measured by Oxenham & Dau (2001b); Shen & Lentz (2009) with the thresholds predicted by the auditory model with the Nobili *et al.* cochlear model.

Stimuli: Oxenham & Dau (2001b) used a pure tone with a frequency, f_s, of 1 kHz as a test tone and a harmonic complex tone with spectral components between $0.4f_s$ and $1.6f_s$ as a masker. The duration of the masker was 320 ms and it was ramped on and off with 30-ms raised-cosine ramps. The duration of the test tone was 260 ms and

it was ramped on and off with 50-ms raised-cosine ramps. The level of the maskers was 40, 60 and 85 dB SPL. Shen & Lentz (2009) used a pure tone with a frequency, f_s, of 2 kHz as a test tone and a harmonic complex tone with spectral components between $0.4f_s$ and $1.6f_s$ as a masker. The duration of the test tone and masker was 300 ms and it was ramped on and off with 30 ms raised-cosine ramps. The level of the masker was 50, 70 and 90 dB SPL. The phases between the individual spectral components of both, 1-kHz and 2-kHz, maskers were set by Eq. 4.1 for C equal to -1, 0, 0.25, 0.5, 0.75 and 1. The starting phase of the test tone was set to be equal to the starting phase of the spectral component with the same frequency as the test tone frequency. This contrasts with the procedure used to get the subjective data in Oxenham & Dau (2001b) and Shen & Lentz (2009), where the test tone starting phase was set randomly in each trial of the psychophysical experiment. The author predicted thresholds also with the the randomly set starting phase, as well as is the case of the psychophysical experiments. However, it did not affected the predicted threshold so much. The thresholds were just few dB above the predicted thresholds shown below in Results. Since it would be necessary to repeat the threshold predictions if the starting phase was set randomly, the fixed starting phase was used instead.

Procedure of the masking threshold prediction: Procedure used to predict the masking thresholds was same as in Section 4.2.3.

Results: Fig. 4.11 shows the test tone thresholds relative to the masker level as a function of the masker phase curvature (parameter C). The panels in the upper row show the subjective data: the left panel shows the mean values across listeners for the 1-kHz maskers (reproduced from Oxenham & Dau (2001b) (Experiment 2, Fig. 6)); and the right panel shows the mean values across listeners for the 2-kHz maskers (reproduced from Shen & Lentz (2009) (Experiment 1, Fig. 2)). Panels in the bottom row show the data predicted by the auditory model with the Nobili *et al.* cochlear model. Each panel shows the data for the same stimuli setting as was used to obtain the subjective data shown in the upper panels. The data shown in the left panels (diamonds, triangles and squares), were measured using the 1-kHz maskers at a level of 40, 60 and 85 dB SPL, respectively. The data shown in the right panels (diamonds, triangles and squares) were measured using the 2-kHz maskers at a level of 50, 70, and 90 dB SPL, respectively.

The predicted and subjective data agree qualitatively. The best qualitative agreement

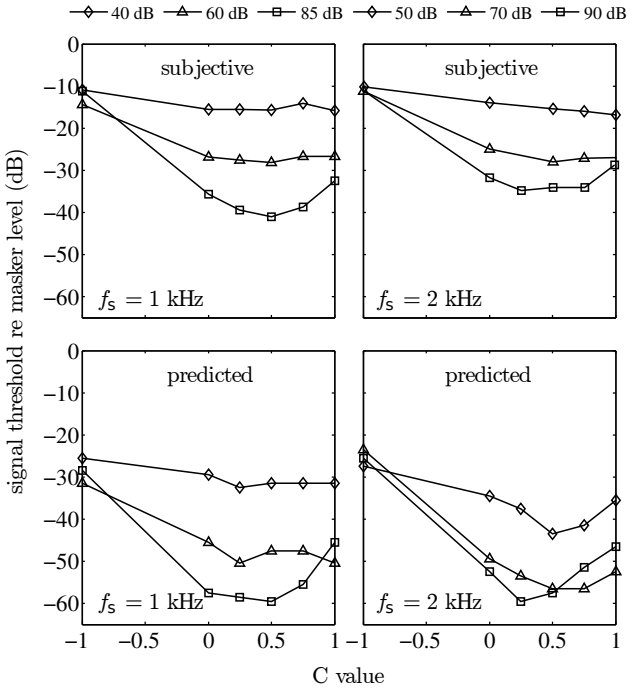

Figure 4.11: Masking thresholds in harmonic complexes relative to the masker level. The thresholds are shown as a function of the masker phase curvature for maskers of various sound pressure levels (SPL). Panels in the upper row show the subjective data. The upper left panel shows the data for a test tone with a frequency, f_s, of 1 kHz and the masker with a fundamental frequency, f_0, of 50 Hz and a level of 40, 60 and 85 dB SPL – the data represent mean values across different listeners and were reproduced from Oxenham & Dau (2001b). The upper right panel shows the data for a test tone with f_s of 2 kHz and masker with f_0 of 100 Hz and a level of 50, 70 and 100 dB SPL – the data represent mean values across different listeners and were reproduced from Shen & Lentz (2009). The bottom row show masking thresholds predicted by the auditory model with the Nobili *et al.* cochlear model. The data were predicted for the maskers and the test tones of the same parameters as were used to measure the subjective data shown in the panels above.

was reached for the 1-kHz maskers. The predicted data are more than 10 dB below the subjective data. This is partly caused by the fixed starting phase of the test tone used to predict the thresholds, but also by the auditory model and the optimal detector. The agreement between the predicted and the subjective data for the 2-kHz maskers is worse. However, the predicted data seems to follow the psychophysically observed trend – increasing difference between the minimal and maximal masking threshold and the shift of the lowest masking threshold as the masker level increases. Shen & Lentz (2009) argued that this shift does not mean a change of the phase curvature of the auditory filters as the level increases[1], but the level dependent broadening of

[1]Physiological data shows that the level affects the isointensity response phases of the BM (see Fig. 4.2). However, the change of the phase curvature is not as high as would indicate the psychophysical

the amplitude transfer functions of auditory filters. As a results, the off-frequency phase curvature of the auditory filter is involved in masking which decreases magnitude of the auditory filter phase curvature. The shown predicted thresholds support this hypothesis.

4.2.6 Schroeder phase maskers: passive cochlear model

Oxenham & Dau (2004) conducted listening tests with normal-hearing listeners and listeners with sensorineural hearing loss. They measured effects of phase in Schroeder phase maskers. The results showed reduced effects of masker phase in hearing-impaired listeners. This may indicate that the masker phase effects are related to the outer hair cell (OHC) function. This Section shows how the active function of the cochlea, which is in the Nobili *et al.* cochlear model simulated by the force term, f_{OHC}, in Eq. (3.1), affects the predicted masker phase effects.

Stimuli: The stimuli setting was given by Oxenham & Dau (2004). A test tone with a frequency, f_s, of 1 kHz was used together with a harmonic complex masker with spectral components between $0.4f_s$ and $1.6f_s$. The fundamental frequency, f_0, of the masker was 100 Hz. The masker duration was 320 ms and it was ramped on and off with 10-ms raised-cosine ramps. The test tone was temporally centered within the masker and its duration was 260 ms. The masker level was 93 dB SPL. The phases between the individual spectral components of the masker were set by Eq. (4.1) for C values between -1 and 1 with a step of 0.25. The starting phase of the test tone was set to be equal to the starting phase of the spectral component of the masker at the same frequency as the test tone. This contrasts with the procedure used to get the subjective data in Oxenham & Dau (2004) were the test tone starting phase was set randomly in each trial of the psychophysical experiment. Since no significant effects of the random phase design on the predicted thresholds have been observed, the fixed phase design was used instead in order to decrease the time of the masking threshold predictions (see also Section 4.2.5).

Procedure of the masking threshold prediction: The procedure used to predict the masking thresholds was same as in Section 4.2.3 and 4.2.5.

data shown by Shen & Lentz (2009).

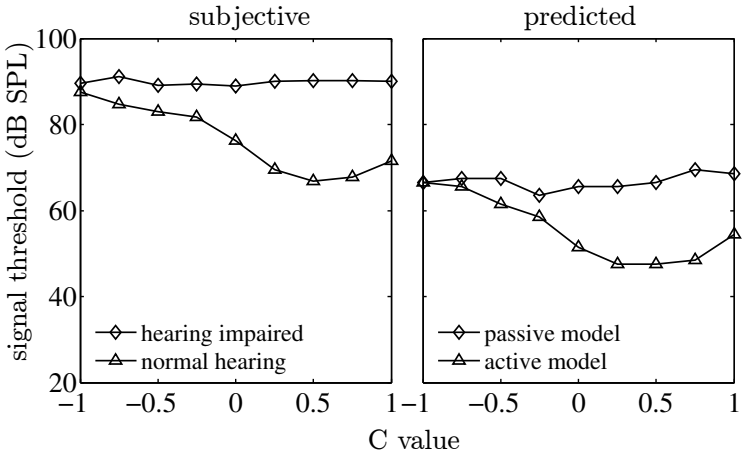

Figure 4.12: Masking thresholds in harmonic complex maskers as a function of the masker phase curvature. The left panel shows subjective data reproduced from Oxenham & Dau (2004): diamonds show data for hearing-impaired listener NH1, triangles show data for normal-hearing listener NH3. The right panel shows the predicted masking thresholds obtained by the Nobili *et al.* cochlear model: diamonds show thresholds for the passive Nobili *et al.* model, triangles show thresholds for the active Nobili *et al.* model.

Results: Oxenham & Dau (2004) showed reduced masker phase effects in hearing-impaired listeners. Their results for hearing-impaired listener HI1 are shown in the left panel (diamonds) of Fig. 4.12. The left panel (triangles) also shows their results for normal-hearing listener NH3. The masking thresholds are shown as a function of parameter C (masker phase curvature).

The hearing impairment caused by disabled OHCs can be simulated in the Nobili *et al.* cochlear model by removing the active force term, f_{OHC}, in Eq. (3.1). Fig. 4.12 (right panel) shows the effect of the active force term on masking thresholds in harmonic complex maskers (the same stimuli as for the data in the left panel). Diamonds show the masking thresholds predicted by the passive cochlear model (removed force term f_{OHC}). Triangles show the masking thresholds predicted by the active cochlear model. The predicted data qualitatively agree with the subjective data. As well as in Section 4.2.5 above, the model predicted masking thresholds more than 10 dB below the subjective thresholds. This can be partly explained by the fixed starting phase of the test tone during the thresholds prediction. However, it must be caused also by the auditory model and the optimal detector. The passive and active model, in agreement with the subjective data, predicted the same masking threshold for the negative Schroeder phase masker ($C = -1$).

4.3 Summary

This Chapter verified the ability of the Nobili *et al.* cochlear model to simulate physiological data and psychophysical data. In other words, verified the ability of the cochlear model to adequately simulate the function of the cochlea. The physiological data (reproduced from the literature) – isointensity responses, input/output (I/O) functions and impulse responses measured in the cochlea of live mammals – were compared with the responses of the Nobili *et al.* cochlear model. In the second part of this Chapter, the Nobili *et al.* cochlear model was extended by signal processing models of inner hair cells (IHCs) and auditory nerve (AN) synapse, and a modulation filterbank. Such created auditory model was then used to predict psychophysically measured masking thresholds (reproduced from the literature) for pure tone and harmonic complex tone maskers.

The responses of the Nobili *et al.* cochlear model qualitatively agreed with those measured in the live mammalian cochleae. Both – experimental and the Nobili *et al.* cochlear model – isointensity responses are level dependent. Functions showing the isointensity response magnitudes (amplitude characteristics of cochlear filters) broadens and its peak shifts as the level of the tone bursts used to measure the responses increases. The response magnitudes of the Nobili *et al.* cochlear model broadens and its peak shifts toward low frequencies also when they are measured in the model outputs with characteristic frequency (CF) below approximately 1.5 kHz. At these CFs, a direction of the peak shift cannot be determined from the mechanical data measured at the basilar membrane (BM) or tectorial membrane (TM) of live mammals (Robles & Ruggero, 2001). The data measured in the AN fibers of the live mammalian cochlea showed that the peak of the response magnitudes does not shift or shifts toward high frequencies at CFs below approximately 1.5 kHz (Cheatham & Dallos, 2001; Anderson *et al.*, 1970; Zinn *et al.*, 2000; Carney *et al.*, 1999). This contradicts the Nobili *et al.* cochlear model responses. On the other hand, the psychophysical tuning curves measured in humans indicate that the peak of the magnitude responses shifts toward low frequencies even at low CFs (at the apical site of the cochlea) (Lopez-Poveda *et al.*, 2007). This agrees with the responses of the Nobili *et al.* cochlear model.

Isointensity response phases of the Nobili *et al.* cochlear model showed qualitatively similar dependency on the level of the tone bursts as the data measured at the BM of live mammals – compare Fig. 4.2 and 4.4.

Input/output (I/O) functions of the Nobili *et al.* cochlear model are compressively nonlinear when measured using tone bursts with a frequency near the CF of the model output. I/O functions measured using tone bursts with a frequency further away from the CF of the model output are more linear. These results agreed with the observations conducted in the live mammalian cochleae (compare Fig. 4.5 and 4.6).

Impulse responses of the BM and AN were shown to be level near-invariant. This condition places a strong constraint on the active function of the OHCs and contradicts many cochlear models (Shera, 2001). The Nobili *et al.* cochlear model was shown to fulfill this condition (see Fig. 4.7).

Psychophysically measured tone on tone masking thresholds show upward spread of masking as the masker level increases (Moore *et al.*, 1998). This upward spread of masking was observed also in the masking thresholds predicted by the auditory model with the Nobili *et al.* cochlear model.

The Nobili *et al.* cochlear model was also used to predict masking thresholds for harmonic complex maskers. The relative phase between the spectral components of the maskers was manipulated in order to show the phase effects on the masking threshold. The starting phases of the masker spectral components were set according to Eq. (4.1) introduced by Schroeder (1970). Such created "Schroeder phase" maskers may produce masking thresholds differing by more than 20 dB (e.g. Kohlrausch & Sander (1995)). Section 4.2.3 showed the effects of relative phase between the spectral components of the masker on the masking thresholds. Section 4.2.4 used harmonic complex maskers to estimate frequency selectivity of the cochlear model. These experimental results were used to set frequency selectivity of the Nobili *et al.* cochlear model. Section 4.2.3 showed the effects of level on the masking thresholds for Schroeder phase maskers and Section 4.2.6 showed how the active function of the cochlea affects the effects of phase on the masking thresholds for Schroeder phase maskers.

The auditory model with the Nobili *et al.* cochlear model predicted the psychophysically observed effects in harmonic complex maskers. Although, an agreement between the model predictions and subjective data (reproduced from the literature) was mostly only qualitative, the results must be looked at in the scope of the fact that not so many cochlear models can predict these effects. As far as the author knows, not many studies showed a cochlear model which could account for the masker phase effects. Such models were, for example, shown by Oxenham & Dau (2001a); Nishimura (2005) and Shen & Lentz (2009). This type of stimuli thus put a strong constraint on cochlear models

(Oxenham & Dau, 2001a).

Chapter 5

Listening tests

The author of this thesis used the roughness model (described in Chapter 3) to predict the roughness of various types of acoustic stimuli (see Chapter 6). Subjective data of roughness for some of the used acoustic stimuli were reproduced from the literature, the roughness of the remaining stimuli was measured by means of the listening tests conducted within the framework of this thesis. This Chapter describes the listening tests and shows the results.

5.1 Roughness of amplitude-modulated harmonic complexes

5.1.1 Method

Stimuli: The stimuli were harmonic complex tones composed of the first three harmonics ($N = 3$) given by

$$p(t) = [1 + m \cdot \cos(2\pi f_m)] \sum_{n=1}^{N} A(n) \cdot \cos(2\pi n f_0 t), \tag{5.1}$$

where f_m is the modulation frequency, m is the modulation index, $A(n)$ is the amplitude of the harmonics and f_0 is the fundamental frequency of the complexes. The fundamental frequency of the harmonics, f_0, was 300 Hz, the modulation frequency, f_m, was 30, 40, 50, 60, and 70 Hz, the modulation index, m, was set to 0, -3, -6, -9 and -12 dB given

by the relation $20\log_{10}(m)$, and the amplitude, $A(n)$, of the first, second and third spectral component was 0, -10 and -20 dB, respectively. The duration of the stimuli was 600 ms and they were ramped on and off with 30-ms raised-cosine ramps. The level of the stimuli was 75 dB SPL. Combinations of the modulation frequencies and the modulation depths led to 25 different stimuli.

Listeners: Five experienced listeners – four men, age ranging between 25 and 44 years, including the author – participated in the experiment. The listeners had normal hearing: pure-tone thresholds below 20 dB HL for frequencies between 250 Hz and 8 kHz.

Procedure and equipment: The listeners rated the roughness of the stimuli on a discrete scale from 1 to 7 in steps of 1, where 1 was for the lowest and 7 for the highest roughness. All 25 stimuli were presented in random order and the listeners rated each stimulus ten times which gave 250 ratings from each listener. The listeners could hear each stimulus as many times as they desired, assign a roughness rating and then proceed to the next stimulus. The test was conducted on a computer. The stimuli were presented diotically – the same signal to both ears – via Sennheisser HD-600 headphones.

The procedure was inspired by Patel *et al.* (2012) where listeners rated the roughness of pathological voice samples on a 5 point scale. Here, a 7 point scale was used instead. The reason for this was that the just noticeable difference of roughness corresponds to about 10% change of the modulation index, m, of a SAM tone (Fastl & Zwicker, 2007). Five chosen values of the modulation depths (0, -3, -6, -9 and -12 dB) of the SAM complexes should thus cause perceptible changes of the roughness. Moreover, the roughness of the stimuli depends as well on the modulation frequency (Fastl & Zwicker, 2007). Hence, for the SAM complex stimuli, a 5-point scale seemed to be too course.

5.1.2 Results

The listeners rated the roughness of each stimulus ten times, but the first two ratings for each stimulus were not taken into account for the final processing of the results. The intrasubject and intersubject reliability was estimated as Cronbach's alpha calculated from the ratings given by each listener. Cronbach's alpha calculated from the ratings

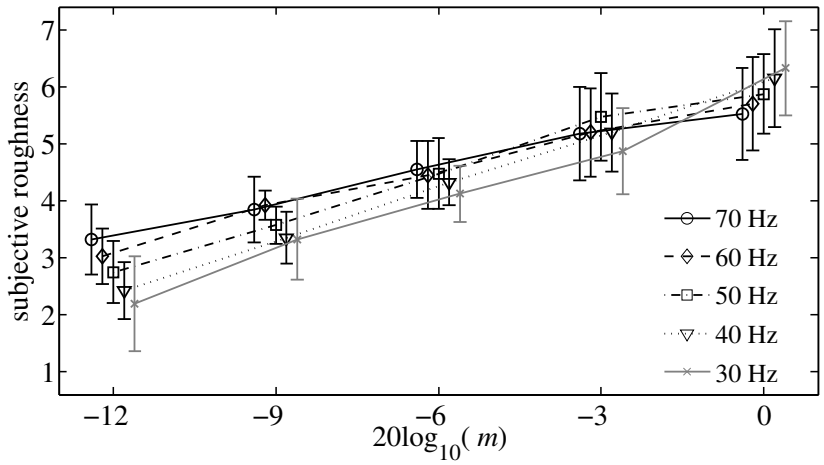

Figure 5.1: Mean values and standard deviations across the listeners of the subjective ratings of roughness for sinusoidally amplitude-modulated (SAM) harmonic complex tones plotted as a function of the modulation index, m.

given by the individual listeners was in all cases higher than 0.8 with 5% level of significance: it means that the listeners were self consistent. Crobnach's alpha calculated from the mean ratings across the listeners was 0.951 with 5% level of significance – there was an agreement between the ratings from individual listeners.

Fig. 5.1 shows the mean values and the standard deviations of the roughness ratings across the listeners. The data are plotted as a function of the modulation index calculated as $20\log_{10}(m)$. Markers connected by lines show the roughness for a specific modulation frequency: circles connected by solid line for the modulation frequency of 70 Hz, diamonds connected by dashed lines for 60 Hz, squares connected by dash-dot lines for 50 Hz, upward triangles connected by dotted lines for 40 Hz and gray crosses connected by gray solid lines for 30 Hz. The data for the SAM complexes with the same modulation index, m, were shifted on the abscissa to be better visible.

5.2 Roughness of synthetic vowels

The synthetic vowels /a/ were generated by the Klatt synthesizer (Klatt, 1980) which was adjusted in order to generate vowels with defined jitter and shimmer. The Klatt synthesizer generates unit impulses which are then filtered by a glottal filter. The temporal positions and amplitudes of the impulses were adjusted which affected jitter

(Jitt) and shimmer (Shim). The jitter of the generated impulses was calculated by

$$\text{Jitt}\% = 100 \frac{\frac{1}{N-1}\sum_{n=1}^{N-1} |T_x(n+1) - T_x(n)|}{\frac{1}{N}\sum_{n=1}^{N} T_x(n)}, \tag{5.2}$$

where T_x is the time difference between the successive adjacent impulses. The shimmer of the generated impulses was calculated by

$$\text{Shim}\% = 100 \frac{\frac{1}{N-1}\sum_{n=1}^{N-1} |A_x(n+1) - A_x(n)|}{\frac{1}{N}\sum_{n=1}^{N} A_x(n)}, \tag{5.3}$$

where A_x is the amplitude of the impulses.

5.2.1 Method

Stimuli: Ten vowels /a/ varying in roughness were generated by means of the Klatt synthesizer (Klatt, 1980). The Klatt synthesizer first generates unit impulses which are then filtered by a glottal filter in order to create a glottal signal. The amplitude of the impulses, A_x and the time difference between the adjacent impulses, T_x was manipulated which affected the jitter (Jitt), shimmer (Shim) calculated by Eq. (5.2) and (5.3). Ten vowels /a/ with different roughness were generated. Nine vowels with a frequency of the glottal pulses – fundamental frequency of the vowels – equal to 125 Hz, and one vowel with a fundamental frequency equal to 63 Hz. The duration of the stimuli was 400 ms and it was ramped on and off with 30-ms raised-cosine ramps. The level of the stimuli was 75 dB SPL.

Listeners: Four normal-hearing experienced listeners participated in the experiment. Their pure-tone hearing thresholds were within a range of 15 dB HL for frequencies between 250 Hz and 8 kHz. The listeners were men aged between 25 and 36 years. The author was among the listeners.

Procedure and equipment: The listeners rated the roughness of the vowels on a discrete 5-point scale from 1 to 5 in steps of 1, where 1 was for the lowest and 5 for the highest roughness. The same scale was used by Patel *et al.* (2012) to measure the roughness of real pathological voice samples of a sustained vowel /a/. The procedure and equipment were the same as in the previous experiment with SAM complex tones

(see Section 5.1). The randomly ordered ten stimuli were rated ten times, giving 100 stimuli per the listening test.

Table 5.1: Synthetic vowels /a/

stimuli	f_0(Hz)	Jitt(%)	Shim(%)	$R_{subj.}$
s_{01}	125	0	0	1.4
s_{02}	125	0	9.7	1.5
s_{03}	125	5	0	1.88
s_{04}	125	0	20	1.91
s_{05}	125	0	33.3	2.8
s_{06}	125	5.1	33.3	3.75
s_{07}	125	12.5	0	4.05
s_{08}	125	9.7	33.3	4.05
s_{09}	125	0	80	4.44
s_{10}	63	0	0	4.7

5.2.2 Results

Each stimulus was rated ten times, but the first two ratings were, as well as in the previous experiment, not taken into account for the final processing of the results. The intrasubject reliability of the obtained roughness ratings estimated as Cronbach's alpha was higher than 0.75 with 5% level of significance for all the listeners – the listeners were self-consistent. The intersubject reliability (Cronbach's alpha) was higher than 0.83 with 5% level of significance which means a good agreement between the ratings from the individual listeners.

Table 5.1 shows the jitter and shimmer calculated by Eq. (5.2) and (5.3) from the generated unit impulses used to synthesize the vowels. The table also shows the subjective roughness (mean values across the listeners), $R_{\text{subj.}}$, of the vowels. The same data are shown in Fig. 5.2.

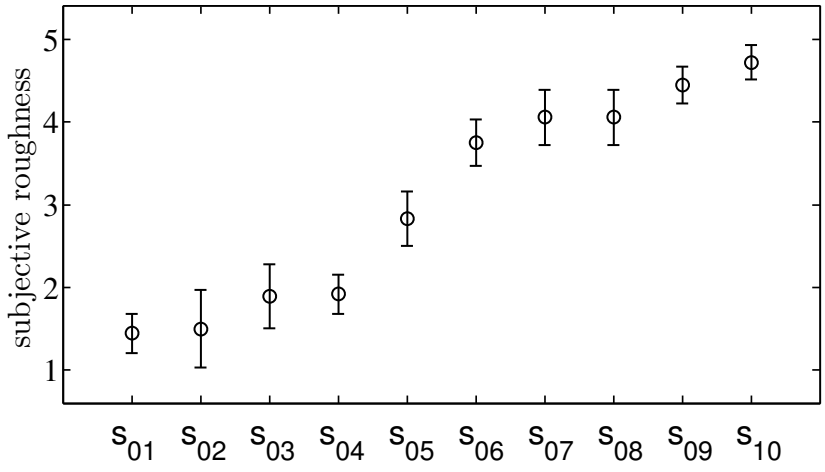

Figure 5.2: Mean values and standard deviations of the subjective ratings of roughness of synthetic vowels /a/.

5.3 Roughness of real vowels

The same method as in the previous experiment was used to rate the roughness of real pathological voice samples – sustained vowels /a/.

5.3.1 Method

Stimuli: 11 real pathological voice samples of a sustained vowel /a/ were used as stimuli. The vowels were extracted from the stimuli recorded from 11 different subjects during the scale singing. The subjects had a pathology affecting their larynx. The stimuli differed in the pitch and in the amount of roughness. The duration of the stimuli was 300 ms and they were ramped on and off with 30-ms raised-cosine ramps. The level of the stimuli was 75 dB SPL.

Listeners: Six experienced listeners – men aged between 25 and 36 years, including the author – participated in the experiment. The listeners had normal hearing: pure-tone thresholds below 20 dB hearing level (HL) for frequencies between 250 Hz and 8 kHz.

Procedure and equipment: Roughness was rated on a discrete 5-point scale from 1 to 5 in steps of 1, where 1 was for the lowest and 5 for the highest roughness. The same method was used by Patel *et al.* (2012) estimating the roughness of the same

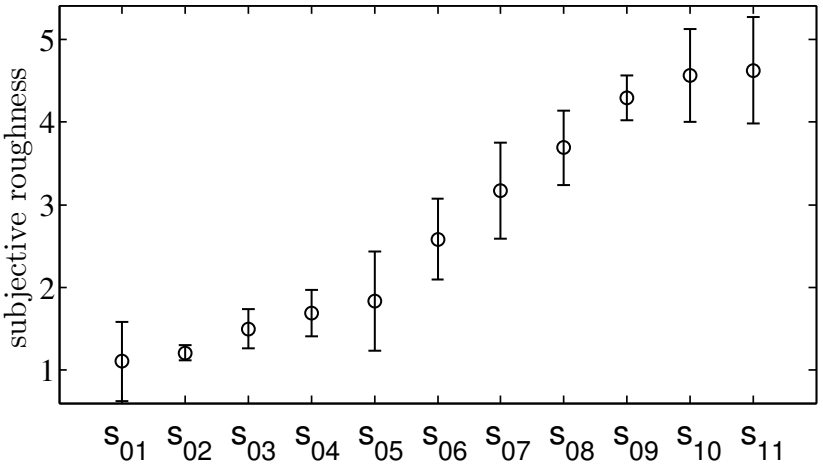

Figure 5.3: Mean values and standard deviations of the subjective ratings of roughness for real vowels /a/. The real voice samples are denoted in the abscissa as s_{01-11}.

type of stimuli – pathological voice samples of a sustained vowel /a/. The procedure and equipment were the same as in the previous two experiments with SAM complexes and synthetic vowels (Section 5.1 and 5.2, respectively). Randomly ordered 11 stimuli were rated 10 times, giving 110 stimuli per the listening test.

5.3.2 Results

Each stimulus was rated ten times, but the first two ratings were not taken into account for the final processing of the results. The intrasubject reliability estimated as Cronbach's alpha was for the listeners higher than 0.9 with 5% level of significance. The intersubject reliability estimated as Cronbach's alpha was 0.983 with 5% level of significance.

Fig. 5.3 shows the mean values and standard deviations from the mean of the subjective ratings of the real voice samples denoted as s_{01--11}.

5.4 Summary

The Chapter described the listening tests conducted by the author within the framework of the thesis. The tests were conducted in order to obtain the roughness ratings for sinusoidally amplitude-modulated (SAM) complex tones, synthetic and real vowels /a/.

The rating method was used in the experiments – listeners rated the roughness on a given discrete scale.

Chapter 6

Prediction of roughness

The roughness model described in Chapter 3 was used to predict roughness of various types of acoustic stimuli. This chapter shows the predicted roughness and compares it with the results of listening tests conducted within the framework of this thesis or reproduced from the literature (Terhardt, 1968, 1974; Kemp, 1982; Aures, 1984; Pressnitzer & McAdams, 1999; Vassilakis, 2005; Vogel, 1975; Miśkiewicz *et al.*, 2006; Fastl & Zwicker, 2007).

6.1 Roughness of sinusoidally amplitude-modulated tones

A sinusoidally amplitude-modulated (SAM) tone is given by

$$p(t) = A\,[1 + m \cdot \cos(2\pi f_m t)]\cos(2\pi f_c t), \tag{6.1}$$

where A is the amplitude, m is the modulation index, f_m is the modulation frequency and f_c is the tone frequency (carrier frequency). SAM tones have been used in a number of listening tests investigating the dependence of roughness on parameters of SAM tones – modulation and tone frequency, modulation index and level (Fastl & Zwicker, 2007).

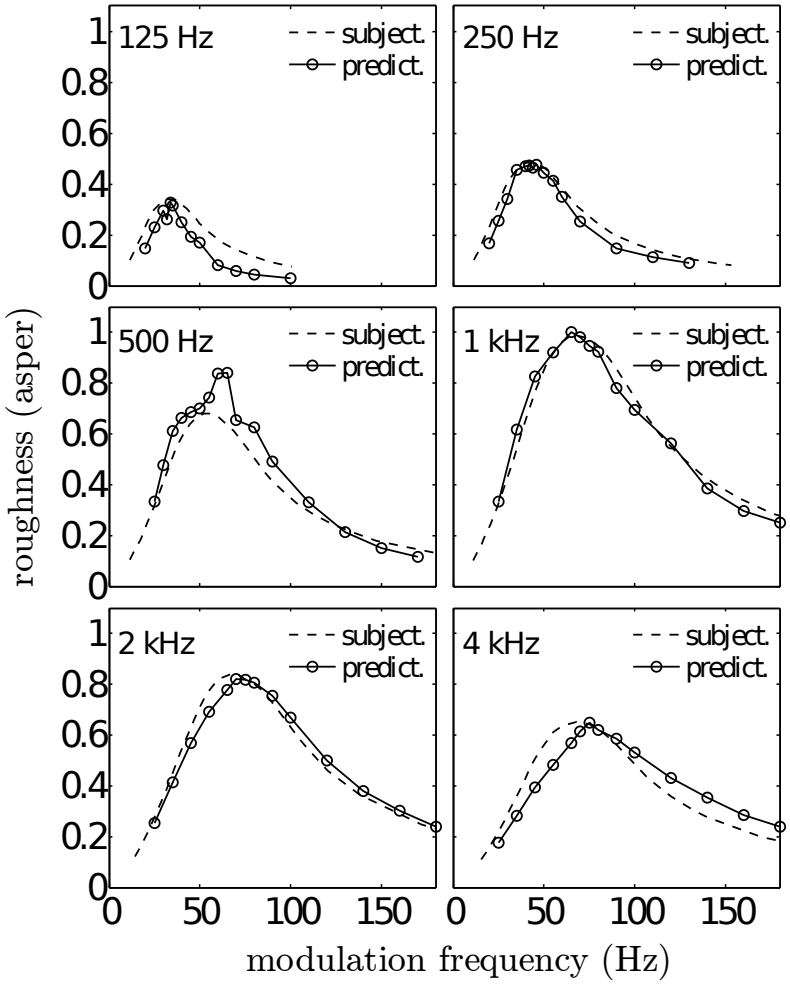

Figure 6.1: Roughness of 100% SAM tones as a function of the modulation frequency. Dashed lines show the subjective data reproduced from Fastl & Zwicker (2007). Circles connected by solid lines show the predicted roughness. The level of SAM tones was 60 dB SPL. Its frequency is given in the upper left corner of each panel.

6.1.1 Dependence on the modulation frequency

Fig. 6.1 shows the roughness of SAM tones plotted as a function of the modulation frequency: each panel shows data for a SAM tone of a specific frequency given in the upper left corner. The level of the SAM tones was 60 dB SPL. Dashed lines in each panel of Fig. 6.1 show the results of listening tests conducted by Aures (1984) and reproduced from Fastl & Zwicker (2007). Circles connected by solid lines show the roughness predicted by the described roughness model.

The subjectively estimated roughness of SAM tones was used to design the roughness model. In other words, parameters of the roughness model were set in order to

quantitatively fit the subjective data showing the dependence of roughness of SAM tones on the modulation frequency (see Section 3.2). However, for the SAM tones with a frequency of 125 Hz, 500 Hz, 4 kHz and 8 kHz, the discrepancies between the data are for some of the modulation frequencies larger than the just noticeable roughness difference of 17% (Fastl & Zwicker, 2007).

A very good agreement between the subjective and predicted roughness of SAM tones showed the Daniel and Weber roughness model described by Daniel & Weber (1997). For the SI roughness model, the predicted roughness shown as a function of the modulation frequency also exhibits a bandpass characteristic. However, the data were not directly compared with subjective roughness (Leman, 2000).

6.1.2 Dependence on the modulation index

Results of the listening tests showed that the dependence of roughness on the modulation index, m, is given by a power-law

$$R \sim m^p, \tag{6.2}$$

where the exponent, p, varies among different perceptual studies: Terhardt (1968) estimated p=2, Vogel (1975) p=1.5 and Fastl & Zwicker (2007) p=1.6. Fig. 6.2 shows the dependence of roughness of a 1 kHz-SAM tone with a level of 70 dB SPL and a modulation frequency of 70 Hz on the modulation index. Dashed, dash-dot and dotted lines in Fig. 6.2 show the roughness given by the relation $R = 1.36 \cdot m^p$, where p was set to 2, 1.5 and 1.6, respectively. This relation was used by Daniel & Weber (1997). It gives the roughness of 1.36 aspers for m=1 which agrees with the subjectively estimated roughness of 100% SAM tone with a frequency of 1 kHz, a level of 70 dB SPL and a modulation frequency of 70 Hz (Terhardt, 1968). Circles connected by the solid line show the predicted roughness. The best agreement between the model predictions and the power-law relations is for the values of m between 0 and 0.8.

The Daniel and Weber roughness model predicted the dependence of roughness on the modulation index in agreement with the subjective data (Daniel & Weber, 1997). Leman (2000) did not show the dependence of the predicted (by the SI roughness model) roughness of SAM tones on the modulation index.

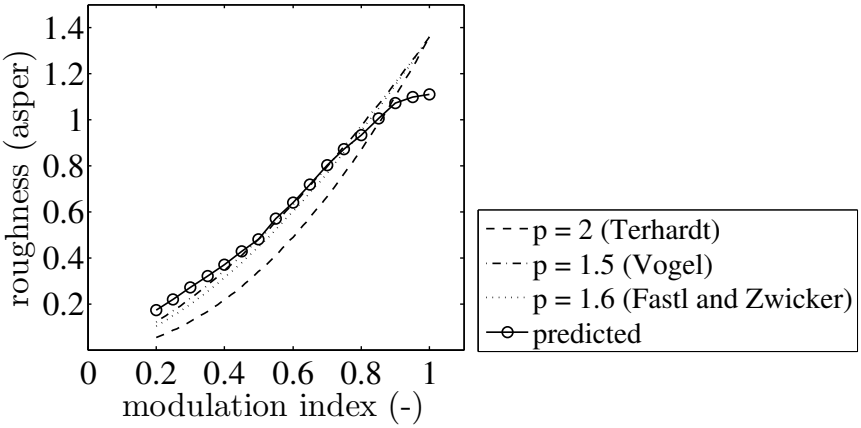

Figure 6.2: Roughness of a SAM tone with a frequency of 1 kHz, a level of 70 dB SPL and a modulation frequency of 70 Hz plotted as a function of the modulation index. The dashed, dashed-dot, and dotted line was obtained by equation $R = 1.36 \cdot m^p$, where m is the modulation index and p equals to 1.6, 2, and 1.5, respectively. The values of p were estimated from the subjective experimental data: Fastl & Zwicker (2007) estimated $p = 1.6$, Terhardt (1968) $p = 2$, and Vogel (1975) $p = 1.5$. The solid line shows the predicted roughness.

6.1.3 Dependence on the level

The dependence of roughness of SAM tones on the level was measured by Terhardt (1968, 1974) using listening tests. Both studies used the same method. Subjects were presented with a pair of stimuli: the first stimulus with a level of 40 or 60 dB SPL and a modulation index, m, of 1; the second stimulus with a level of 80 dB SPL and a modulation index set randomly to a value between 0.2 and 1. The subject's task was to mark a stimulus with more roughness. The results were then shown as the values of the modulation index of an 80-dB SAM tone which was perceived with the same roughness as 40 and 60-dB SAM tones. Fastl & Zwicker (2007) reproduced the results of Terhardt (1968) and showed them as the relative roughness given by the relation $R_\mathrm{r} = 100 \cdot m^{1.6}$ (%), where m is the modulation index of an 80-dB SAM tone estimated by Terhardt (1968).

Fig. 6.3 shows the level dependence of roughness of a 100% SAM tone with a frequency of 1 kHz and a modulation frequency of 70 Hz. Squares connected by dashed lines and crosses connected by dash-dot lines show the data reproduced from Terhardt (1968) and Terhardt (1974), respectively. Both sets of the data were measured binaurally. The SAM tone with higher level (80 dB SPL) was placed after the SAM tone with lower level (40 or 60 dB SPL). Plus signs connected by dotted lines show the data measured by Terhardt (1974). The data were measured monaurally. The SAM tone with

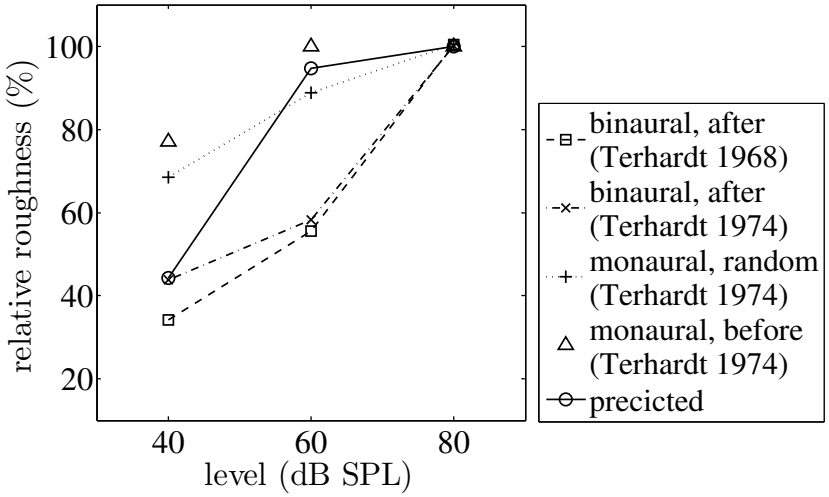

Figure 6.3: Dependence of roughness of a 100% SAM tone with a frequency of 1 kHz on the level. Squares connected by dashed lines, crosses connected by dash-dot lines, plus signs connected by dotted lines and triangles show the subjective data reproduced from Terhardt (1968, 1974). Circles connected by solid lines show the predicted roughness. The data are shown as the relative roughness to the roughness of a SAM tone with a level of 80 dB SPL.

higher level was placed randomly in one of the two intervals. Triangels show the data measured by Terhardt (1974). The data were measured monaurally. The SAM tone with higher level was placed before the SAM tone with lower level. All of the subjective data were in the mentioned studies shown as the values of the modulation index of an 80-dB SAM tone. Fig. 6.3 shows the data as the relative roughness calculated using the aforementioned relation given by Fastl & Zwicker (2007) – $R_\mathrm{r} = 100 \cdot m^{1.6}$.

The predicted roughness is shown in Fig. 6.3 as open circles connected by solid lines. The data are shown as the relative roughness to the predicted roughness of the 80-dB SAM tone. The predicted roughness increases approximately three times for the 40-dB level increment which agrees with the subjective data (squares and crosses in Fig. 6.3). However, the highest increase of the predicted roughness is for the stimuli with levels between 40 and 60 dB SPL, which is not the case for the subjective data shown as squares and crosses.

The Daniel and Weber model predicted the dependence of roughness on the level of SAM tones in agreement with the data measured by Terhardt (1968). The predicted roughness showed approximately threefold increase of the roughness when the SAM tone level was increased from 40 to 80 dB SPL (Daniel & Weber, 1997). Leman (2000) did not show the level dependence of roughness predicted by the SI roughness

model.

6.2 Roughness of two tone stimuli

This Section shows the predicted and subjective roughness of two tone stimuli (dyads). The dyads were composed of pure tones and harmonic complex tones.

6.2.1 Pure tone dyads

Pure-tone dyads are the stimuli composed of two added pure tones as is given by

$$p(t) = A_1 \cos(2\pi f_1 t) + A_2 \cos(2\pi f_2 t), \tag{6.3}$$

where A_1 and A_2 are the amplitudes, and f_1 and f_2 are the frequencies of the pure tones.

Miśkiewicz *et al.* (2006) conducted listening tests to measure the roughness of pure tone dyads: circles connected by dashed lines in Fig. 6.4 show the results of the listening tests. Miśkiewicz *et al.* (2006) measured the roughness by the method of magnitude estimation. The listeners assigned a number to the perceived stimulus according to its roughness. Each panel in Fig. 6.4 shows the roughness of pure tone dyads with a specific center frequency (given in the upper left corner), abscissa shows the beat rate which equals the frequency difference between the tones. The roughness was measured using pure tone dyads with a center frequency of 0.125, 0.25, 0.5, 1, 2, and 4 kHz and a level of 67, 56, 49, 46, 50, 48, and 43 dB SPL, respectively.

Squares connected by solid lines in Fig. 6.4 show the predicted roughness. The data were scaled to give the maximal predicted roughness of 1 for values in the respective panels. The predicted roughness, as well as the subjective roughness, exhibits a bandpass characteristic with maximum which shifts to higher frequencies as the dyad center frequency increases. However, the predicted and the subjective roughness differs especially for higher beat rates. The best agreement between the predicted and the subjective data was reached for the dyad with a center frequency of 1 kHz.

Daniel & Weber (1997) and Leman (2000) did not show the predicted roughness of two tone stimuli composed of two pure tones for the Daniel and Weber roughness

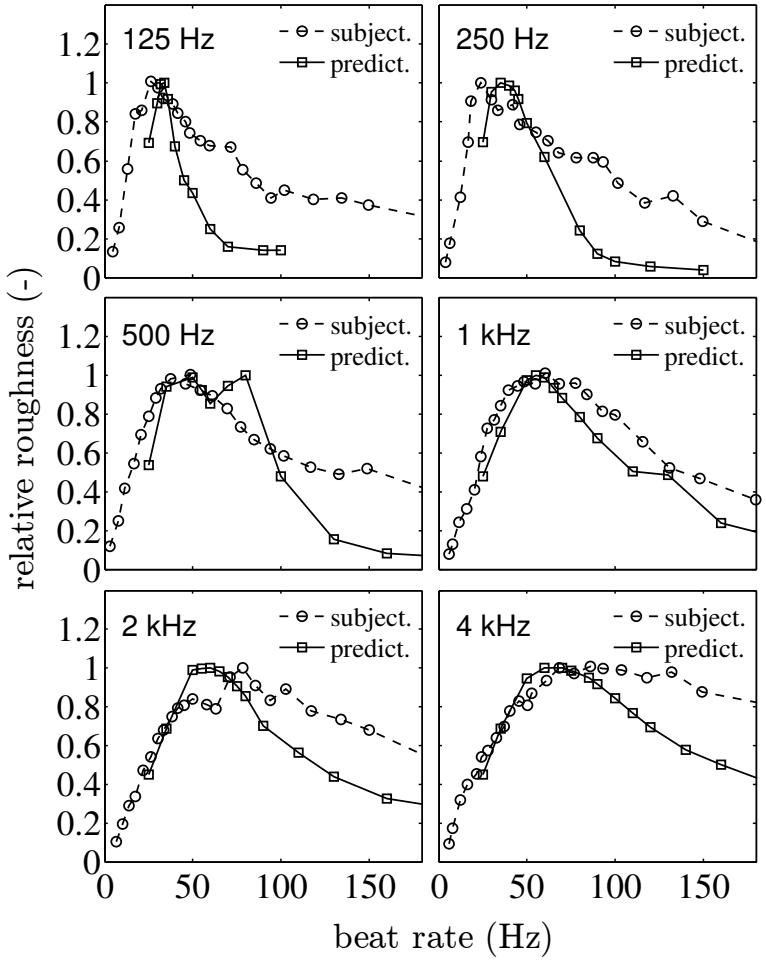

Figure 6.4: Roughness of two tone stimuli composed of pure tones as a function of the frequency difference. Circles connected by dashed lines show the subjective data reproduced from Miśkiewicz *et al.* (2006). Squares connected by solid lines show the predicted roughness. The center frequency of the pure tone dyads is given in the upper left corner of each panel, the level of the stimuli was 67, 56, 49, 46, 50, 48, and 43 dB SPL, respectively for increasing values of the center frequencies.

model and the SI roughness model, respectively.

6.2.2 Dyads of harmonic complex tones – intervals of the chromatic scale

Vassilakis (2005) conducted listening tests to measure the roughness of dyads composed of two harmonic complexes with fundamental frequencies set to create the intervals of

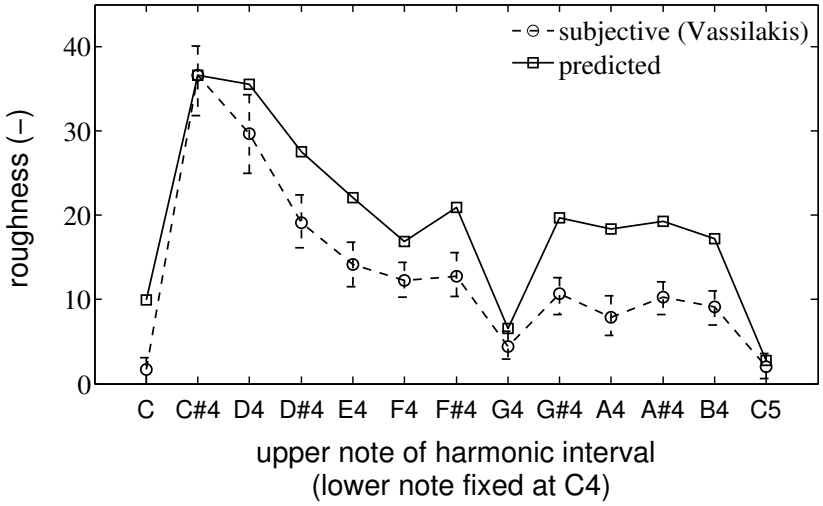

Figure 6.5: Roughness ratings of harmonic intervals of the chromatic scale constructed from the harmonic complex tones. Circles connected by dashed lines show the mean values and the standard deviations of the subjective data across ten listeners (reproduced from Vassilakis (2005)). Squares connected by solid lines show the predicted roughness normalized by its maximal value and scaled by the maximal value of the subjective roughness to be in the range of the used rating scale.

the chromatic scale. The dyads were given by

$$p(t) = \sum_{n=1}^{N} \frac{A}{n} \cos(2\pi n f_{0_1} t) + \sum_{n=1}^{N} \frac{A}{n} \cos(2\pi n f_{0_2} t), \tag{6.4}$$

where A is the amplitude of the first spectral component in each harmonic complexes; f_{0_1} and f_{0_2} are the fundamental frequencies of the first and the second harmonic complexes, respectively; and N is the number of harmonics which was set to 6. The fundamental frequency of the lower harmonic complexes, f_{0_1}, was set to middle C (C4, fundamental frequency 256 Hz, equal temperament). The level of the dyads was 75 dB SPL.

Ten listeners rated the roughness of the dyads on a continuous scale ranging from 0 (not rough) to 42 (rough) (Vassilakis, 2005). Fig. 6.5 (circles connected by dashed lines) shows the mean values and the standard deviations of the roughness ratings calculated across responses from ten listeners. The abscissa shows a frequency of the higher tone in the dyads. The predicted roughness of the dyads is shown as squares connected by solid lines. The predicted data were normalized by its maximal value and then scaled by the maximal value of the subjective roughness in order to visualize the predicted and the subjective data in the same graph. The predicted data agree with the subjective data: Spearman's correlation coefficient $r = 0.92, p = 0$; Pearson's correlation coefficient $r = 0.94, p = 1.5 \cdot 10^{-6}$.

Vencovský (2014c) processed the dyads composed of harmonic complex tones by the Daniel and Weber roughness model (Daniel & Weber, 1997) and the SI roughness model (Leman, 2000). The Daniel and Weber roughness model was implemented in the PsySound3 sound analyses software (PsySound3, 2008) and the SI roughness model in the IPEM toolbox (IPEM, 2003). The roughness predicted using the Daniel and Weber model did not agree well with the subjective data: Spearman's correlation coefficient $r = 0.224, p = 0.46$; Pearson's correlation coefficient $r = 0.649, p = 0.016$. The roughness predicted using the SI roughness model gave better predictions than the Daniel and Weber roughness model. However the agreement was worse than for the presented roughness model: Spearman's correlation coefficient: $r = 0.852, p = 3.4 \cdot 10^{-5}$; Pearson's correlation coefficient: $r = 0.846, p = 2.7 \cdot 10^{-4}$.

6.3 Roughness of stimuli with envelopes that are not sinusoidal

Mathes & Miller (1947) showed that the stimuli with equal amplitude spectrum and different phase spectrum can have different roughness. This phenomenon was then later studied by Pressnitzer & McAdams (1999) with stimuli whose time envelopes were not sinusoidal – pseudo amplitude-modulated (pAM) tones and stimuli with asymmetrical temporal envelopes. Pressnitzer & McAdams (1999) conducted listening tests to measure the roughness of these stimuli and then analyzed the stimuli of different roughness by a time-domain model of cochlear frequency selectivity. They showed that the stimuli of different roughness may have equal root mean square (RMS) values but different shape of the signal envelope after cochlear filtration. Therefore they advised that the shape of the envelope should be taken into account when the roughness is predicted. The roughness model described in this thesis allows to take into account the shape of the envelope after auditory filtering.

6.3.1 Pseudo amplitude-modulated tones

A 100% SAM tone modulated with modulation frequency f_m is composed of three spectral components: the central component with frequency f_c (frequency of the modulated tone) and amplitude A; and two side components with frequencies $f_c - f_m$ and $f_c + f_m$, and amplitudes equal to $A/2$. Setting the starting phases of the side

components to zero and adjusting the starting phase of the central component, ϕ, creates so called pseudo amplitude-modulated (pAM) tones. The pAM tones are thus given by

$$p(t) = A\cos(2\pi f_c t + \phi) + \frac{A}{2}\cos\left[2\pi(f_c \pm f_m)t\right]. \tag{6.5}$$

If ϕ is nonzero, the time waveform envelope of the pAM tone is not sinusoidal. If ϕ of two pAM tones is opposite (e.g. $-\pi/6$ and $\pi/6$), the pAM tones have the same waveform envelope but different temporal fine structure (Pressnitzer & McAdams, 1999).

Pressnitzer & McAdams (1999) conducted listening tests to estimate the roughness of pAM tones. They presented the listeners with a pair of stimuli and asked them to judge which stimulus has more roughness. Pressnitzer & McAdams (1999) then transformed the judgments into a linear interval scale of the roughness by means of the Bradley-Terry-Luce (BTL) method (David, 1988). The standard deviations of the roughness data were estimated by the bootstrap technique (Pressnitzer & McAdams, 1999). Fig. 6.6 (the panels with gray lines) shows the results of the listening test. Each panel shows roughness of a pAM tone with frequency f_c and modulation frequency f_m given in the upper right corner. The roughness is shown as a function of the starting phase absolute value, $|\phi|$: circles connected by solid lines show the roughness of pAM tones with negative values of ϕ; and crosses connected by dashed lines show the roughness of pAM tones with positive values of ϕ.

The predicted roughness is shown in Fig. 6.6 as black lines and symbols: the panels show the roughness of pAM tones with the same parameters as had the pAM tones used to get the subjective data plotted in the panels above them. The predicted roughness was, in order to better visualize the data, normalized by the maximal value of the data for each panel such that the individual data are not higher than 1. Since the subjective roughness data are on an interval scale and the predicted roughness data on a ratio scale, the subjective and predicted roughness cannot be compared neither quantitatively nor qualitatively. Instead of that, the data can be compared as ranking scale data. In other words, only the agreement in ranking of the data is important. The roughness model is sensitive to the phase changes between the spectral components of the pAM tones. The predicted data show the same tendency as the subjective data – decrease of the roughness difference for positive and negative value of ϕ at frequencies f_c above 1 kHz. However, there are also discrepancies between the subjective and predicted data: the most obvious are at f_c of 125 Hz and 500 Hz, where the roughness model predicted higher roughness for $\phi = -\pi/6$ than for $\phi = +\pi/3$.

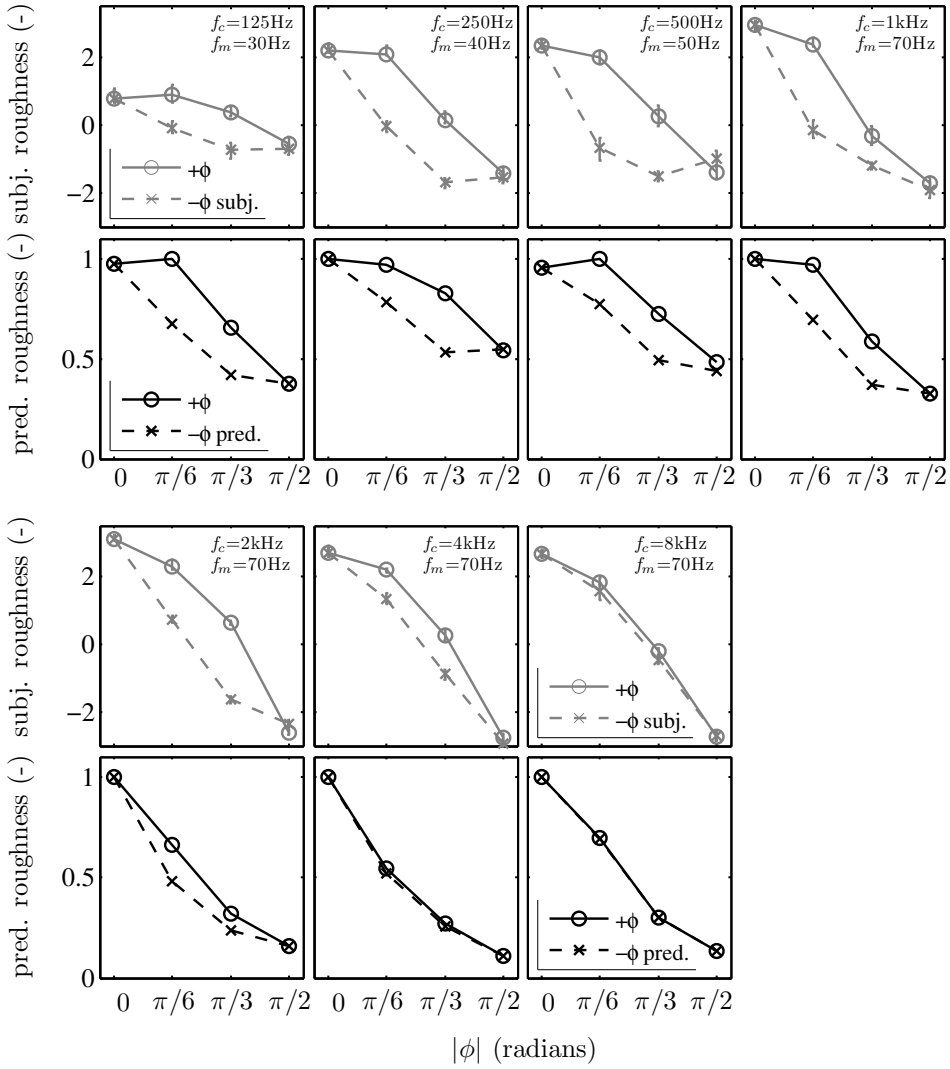

Figure 6.6: Subjective and predicted roughness of pAM tones as a function of the starting phase absolute value, $|\phi|$, (dashed and solid lines show the data for negative and positive values of ϕ, respectively). The subjective data plotted as gray lines and markers were reproduced from Pressnitzer & McAdams (1999). The predicted roughness is plotted as black lines and markers. The predicted data shown in each panel were normalized by its maximal value. The frequency and the modulation frequency of each pAM tone is shown in the upper right corner of the panels showing the subjective data. The predicted data shown in the panels were obtained for the pAM tones of the same parameters as had the pAM tones used to measure the subjective data in the panels placed above them.

The roughness model is sensitive to the phase of the spectral components mainly because of the algorithms used in the central stage (see Chapter 3). The rising parts of the envelope of the tones with the positive ϕ values processed by the peripheral stage of the roughness model are shorter than in the case of the negative ϕ values. The central stage thus estimates higher value of F_{sat} parameter (see Eq. (3.20) in Chapter 3) and, in turn, predicts more roughness. The similar observation was done by Pressnitzer & McAdams (1999) after they processed the pAM tones by a model of cochlear frequency selectivity.

The Daniel and Weber roughness model and SI roughness model cannot account for the effect of the sign of ϕ on the roughness of pAM tones (Kohlrausch *et al.*, 2005; Leman, 2000).

6.3.2 Sawtooth and reversed stimuli

Pressnitzer & McAdams (1999) showed the effect of the shape of the waveform envelope on the perceived roughness of so called "sawtooth" and "reversed" stimuli (also called "ramped" and "damped", respectively). The "sawtooth" stimuli are amplitude-modulated tones given by

$$p_{st}(t) = \left(1 + m\frac{E_{st}(t)}{\max\{E_{st}(t)\}}\right)\cos\left(2\pi f_c t - \frac{\pi}{2}\right), \tag{6.6}$$

where f_c is the frequency of the stimuli, m is the modulation index and $E_{st}(t)$ is the modulation signal which is a harmonic complex tone given by

$$E_{st}(t) = \sum_{n=1}^{N}\frac{1}{n}\cos\left(2\pi n f_m t - \frac{\pi}{2}\right), \tag{6.7}$$

where f_m is the modulation frequency and N is the number of harmonics. The value of N is set to fulfill the condition $N \cdot f_m \leq 0.5B(f_c)$, where $B(f_c)$ is the psychophysically estimated equivalent rectangular bandwidth (ERB) of the cochlear filter with a characteristic frequency equal to f_c given by Eq. (3.4). The equation was taken from Moore & Glasberg (1996). The condition ensures that the spectral components of the harmonic complex tone E_{st} are within a range of one critical band. The "reversed" stimuli are time reversals of the "sawtooth" stimuli and can be created by inverting a sign of the phase shift in the argument of the cosine functions in Eq. (6.6) and Eq. (6.7) to $+\pi/2$.

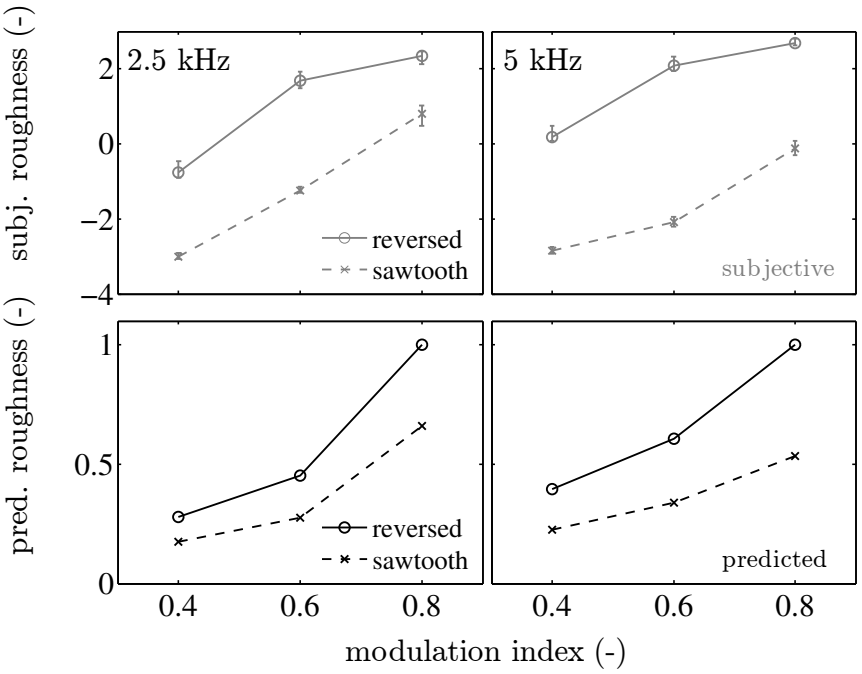

Figure 6.7: Roughness of tones with asymmetrical temporal envelopes as a function of the modulation index. The top panels show the subjective roughness of "sawtooth" (crosses connected by dashed lines) and "reversed" (circles connected by solid lines) stimuli. The data show the mean values and standard errors from the mean and were reproduced from Pressnitzer & McAdams (1999). The bottom panels show the predicted roughness for the same stimuli as in the top panels. The data were normalized by the corresponding maximal value which gave the maximal predicted roughness (shown in each panel) equal to 1. The data were obtained for "sawtooth" and "reversed" stimuli with $f_c = 2.5\,\mathrm{kHz}$, $f_m = 70\,\mathrm{Hz}$ and $N = 2$ (the left panels); and $f_c = 5\,\mathrm{kHz}$, $f_m = 70\,\mathrm{Hz}$ and $N = 4$ (right panels). The level of the stimuli was 60 dB SPL.

Fig. 6.7 shows the roughness of the "sawtooth" (crosses connected by dashed lines) and "reversed" (circles connected by solid lines) stimuli as a function of the modulation index, m. The subjective data reproduced from Pressnitzer & McAdams (1999) are shown as gray lines and markers in the top panels. The data were obtained by means the method of pair wise comparisons and the BTL method – the same method as for the pAM tones (Pressnitzer & McAdams, 1999). The panels in the bottom row show the normalized predicted roughness. The predicted roughness data were normalized such that the plotted data in each panel are not higher than 1. The subjective and predicted roughness data in the left panels of Fig. 6.7 show the roughness of the stimuli with $f_m = 70\,\mathrm{Hz}$, $f_c = 2.5\,\mathrm{kHz}$ and $N = 2$. The right panels show the roughness of the stimuli with $f_m = 70\,\mathrm{Hz}$, $f_c = 5\,\mathrm{kHz}$ and $N = 4$. The level of the stimuli was 60 dB SPL.

The results show that the roughness model can account for the effect of the shape of the waveform envelope. However, the "sawtooth" stimulus with frequency $f_c = 5\,\text{kHz}$ and $m = 0.8$ was perceived to be less rough than the "reversed" stimulus with $m = 0.4$. The same discrepancy is between the subjective and predicted data for the stimuli with $f_c = 2.5\,\text{kHz}$ where the "sawtooth" stimulus with $m = 0.8$ was perceived to be less rough than the "reversed" stimulus with $m = 0.6$.

The Daniel and Weber roughness model (Daniel & Weber, 1997) and the SI roughness model (Leman, 2000) cannot account for the effect of the shape of the waveform envelope on roughness (Kohlrausch *et al.*, 2005; Leman, 2000).

6.4 Roughness of frequency-modulated tones

Sinusoidally frequency-modulated (SFM) tones are given by

$$p(t) = A \cdot \sin\left[2\pi f_c t - \frac{\Delta f}{f_m}\cos(2\pi f_m t)\right], \tag{6.8}$$

where A is the amplitude, f_c is the tone frequency, Δf is the frequency deviation and f_m is the modulation frequency. Kemp (1982) conducted listening tests to estimate the roughness of SFM tones: he measured the dependence of roughness of SFM tones on the modulation frequency, f_m, and on the frequency deviation, Δf. The roughness was measured by the method of magnitude estimation. A pair of stimuli, a standard and a comparison – stimulus under test – was presented to a listener. The listener was given a number reflecting the roughness of the standard and was asked to assign a number reflecting the roughness of the comparison relative to the roughness of the standard (Kemp, 1982).

6.4.1 Dependence on the modulation frequency

Fig. 6.8 shows the roughness of SFM tones as a function of the modulation frequency, f_m. Circles connected by dashed lines show the medians and quartiles of the listening test results reproduced from Fig. 1 in Kemp (1982). Kemp (1982) measured the roughness by the method of magnitude estimation with a SFM tone at a frequency of 1.6 kHz, a modulation frequency of 70 Hz, a frequency deviation of 800 Hz and a level of 60 dB SPL used as a standard. Squares connected by solid lines show the predicted

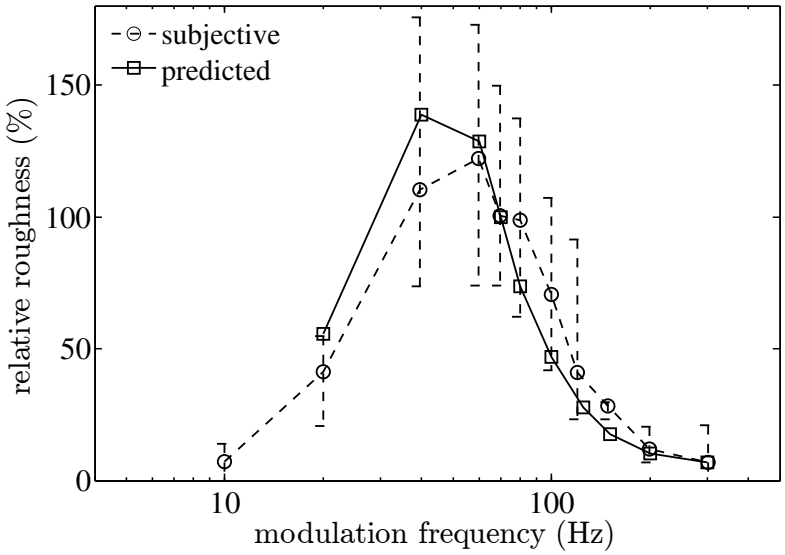

Figure 6.8: Dependence of roughness of SFM tones on the modulation frequency, f_m. Crosses connected by dashed lines show medians and quartiles of the subjective data reproduced from KEMP (1982). Circles connected by solid lines show the roughness predicted by the roughness model. The frequency of the SFM tones was 1.6 kHz, the modulation index, Δf, was 800 Hz and its level was 60 dB SPL. The data are shown as the relative roughness to the roughness of the SFM tone with a modulation frequency of 70 Hz

roughness. Both, subjective and predicted, data are shown as the relative roughness to the roughness of a SFM tone with a modulation frequency of 70 Hz. The SFM tones had a frequency, f_c, of 1.6 kHz, a level of 60 dB SPL and a frequency deviation (modulation index), Δf, of 800 Hz.

The dependence of roughness of the SFM tones on the modulation frequency exhibits a bandpass characteristic. This is similar to the dependence of roughness of SAM tones and SAM noise stimuli on the modulation frequency (FASTL & ZWICKER, 2007). The comparable bandpass characteristic shows also the predicted roughness: the subjective and predicted data agree within a range of quartiles.

The Daniel and Weber roughness model predicted the dependence of roughness on the modulation frequency for SFM tones in agreement with the subjective data (DANIEL & WEBER, 1997). LEMAN (2000) did not show the SI roughness model performance using SFM tones.

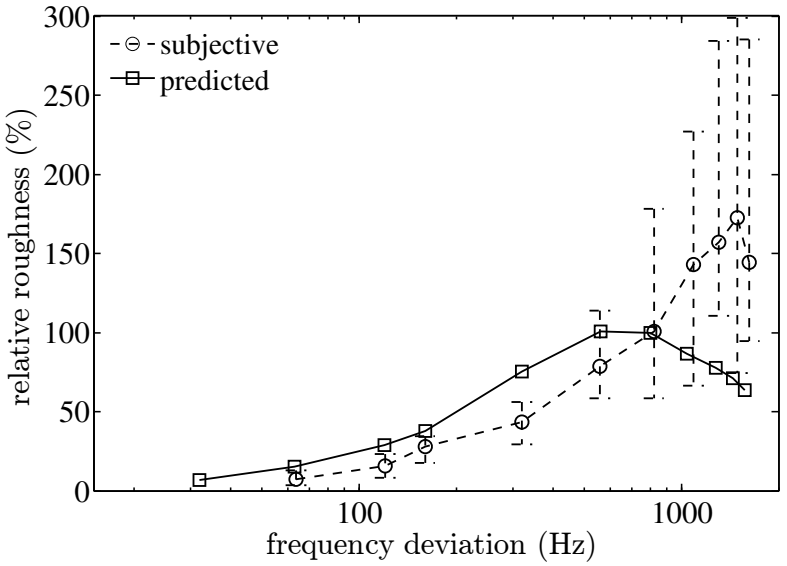

Figure 6.9: Dependence of the relative roughness of SFM tones on the frequency deviation (modulation index), Δf. Crosses connected by dashed lines show medians and quartiles of the subjective data reproduced from Kemp (1982). Circles connected by solid lines show comparable data obtained by means of the roughness model. The frequency of the SFM tone was 1.6 kHz, the modulation frequency was 70 Hz and the level was 60 dB SPL. The data are shown as the relative roughness to the roughness of a SFM tone with a frequency deviation of 800 Hz.

6.4.2 Dependence on the frequency deviation

Fig. 6.9 shows the dependence of roughness of SFM tones on the frequency deviation (modulation index), Δf. Circles connected by dashed lines show medians and quartiles of the roughness of SFM tones reproduced from Fig. 3 in Kemp (1982). Kemp (1982) used the method of magnitude estimation with a 100% SAM tone at a frequency of 1.6 kHz, a modulation frequency of 70 Hz and a level of 60 dB SPL used as a standard. The predicted roughness of the SFM tones is shown in Fig. 6.9 as squares connected by solid lines. Both, subjective and predicted, data show the relative roughness to the roughness of a SFM tone with a frequency deviation, Δf, of 800 Hz. The SFM tones had a frequency, f_c, of 1.6 kHz, a level of 60 dB SPL and a modulation frequency, f_m, of 70 Hz.

The subjective roughness of SFM tones increases as the frequency deviation, Δf, increases. The similar increase shows the predicted roughness. However, qualitative agreement between the predicted and the subjective data is only for the frequency deviations up to 800 Hz. The predicted roughness for $\Delta f > 800$ Hz decreases as Δf increases. For the highest values of Δf, the predicted roughness is out of the range of

quartiles of the subjective roughness.

The Daniel and Weber roughness model predicted the dependence of roughness of SFM tones on the modulation frequency in a good agreement with the subjective data (Daniel & Weber, 1997). Leman (2000) did not show the SI model performance for SFM tones.

6.5 Roughness of unmodulated bandpass noise

Aures (1985) conducted listening tests to estimate the roughness of unmodulated bandpass noise stimuli of various bandwidths and center frequencies. He used the method of adjustment. The listener's task was to adjust the modulation index of a SAM tone with a frequency of 1 kHz, a level of 70 dB SPL and a modulation frequency of 70 Hz in order to set the perceived roughness of the SAM tone to be equal to the perceived roughness of unmodulated bandpass noise (stimuli under test). The level of the unmodulated bandpass noise was 70 dB SPL. Aures (1985) (Fig. 6) expressed the obtained data – the values of the modulation index of the 1-kHz SAM tone – as a function of the bandwidth of unmodulated bandpass noise. Daniel & Weber (1997) transformed these data to aspers by the relation $R = 1.36 \cdot m^{1.6}$, where m is the measured modulation index (reproduced from Aures (1985)) of the 1-kHz SAM tone.o

Fig. 6.10 shows the roughness of unmodulated bandpass noise stimuli. Circles connected by dashed lines show the medians and quartiles of the subjective roughness reproduced from Daniel & Weber (1997). Squares connected by solid lines show the medians and quartiles of the predicted roughness. The data were predicted from ten realizations of each stimulus. Each panel shows the roughness of unmodulated bandpass noise of a specific center frequency (given in the upper part of each panel) – of 0.25, 1 and 4 kHz. The level of the unmodulated bandpass noises was 60 dB SPL and the bandwidth is shown in the abscissa of the graphs. A good agreement between the subjective and predicted data was reached only for the 4-kHz bandpass noise (see the bottom panel of Fig. 6.10). Since a lot of natural sounds contain noise, this issue should be studied in the future research.

The Daniel and Weber roughness model predicted the roughness of the unmodulated bandpass noise in a good agreement with the subjective data (Daniel & Weber, 1997). Leman (2000) did not show the SI roughness model performance for these stimuli.

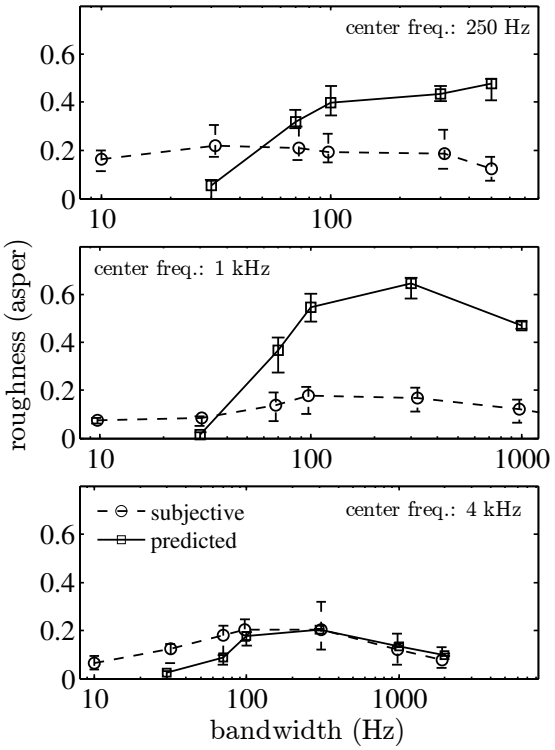

Figure 6.10: Roughness of unmodulated bandpass noise stimuli as a function of the bandwidth. Circles connected by dashed lines show the medians and quartiles of the subjective data reproduced from Daniel & Weber (1997). Squares connected by solid lines show the medians and quartiles of the predicted roughness calculated from ten realizations of the stimuli. Abscissa shows the noise bandwidth. Each panel shows the roughness of the unmodulated bandpass noise with a center frequency, f_c, given in the upper corner and a level of 70 dB SPL.

6.6 Roughness of sinusoidally amplitude-modulated harmonic complexes

The author of this thesis conducted listening tests to measure the roughness of sinusoidally amplitude-modulated (SAM) complex tones (see Section 5.1 in Chapter 5). The SAM complexes described in Section 5.1 are in this Section processed by the roughness model and the predicted roughness is compared with the listening test results shown in Fig. 5.1, Section 5.1.

Fig. 6.11 shows the mean values and standard deviations of the roughness ratings across all listeners – the same data as in Fig. 5.1. The data are plotted as a function of the predicted roughness in aspers. The predicted roughness data are in a good agreement with the subjective roughness data: Spearman's correlation coefficient

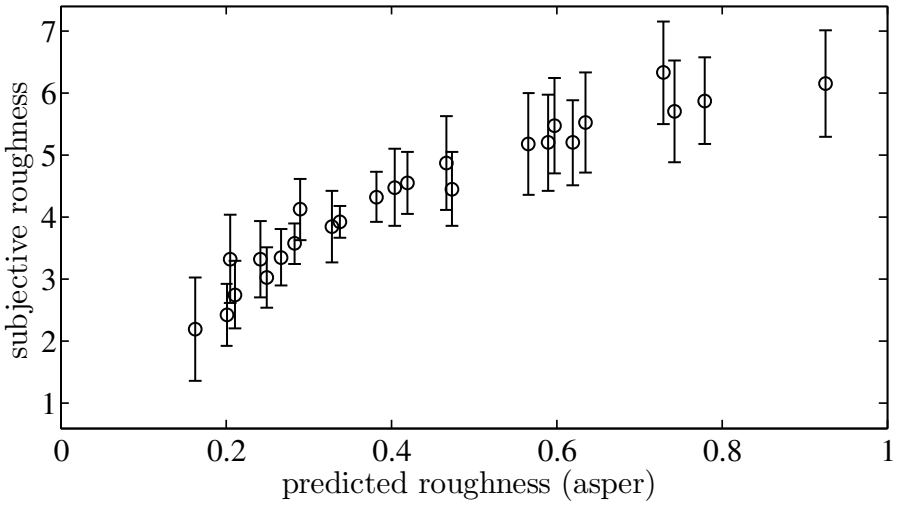

Figure 6.11: Mean values of the subjective ratings of the roughness of sinusoidally amplitude-modulated (SAM) harmonic complex tones as a function of the predicted roughness in aspers. The subjective roughness ratings were obtained by the listening test described in Section 5.1, Chapter 5.

$r = 0.98$, $p = 2.45 \cdot 10^{-18}$; and Pearson's correlation coefficient $r = 0.95$, $p = 2.1 \cdot 10^{-13}$.

Vencovský (2014c) processed the SAM complex tones by the Daniel and Weber roughness model (Daniel & Weber, 1997) and the SI roughness model (Leman, 2000). The Daniel and Weber roughness model was implemented in the PsySound3 sound analyses software (PsySound3, 2008) and the SI roughness model in the IPEM toolbox (IPEM, 2003). Both models process acoustic stimuli and predict the roughness of signals in short time frames. Medians over the time frames were calculated and taken as the resulting roughness of each stimulus (Vencovský, 2014c). The roughness predicted by both roughness models did agree very well with the subjective data. For the Daniel and Weber roughness model: Spearman's correlation coefficient $r = 0.961, p = 2.3 \cdot 10^{-14}$; Pearson's correlation coefficient $r = 0.863, p = 2.8 \cdot 10^{-14}$. For the SI roughness model: Spearman's correlation coefficient: $r = 0.977, p = 6.2 \cdot 10^{-18}$; Pearson's correlation coefficient: $r = 0.956, p = 9.4 \cdot 10^{-14}$.

6.7 Roughness of synthetic vowels

Section 5.2 in Chapter 5 describes a listening test and its results which represent the roughness of synthetic vowels /a/ generated by the Klatt synthesizer (Klatt, 1980). The subjective roughness of the synthetic vowels /a/ is in this Section compared with

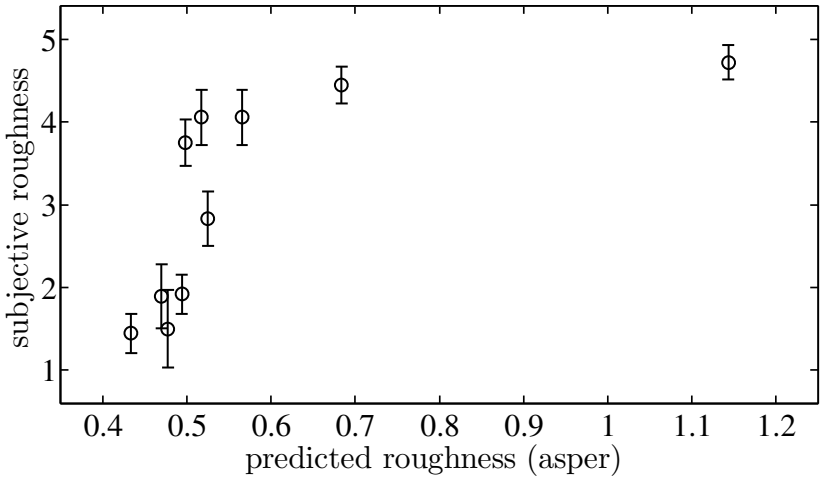

Figure 6.12: Mean values and standard deviations of the subjective ratings of roughness of synthetic vowels /a/ plotted as a function of the predicted roughness. The subjective roughness ratings were obtained by the listening test described in Section 5.2, Chapter 5.

predictions of the roughness model. All details about the stimuli are given in Section 5.2.

Table 6.1 shows the jitter (Jitt) and shimmer (Shim) calculated from the generated unit impulses used to synthesize the vowels (see Section 5.2); the subjective roughness, $R_{\text{subj.}}$, (calculated as the mean values across the listeners ratings); and the predicted roughness, $R_{\text{pred.}}$, of the vowels. The predicted roughness is shown in aspers. The mean values of the subjective ratings across the listeners together with the standard deviations are shown in Fig. 6.12. The data are plotted as a function of the predicted roughness in aspers. The predicted roughness is in a good agreement with the subjective data: Spearman's correlation coefficient $r = 0.94$, $p = 4.5 \cdot 10^{-5}$; and Pearson's correlation coefficient $r = 0.65$, $p = 0.04$.

The author processed the synthetic vowels by the Daniel and Weber roughness model (Daniel & Weber, 1997) and the SI roughness model (Leman, 2000). The Daniel and Weber roughness model was implemented in the PsySound3 sound analyses software (PsySound3, 2008) and the SI roughness model in the IPEM toolbox (IPEM, 2003). The stimuli were processed in the same way as is given above in Section 6.6. The agreement between the predicted and subjective data was worse than for the roughness model described in the thesis. For the Daniel and Weber roughness model: Spearman's correlation coefficient $r = 0.63, p = 0.051$; Pearson's correlation coefficient $r = 0.594, p = 0.07$. For the SI roughness model: Spearman's correlation coefficient: $r = 0.839, p = 2.4 \cdot 10^{-3}$; Pearson's correlation coefficient: $r = 0.754, p = 0.01$.

Table 6.1: Subjective and predicted roughness of the synthetic vowels /a/.

stimuli	f_0(Hz)	Jitt(%)	Shim(%)	$R_{subj.}$	$R_{pred.}$(asper)
s_{01}	125	0	0	1.4	0.43
s_{02}	125	0	9.7	1.5	0.48
s_{03}	125	5	0	1.88	0.47
s_{04}	125	0	20	1.91	0.49
s_{05}	125	0	33.3	2.8	0.53
s_{06}	125	5.1	33.3	3.75	0.50
s_{07}	125	12.5	0	4.05	0.52
s_{08}	125	9.7	33.3	4.05	0.57
s_{09}	125	0	80	4.44	0.68
s_{10}	63	0	0	4.7	1.14

6.8 Roughness of real vowels

Section 5.3 in Chapter 5 describes a listening test and its results – the roughness of real vowels /a/ recorded during scale signing from subjects with pathology on their larynx. The subjective roughness of real vowels /a/ is compared with the predictions of the roughness model. Details about the stimuli (real vowels /a/) are given in Section 5.3.

Fig. 6.13 shows the mean values and standard deviations from the mean of the subjective roughness ratings, $R_{subj.}$, as a function of the predicted roughness, $R_{pred.}$, in aspers. The agreement between the subjective and the predicted data is poor: Spearman's correlation coefficient $r = 0.63$, $p = 0.03$; and Pearson's correlation coefficient $r = 0.52$, $p = 0.08$.

One possible explanation for the poor performance of the roughness model could be that the voice samples were not only rough but also breathy. Speech synthesizers, for example (Klatt, 1980), simulates breathiness by addition of the noise to the generated glottal signal. Section 6.5 showed that the roughness model performs poor for the unmodulated bandpass noise stimuli. Hence, a listening test which gave the ratings of breathiness of the real vowels was conducted.

The listening test was conducted with four normal hearing (pure tone thresholds within

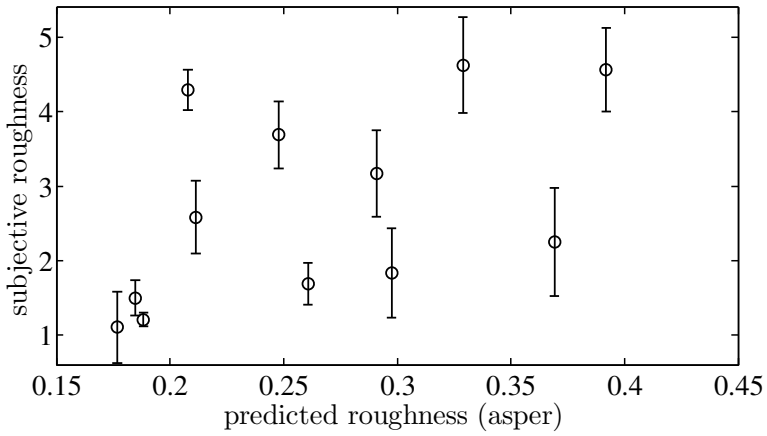

Figure 6.13: Mean values and standard deviations of the subjective ratings of roughness of real vowels /a/ as a function of the predicted roughness. The predicted roughness is shown as relative roughness to the roughness of 100% SAM tone with a frequency of 1 kHz, a level of 60 dB SPL and a modulation frequency of 70 Hz.

a range of 15 dB HL between 0.25 and 8 kHz) experienced listeners aged between 26 and 37 years – men, including the author. The procedure and equipment was same as for the roughness listening test (see Section 5.3). The listeners were asked to rate breathiness on a 5-point discrete scale: 1 for the lowest breathiness, 5 for the highest breathiness. Each stimulus was rated ten times giving 110 ratings per the test. The stimuli were presented in random order. Intrasubject reliability was estimated by Cronbach's alpha. It was for all the listeners higher than 0.807 with 5% level of significance. Intersubject reliability estimated as Cronbach's alpha was 0.854 with 5% level of significance. The breathiness data are, together with the subjective and predicted roughness data, summarized in Table 6.2.

Fig. 6.14 shows the subjective roughness ratings as a function of the predicted roughness ratings (same as in Fig. 6.13) – only the stimuli with breathiness below 3 were taken into account. This improved the model performance: Spearman's correlation coefficient $r = 0.79$, $p = 0.48$; and Pearson's correlation coefficient $r = 0.67$, $p = 0.1$. Breathiness of the stimuli s_{09} and s_{11} was 2.44 and 2.59, respectively. It is very close to the chosen threshold of 3. Excluding these two stimuli would further improve the agreement between the predicted and the subjective roughness of the real vowels.

These results support the previous observations about the poor performance of the roughness model for noise stimuli. The central stage of the model calculates cross-correlation coefficients between the signals in the individual channels of the central stage. This was inspired by the Daniel and Weber roughness model where it helped to

Table 6.2: Subjective and predicted roughness of real vowels /a/

stimuli	$R_{subj.}$	$R_{pred.}$(asper)	subj. breathiness
s_{01}	1.10	0.18	1.19
s_{02}	1.21	0.19	1.06
s_{03}	1.50	0.18	3.31
s_{04}	1.69	0.26	1.91
s_{05}	1.83	0.30	2.15
s_{06}	2.25	0.37	1.50
s_{07}	2.58	0.21	3.41
s_{08}	3.17	0.29	4.28
s_{09}	3.69	0.25	2.44
s_{10}	4.29	0.21	4
s_{11}	4.56	0.39	2.59
s_{12}	4.63	0.33	4.15

improve the model performance for unmodulated noise stimuli (Daniel & Weber, 1997). Another possible explanation for the poor model performance could be in the used method of the roughness listening test – rating listening test (see Section 5.3). Patel *et al.* (2012) described a different method to measure the roughness of voice stimuli, the method of adjustment. They compared the roughness ratings obtained by the method of adjustment and by the rating method and found differences. This may also contribute to the differences between the model predictions and the subjective data.

Vencovský (2014c) processed the real vowels /a/ by the Daniel and Weber roughness model (Daniel & Weber, 1997) and the SI roughness model (Leman, 2000). The Daniel and Weber roughness model was implemented in the PsySound3 sound analyses software (PsySound3, 2008) and the SI roughness model in the IPEM toolbox (IPEM, 2003). The roughness predicted using the Daniel and Weber roughness model did not agree well with the subjective data for all 11 vowels: Spearman's correlation coefficient $r = -0.077, p = 817$, Pearson's correlation coefficient $r = 0.114, p = 0.725$. The roughness predicted using the SI roughness model quite agreed with the subjective data for all 11 stimuli: Spearman's correlation coefficient: $r = 0.727, p = 4.9 \cdot 10^{-3}$, Pearson's correlation coefficient: $r = 0.680, p = 0.011$.

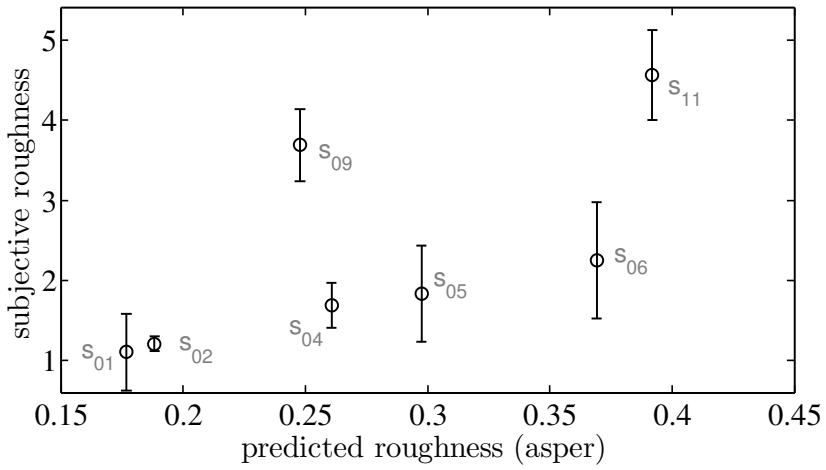

Figure 6.14: Mean values and standard deviations of the subjective ratings of roughness of real vowels /a/ (chosen subset of the stimuli with lowest breathiness) as a function of the predicted roughness. The predicted roughness is shown as the relative roughness to the roughness of 100% SAM tone with a frequency of 1 kHz, a level of 60 dB SPL and a modulation frequency of 70 Hz.

6.9 Summary

The roughness model described in Chapter 3 was used to predict the roughness of sinusoidally amplitude-modulated (SAM) tones; two tone stimuli (dyads) composed of pure-tones and harmonic complex tones; stimuli with envelopes that are not sinusoidal – pseudo amplitude-modulated (pAM) tones and stimuli with asymmetrical temporal envelopes; sinusoidally frequency-modulated (SFM) tones; unmodulated broadband noise; SAM complex tones; synthetic vowels; and real vowels. Results of listening tests conducted to measure the roughness of these stimuli were reproduced from the literature (Terhardt, 1968, 1974; Vogel, 1975; Kemp, 1982; Aures, 1985; Pressnitzer & McAdams, 1999; Vassilakis, 2005; Miśkiewicz *et al.*, 2006; Fastl & Zwicker, 2007) or obtained by means of the listening tests conducted within the framework of the thesis – for SAM complexes, synthetic and real vowels (see Chapter 5).

The predicted roughness agreed with the subjective data for most of the used stimuli. A very good agreement was shown between the predicted and subjective roughness data showing: the dependence of roughness of SAM tones and SFM tones on the modulation frequency (see Fig. 6.1 and Fig.6.8, respectively); the dependence of roughness of SAM tones on the modulation index (see Fig. 6.2); the dependence of roughness of dyads (composed of harmonic complex tones) on the frequency difference between the tones (see Fig. 6.5); the dependence of roughness of pAM tones on the relative phase between

the spectral components (see Fig. 6.6); roughness of SAM harmonic complex tones (see Fig. 6.11); and roughness of synthetic vowels /a/ (see Fig. 6.12). The advantage of the described roughness model in comparison with the roughness models known to the author is that it accounts for the roughness of stimuli with envelopes that are not sinusoidal – pAM tones and stimuli with asymmetrical temporal envelopes (see Section 6.3). Beside this, the roughness model was shown to perform better than the Daniel and Weber roughness model (Daniel & Weber, 1997) and the SI roughness model (Leman, 2000) for two harmonic complex tones (see Section 6.2.2 and also Vencovský (2014c)) and synthetic vowels /a/ (see Section 6.7).

The roughness model performed poor for unmodulated bandpass noise stimuli (see Section 6.5) and for real vowels /a/ (see Section 6.8). Section 6.8 showed that the additional listening test revealed that some of the real voice samples are breathy. If only the stimuli rated on the lower half of the breathiness scale are used, the roughness model performs better (see Fig. 6.14). Speech synthesizers (e.g. Klatt (1980)) add noise to the glottal signal in order to simulate breathiness. This together with the poor model performance for unmodulated bandpass noise stimuli indicates that the roughness model performs poor for stimuli with added noise. The central stage of the roughness model calculates crosscorrelation coefficients between the signals in the individual channels (see Section 3.2). This was inspired by the Daniel and Weber roughness model which performed very well for unmodulated bandpass noise stimuli (Daniel & Weber, 1997). However, the Daniel and Weber roughness model performed poor for the real vowels /a/ – see Section 6.8 and Vencovský (2014c). It is also possible that the rating listening test used to measure the roughness of the real vowels (see Section 5.3) is not suitable for this type of stimuli. Patel *et al.* (2012) used the rating listening test and the method of adjustment to measure the roughness of real vowels. The results obtained using both methods were different.

Chapter 7

Conclusion

7.1 General discussion

This thesis described a new roughness model. The roughness model was used to predict roughness of a large number of various types of acoustic stimuli. The predicted roughness was compared with results of listening tests (reproduced from the literature or conducted within the framework of the thesis).

The designed roughness model (Chapter 3) is composed of two successive stages: a peripheral stage and a central stage. The peripheral stage simulates the function of the peripheral ear: outer-/middle-ear, cochlear mechanics, inner hair cells and auditory nerve synapse. The algorithms simulating the function of the individual parts of the peripheral ear were adapted from the literature and composed into one model. The central stage predicts roughness from the stimuli processed by the peripheral stage. It employs algorithms designed within the framework of this thesis.

The peripheral stage of the roughness model simulates the limited frequency resolution of the peripheral ear. This is important for the prediction of roughness. The limited frequency resolution of the peripheral stage is accounted for by a model of the basilar membrane (BM) response and cochlear hydrodynamics designed by Mammano & Nobili (1993); Nobili & Mammano (1996) and Nobili *et al.* (2003). Specifically, the model variant with realistic parameters and dimensions of the human cochlea was used (Nobili *et al.*, 2003). The model is in this thesis called the Nobili *et al.* cochlear model.

The Nobili *et al.* cochlear model is a physical model which can simulate otoacoustic emissions (Nobili *et al.*, 2003). This thesis verified the ability of the Nobili *et al.* cochlear model to account for physiological and psychophysical data (Chapter 4). This thesis showed that the cochlear model is, in agreement with the experimental data observed in the mammalian cochlea, active, isointensity responses of the model are level dependent, input/output (I/O) function of the responses are compressively nonlinear and impulse responses are level near-invariant. Moreover, the cochlear model was verified also using the subjective data from masking experiments with pure tone and harmonic complex tone maskers. The masking data for harmonic complex maskers showing the frequency selectivity of the human hearing system were used to adjust the parameters affecting the frequency selectivity of the cochlear model. The model qualitatively accounted for the phenomena observed with pure tone and harmonic complex maskers: the upward spread of masking observed with pure tone maskers; the phase effects observed with harmonic complex (Schroeder phase) maskers; and the level effects observed with Schroeder phase maskers. These physiological and psychophysical phenomena were chosen since it is not accounted for by many cochlear models. It thus places a strong constraint on cochlear models (Shera, 2001; Oxenham & Dau, 2001a).

The Nobili *et al.* cochlear model thus could potentially serve as a front end in other applications, for example, assessment of sound quality, speech recognition, etc. Since it is a physical model accounting for otoacoustic emissions (Nobili *et al.*, 2003) and, as was shown in the thesis, for other physiological and psychophysical phenomena, the model could be used also to study how the cochlear mechanics affects perception (Epp *et al.*, 2010).

The roughness model was used in this thesis to predict roughness of: sinusoidally amplitude-modulated (SAM) tones, two tone stimuli (dyads) composed of pure tones and harmonic complex tones, pseudo amplitude-modulated (pAM) tones, stimuli with assymetrical temporal envelopes, sinusoidally frequency-modulated (SFM) tones, unmodulated bandpass noise stimuli, SAM harmonic complex tones, synthetic and real vowels /a/. The predicted roughness was compared with results of listening tests. The results were reproduced from the literature or measured within the framework of this thesis – for SAM complexes, synthetic and real vowels /a/.

The predicted roughness agreed with the subjective data for most of the used stimuli. The model covered the effect of phase of the spectral components and the shape of the temporal envelope on roughness which is its main advantage in comparison with the

roughness models known to the author of this thesis. The roughness model covered the roughness of theses stimuli because the statistics used in the central stage allows to take into account the shape of the signal envelope after it is processed by the peripheral stage. The largest discrepancies between the model predictions and the subjective data were for the unmodulated bandpass noise stimuli and for real vowels. These stimuli contained noise which worsened the model performance. This disadvantage of the roughness model should be fixed in the future work.

7.2 Overview of results

- The thesis describes a new roughness model designed within the framework of the thesis. The roughness model is composed of two successive stages: a peripheral and a central stage. The peripheral stage simulates the function of peripheral ear – algorithms simulating individual parts of the peripheral ear were adapted from the literature and composed into one model. The central stage was designed by the author. It predicts roughness from the output signal of the peripheral stage.

- The peripheral stage of the roughness model contains a physical model of the basilar membrane (BM) response and cochlear hydrodynamics (the Nobili *et al.* cochlear model). The thesis shows that the responses of the Nobili *et al.* cochlear model agree with similar responses measured in live mammalian cochlea – isointensity responses are level dependent, input/output (I/O) functions are compressively nonlinear and impulse responses are level near-invariant. The model was verified also using psychophysical masking thresholds for pure tone and harmonic complex maskers. The Nobili *et al.* cochlear model predicted the upward spread of masking thresholds observed during tone on tone masking. It also qualitatively predicted the phase effects in masking experiments with Schroeder phase maskers. These phenomena are not accounted for by many cochlear models (Shera, 2001; Oxenham & Dau, 2001a). These results thus show that the Nobili *et al.* cochlear model could be applicable as a front end in the roughness model and also in other possible applications.

- The thesis shows the results of roughness listening tests conducted with sinusoidally amplitude-modulated (SAM) harmonic complexes, synthetic vowels /a/, and samples of real vowels /a/. The tests were conducted within the framework of the thesis.

- The described roughness model was used to predict the roughness of a large number of various stimuli: sinusoidally amplitude-modulated (SAM) tones; two tone stimuli (dyads) composed of pure tones and harmonic complex tones; stimuli with envelopes that are not sinusoidal – pseudo amplitude-modulated (pAM) tones and stimuli with asymmetrical temporal envelopes; sinusoidally frequency-modulated (SFM) tones; unmodulated bandpass noise stimuli; SAM harmonic complexes; and synthetic and real vowels /a/.

- The predicted roughness agreed with the results of the listening tests conducted within the framework of the thesis or reproduced from the literature. Since the central stage contains a new algorithm which takes into account the shape of the signal envelope at the output of the peripheral stage, the roughness model also predicted the effect of phase of the spectral components and the shape of the temporal envelope on roughness. These effects are not well covered by the roughness models known to the author. The worst agreement between the predicted and subjective roughness was for unmodulated bandpass noise and real vowels /a/. These stimuli contained noise.

7.3 Future work

Since the worst agreement between the results of listening tests and predictions of the roughness model was in case of unmodulated bandpass noise stimuli and real voice samples, it should be focused on these stimuli in the future work.

Roughness of real vowels should be measured again using the method of adjustment as was described by Patel *et al.* (2012). This method used to measure the roughness of pathological voice samples (vowels) led to slightly different results than the rating method (Patel *et al.*, 2012).

Kreiman *et al.* (1994) showed that it is difficult for listeners to rate roughness of stimuli if they differ also in other perceptual quantities. For example, voice stimuli may differ in roughness and also in breathiness (Kreiman *et al.*, 1994), and violin tones played on the same string may differ in more than one unidimensional perceptual quantity (Otčenášek & Otčenášek, 2014). This raises a question whether it is possible to construct a roughness model which would account for roughness of all types of acoustic stimuli. This issue requires future research.

References

Aertsen, A.M.J.H. & Johannesma, P.I.M. (1980). Spectro-temporal receptive fields of auditory neurons in the grassfrog. I. characterization of tonal and natural stimuli. *Biol. Cybernetics*, **38**, 223–234.

Allen, J.B. (1977). Two-dimensional cochlear fluid model: New results. *J. Acoust. Soc. Am.*, **61**(1), 110–119.

Anderson, D.J., Rose, J.E., Hind, J.E. & Brugge, J.F. (1970). Temporal Position of Discharges in Single Auditory Nerve Fibers within the Cycle of a Sine-Wave Stimulus: Frequency and Intensity Effects. *J. Acoust. Soc. Am.*, **49**(4), 1131–1139.

Aures, W. (1984). *Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse*. PhD Thesis, TU München.

Aures, W. (1985). Ein Berechnungsverfahren der Rauhigkeit. *Acustica*, **58**(5), 268–281.

Bidelman, G.M. & Heinz, M.G. (2011). Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. *J. Acoust. Soc. Am.*, **130**(3), 1488–1502.

de Boer, E. (1975). Synthetic whole-nerve action potentials for the cat. *J. Acoust. Soc. Am.*, **58**(5), 1030–1045.

de Boer, E. (1995). The "inverse problem" solved for a three-dimensional model of the cochlea. I. Analysis. *J. Acoust. Soc. Am.*, **98**(2), 896–903.

de Boer, E. & Nuttall, A.L. (1997). The mechanical waveform of the basilar membrane. I. Frequency modulations ("glides") in impulse responses and cross-correlation functions. *J. Acoust. Soc. Am.*, **101**(6), 3583–3592.

Bouše, J. & Vencovský, V. (2011). Matlab Implementation of the Count-comparison LSO Model. In: *19th Annual Conference Proceedings Technical Computing Prague 2011*. Prague, 1–7.

Bouše, J. & Vencovský, V. (2012). Implementation of Binaural Processing Model. In: *Poster 2012*. Prague; CTU, 1–5.

Bouše, J. & Vencovský, V. (2013). The Matlab Implementation of Binaural Processing Model Simulating Lateral Position of Tones with Interaural Time Differences. In: *21th Annual Conference Proceedings Technical Computing Prague 2011*. Prague, 1–6.

Carney, L.H., McDuffy, M.J. & Shekhter, I. (1999). Frequency glides in the impulse responses of auditory-nerve fibers. *J. Acoust. Soc. Am.*, **105**(4), 2384–2391.

Cheatham, M.A. & Dallos, P. (2001). Inner hair cell response patterns: implications for low-frequency hearing. *J. Acoust. Soc. Am.*, **110**(4), 2034–2044.

Cooper, N.P. & Rhode, W.S. (1992). Basilar membrane mechanics in the hook region of cat and guinea-pig cochleae: sharp tuning and nonlinearity in the absence of baseline position shifts. *Hear. Res.*, **63**, 163–190.

Daniel, P. & Weber, R. (1997). Psychoacoustical Roughness: Implementation of an Optimized Model. *Acustica*, **83**(1), 113–123.

Dau, T., Kollmeier, B. & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.*, **102**(5), 2892–2905.

Dau, T., Püschel, D. & Kohlrausch, A. (1996). A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. *J. Acoust. Soc. Am.*, **99**(6), 3615–3622.

David, H.A. (1988). *The Method of Paired Comparisons*. New York: Oxford University Press, 2nd edition.

Ebeling, M. (2008). Neuronal periodicity detection as a basis for the perception of consonance: a mathematical model of tonal fusion. *J. Acoust. Soc. Am.*, **124**(4), 2320–2329.

Elliott, S.J., Lineton, B. & Ni, G. (2011). Fluid coupling in a discrete model of cochlear mechanics. *J. Acoust. Soc. Am.*, **130**(3), 1441–1451.

Epp, B., Verhey, J.L. & Mauermann, M. (2010). Modeling cochlear dynamics: interrelation between cochlea mechanics and psychoacoustics. *J. Acoust. Soc. Am.*, **128**(4), 1870–1883.

Fastl, H. & Zwicker, E. (2007). *Psychoacoustics: Facts and Models*. Berlin, Heidelberg: Springer.

Giguére, C. & Woodland, P.C. (1994). A computational model of the auditory periphery for speech and hearing research. I. Ascending path. *J. Acoust. Soc. Am.*, **95**(1), 331–342.

Goblick, T.J. & Pfeiffer, R.R. (1969). Time-domain measurements of cochlear nonlinearitites using combination click stimuli. *J. Acoust. Soc. Am.*, **46**(4), 924–938.

Goldstein, J.L. (1995). Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: data and MBPNL model. *Hear. Res.*, **89**, 52–68.

Goode, R.L., Killion, M., Nakamura, K. & Nishihara, S. (1994). New knowledge about the function of the human middle ear: development of an improved analog model. *J. Acoust. Soc. Am.*, **15**(2), 145–154.

Guirao, M. & Garavilla, J.M. (1976). Perceived roughness of amplitude-modulated tones and noise. *J. Acoust. Soc. Am.*, **60**(6), 1335–1338.

Guski, R. (1997). Psychological Methods for Evaluating Sound Quality and Assessing Acoustic Information. *Acta Acustica united with Acustica*, **83**(5), 765–774.

Hammershøi, D. & Møller, H. (1996). Sound transmission to and within the human ear canal. *J. Acoust. Soc. Am.*, **100**(1), 408–427.

Harlander, N., Huber, R. & Ewert, S. (2014). Sound Quality Assessment Using Auditory Models. *J. Audio Eng. Soc.*, **62**(5), 324–336.

von Helmholtz, H.L.F. (1877). *On the Sensations of Tone as the Physiological Basis for the Theory of Music*. 2nd ed., translated by A. J. Ellis (1885), 4th ed. (Dover, New York, 1954; from German).

Holmes, M.H. (1980). Ananalysis of a low-frequency model of the cochlea. *J. Acoust. Soc. Am.*, **68**(2), 482–488.

Howard, D.M., Abberton, E. & Fourcin, A. (2012). Disordered voice measurement and auditory analysis. *Speech Communication*, **54**, 611–622.

Huber, R. & Kollmeier, B. (2006). PEMO-Q A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. *IEEE Transactions on Audio, Speech and Language Processing*, **14**(6), 1902–1911.

Huber, A., Linder, T., Ferrazzini, M., Schmid, S., Dillier, N., Stoeckli, S. & Fisch, U. (2001). Intraoperative assessment of stapes movement. *Ann. Otol. Rhinol. Laryngol.*, **110**(1), 31–35.

Hudde, H. & Engel, A. (1998a). Measuring and Modeling Basic Properties of the Human Middle Ear and Ear Canal. Part I: Model Structure and Measuring Techniques. *Acustica*, **84**(4), 720–738.

Hudde, H. & Engel, A. (1998b). Measuring and Modeling Basic Properties of the Human Middle Ear and Ear Canal. Part II: Ear Canal, Middle Ear Cavities, Eardrum, and Ossicles. *Acustica*, **84**(5), 894–913.

Hudde, H. & Engel, A. (1998c). Measuring and Modeling Basic Properties of the Human Middle Ear and Ear Canal. Part III: Eardrum Impedances, Transfer Functions and Model Calculations. *Acustica*, **84**(6), 1091–1108.

IPEM (2003). Toolbox for perception-based music analysis. `http://www.ipem.ugent.be/?q=node/27`.

Irino, T. & Patterson, R.D. (1997). A time-domain, level-dependent auditory filter: The gammachirp. *J. Acoust. Soc. Am.*, **101**(1), 412–419.

Irino, T. & Patterson, R.D. (2001). A compressive gammachirp auditory filter for both physiological and psychophysical data. *J. Acoust. Soc. Am.*, **109**(5), 2008–2022.

Jenison, R.L., Greenberg, S., Kluender, K.R. & Rhode, W.S. (1991). A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers. *J. Acoust. Soc. Am.*, **90**(2), 773–786.

Jepsen, M.L., Ewert, S.D. & Dau, T. (2008). A computational model of human auditory signal processing and perception. *J. Acoust. Soc. Am.*, **124**(1), 422–438.

Johnson, D.H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *J. Acoust. Soc. Am.*, **68**(4), 1115–1122.

Kameoka, A. & Kuriyagawa, M. (1969). Consoance Theory Part I: Consonance of Dyads. *J. Acoust. Soc. Am.*, **45**(6), 1451–1459.

Kemp, S. (1982). Roughness of Frequency-Modulated Tones. *Acustica*, **50**(2), 126–133.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.*, **67**(3), 971–995.

Kohlrausch, A., Hermes, D. & Duisters, R. (2005). Modeling roughness perception for sounds with ramped and damped temporal envelopes. In: *Forum Acusticum*. Budapest, Hungary, 1719–1724.

Kohlrausch, A. & Sander, A. (1995). Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. *J. Acoust. Soc. Am.*, **97**(3), 1817–1829.

Kreiman, J., Gerratt, B.R. & Berke, G.S. (1994). The multidimensional nature of pathologic vocal quality. *J. Acoust. Soc. Am.*, **96**(3), 1291–1302.

Kringlebotn, M. (1988). Network model for the human middle ear. *Scand. Audiol.*, **17**(2), 75–85.

Leman, M. (2000). Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (SIM). In: *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*. Verona, Italy, DAFX 1–DAFX 6.

Lentz, J.J. & Leek, M.R. (2001). Psychophysical estimates of cochlear phase response: masking by harmonic complexes. *J. Assoc. Res. Otolaryngol*, **2**(4), 408–422.

Lin, T. & Guinan, J.J. (2000). Auditory-nerve-fiber responses to high-level clicks: interference patterns indicate that excitation is due to the combination of multiple drives. *J. Acoust. Soc. Am.*, **107**(5), 2615–2630.

Lopez-Najera, A., Lopez-Poveda, E.A. & Meddis, R. (2007). Further studies on the dual-resonance nonlinear filter model of cochlear frequency selectivity: responses to tones. *J. Acoust. Soc. Am.*, **122**(4), 2124–2134.

Lopez-Poveda, E.A., Barrios, L.F. & Alves-Pinto, A. (2007). Psychophysical estimates of level-dependent best-frequency shifts in the apical region of the human basilar membrane. *J. Acoust. Soc. Am.*, **121**(6), 3646–3654.

Lopez-Poveda, E.A. & Eustaquio-Martín, A. (2006). A Biophysical Model of the Inner Hair Cell: The Contribution of Potassium Currents to Peripheral Auditory Compression. *J. Assoc. Res. Otolaryngol.*, **7**(3), 218–235.

Lyon, R.F., Katsiamis, A.G. & Drakakis, E.M. (2010). History and Future of Auditory Filter Models. In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*. Paris, 3809–3812.

Mammano, F. & Nobili, R. (1993). Biophysics of the cochlea: linear approximation. *J. Acoust. Soc. Am.*, **93**(6), 3320–3332.

MAP (2014). Matlab auditory periphery (MAP), University of Essex: Hearing Research Laborathory. `http://www.essex.ac.uk/psychology/department/hearinglab/modelling.html`.

Mathes, R.C. & Miller, R.L. (1947). Phase Effects in Monaural Perception. *J. Acoust. Soc. Am.*, **19**(5), 780–797.

McDermott, J., Lehr, A.J. & Oxenham, A.J. (2010). Individual Differences Reveal the Basis of Consonance. *Curr Biol.*, **20**(11), 1035–1041.

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

Meddis, R. (2006). Auditory-nerve first-spike latency and auditory absolute threshold: a computer model. *J. Acoust. Soc. Am.*, **119**(1), 406–417.

Meddis, R. & Lopez-Poveda, E.A. (2010). Auditory periphery: From pinna to auditory nerve. In: R. Meddis, E. A.Lopez-Poveda, A.N. Popper & R.R. Fay (editors), *Computational Models of the Auditory System (Springer Handbook of Auditory Research)*. New York, Dordrecht, Heidelberg, London: Springer.

Meddis, R., O'Mard, L.P. & Lopez-Poveda, E.A. (2001). A computational algorithm for computing nonlinear auditory frequency selectivity. *J. Acoust. Soc. Am.*, **109**(6), 2852–2861.

Miśkiewicz, A., Rakowski, A. & Rościszewska, T. (2006). Perceived Roughness of Two Simultaneous Pure Tones. *Acta Acustica united with Acustica*, **96**(2), 331–336.

Moore, B.C.J., Alcántara, J.I. & Dau, T. (1998). Masking patterns for sinusoidal and narrow-band noise maskers. *J. Acoust. Soc. Am.*, **104**(2), 1023–1038.

Moore, B.C.J. & Glasberg, B.R. (1996). A revision of Zwicker's loudness model. *Acta Acustica united with Acustica*, **82**(2), 335–345.

Nishimura, A. (2005). An auditory model that can account for frequency selectivity and phase effects on masking. *Acoust. Sci. & Tech.*, **25**(5), 330–339.

Nobili, R. & Mammano, F. (1996). Biophysics of the cochlea. II: Stationary nonlinear phenomenology. *J. Acoust. Soc. Am.*, **99**(4), 2244–2255.

Nobili, R., Vetešník, A., Turicchia, L. & Mammano, F. (2003). Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model. *J. Assoc. Res. Otolaryngol.*, **4**(4), 478–494.

O'Connor, K.N. & Puria, S. (2008). Middle-ear circuit model parameters based on a population of human ears. *J. Acoust. Soc. Am.*, **123**(1), 197–211.

Otčenášek, Z. & Otčenášek, Z. (2014). Perception of different types of roughness of violin tones. In: *Proceedings of International Symposium on Musical Acoustics (ISMA2014)*. Le Mans, France, 557–562.

Oxenham, A.J. (2010). The Perception of Musical Tones. In: D. Deutsch (editor), *The Psychology of Music*. London: Academic Press, Third edition, 1–33.

Oxenham, A.J. & Dau, T. (2001a). Reconciling frequency selectivity and phase effects in masking. *J. Acoust. Soc. Am.*, **110**(3), 1525–1538.

Oxenham, A.J. & Dau, T. (2001b). Towards a measure of auditory-filter phase response. *J. Acoust. Soc. Am.*, **110**(6), 3169–3178.

Oxenham, A.J. & Dau, T. (2004). Masker phase effects in normal-hearing and hearing-impaired listeners: evidence for peripheral compression at low signal frequencies. *J. Acoust. Soc. Am.*, **116**(4), 2248–2257.

Patel, S.A., Shrivastav, R. & Eddins, D.A. (2012). Identifying a Comparison for Matching Rough Voice Quality. *J. Speech Lang. Hear. Res.*, **55**(5), 1407–1422.

Patterson, R.D. & Nimmo-Smith, I. (1980). Off-frequency listening and auditory-filter asymmetry. *J. Acoust. Soc. Am.*, **67**(1), 229–245.

Patterson, R.D., Nimmo-Smith, I., Weber, D.L. & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J. Acoust. Soc. Am.*, **72**(6), 1788–1803.

Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C. & Allerhand, M. (1992). Complex sounds and auditory images. In: Y. Cazals, L. Demany & K. Horner (editors), *Auditory Physiology and Perception, Proc. 9th International Symposium on Hearing*. Pergamon, Oxford, 429–446.

Patterson, R.D., Unoki, M. & Irino, T. (2003). Extending the domain of center frequencies for the compressive gammachirp auditory filter. *J. Acoust. Soc. Am.*, **114**(3), 1529–1542.

Peterson, L.C. & Bogert, B.P. (1950). A Dynamical Theory of the Cochlea. *J. Acoust. Soc. Am.*, **22**(3), 369–381.

Pfeiffer, R.R. (1970). A model for two-tone inhibition of single cochlear-nerve fibers. *J. Acoust. Soc. Am.*, **48**(6), 1373–1378.

Pickles, J.O. (2008). *An Introduction to the Physiology of Hearing*. Howard House, Wagon Lane, Bingley, United Kingdom: Emerald Group Publishing Limited, 3rd edition.

Plomp, R. & Levelt, W.J.M. (1965). Tonal Consonance and Critical Bandwidth. *J. Acoust. Soc. Am.*, **38**(4), 548–560.

Plomp, R. & Steeneken, H.J.M. (1968). Interference between two simple tones. *J. Acoust. Soc. Am.*, **43**(4), 883–884.

Pressnitzer, D. & McAdams, S. (1999). Two phase effects in roughness perception. *J. Acoust. Soc. Am.*, **105**, 2773–2782.

Pressnitzer, D., McAdams, S., Winsberg, S. & Fineberg, J. (2000). Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness. *Perception & Psychophysics*, **62**, 66–80.

PsySound3 (2008). Sound analyses software. `http://psysound.wikidot.com/`.

Püschel, D. (1988). *Prinzipien der zeitlichen Analyse beim Hören*. Ph.D. thesis, University of Göttingen, Göttingen, Germany.

Ranke, O.F. (1950). Theory of Operation of the Cochlea: A Contribution to the Hydrodynamics of the Cochlea. *J. Acoust. Soc. Am.*, **22**(6), 772–777.

Recio, A. & Rhode, W.S. (2000). Basilar membrane responses to broadband stimuli. *J. Acouts. Soc. Am.*, **108**(5), 2281–2298.

Rhode, W.S. & Cooper, N.P. (1996). Nonlinear Mechanics in the Apical Turn of the Chinchilla Cochlea In Vivo. *Auditory Neuroscience*, **3**, 101–121.

Rhode, W. & Recio, A. (2000). Study of mechanical motion in the basal region of the chinchilla cochlea. *J. Acoust. Soc. Am.*, **107**(6), 3317–3332.

Robert, A. & Eriksson, J.L. (1999). A composite model of the auditory periphery for simulating responses to complex sounds. *J. Acoust. Soc. Am.*, **106**(4), 1852–1864.

Robles, L., Rhode, W.S. & Geisler, C.D. (1976). Transient response of the basilar membrane measured in squirrel monkeys using the Mössbauer effect. *J. Acoust. Soc. Am.*, **59**(4), 926–939.

Robles, L. & Ruggero, M.A. (2001). Mechanics of the mammalian cochlea. *Physiological Reviews*, **81**(3), 1305–1352.

Ruggero, M.A., Rich, N.C., Recio, A., Narayan, S.S. & Robles, L. (1997). Basilar-membrane responses to tones at the base of the chinchilla cochlea. *J. Acoust. Soc. Am.*, **101**(4), 2151–2163.

Ruggero, M.A., Rich, N.C., Robles, L. & Shivapuja, B.G. (1990). Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea. *J. Acoust. Soc. Am.*, **87**(4), 1612–1629.

Rund, F. (2004). *Přenos akustického tlaku vnějším zvukovodem*. Ph.D. thesis, Czech Technical University in Prague, Department of Radioelectronics, Prague, Czech Republic.

Schroeder, M.R. (1970). Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans. Inf. Theory.*, **16**(1), 85–89.

Schroeder, M.R. (1978). Models of hearing. In: *Proceedings of IEEE*, volume 63. 1332–1350.

Sellick, P.M., Yates, G.K. & Patuzzi, R. (1983). The influence of Mössbauer source size and position on phase and amplitude measurements of the guinea pig basilar membrane. *Hear. Res.*, **10**(1), 101–108.

Sethares, W.A. (2005). *Tuning, Timbre, Spectrum, Scale*. London, Berlin, Heidelberg: Springer-Verlag, 2nd edition.

Shamma, S.A., Chadwick, R.S., Wilbur, W.J., Morrish, K.A. & Rinzel, J. (1986). A biophysical model of cochlear processing: intensity dependence of pure tone responses. *J. Acoust. Soc. Am.*, **80**(1), 133–145.

Shen, Y. & Lentz, J.J. (2009). Level dependence in behavioral measurements of auditory-filter phase characteristics. *J. Acoust. Soc. Am.*, **126**(5), 2501–2510.

Shera, C.A. (2001). Intensity-invariance of fine time structure in basilar-membrane click responses: implications for cochlear mechanics. *J. Acoust. Soc. Am.*, **110**(1), 332–348.

Shera, C.A., Tubis, A. & Talmadge, C.L. (2004). Do Forward- and Backward-Traveling Waves Occur Within the Cochlea? Countering the Critique of Nobili et al. *J. Assoc. Res. Otolaryngol.*, **5**(4), 349–359.

Shrivastav, R., Sapienza, C.M. & Nandur, V. (2005). Application of psychometric theory to the measurement of voice quality using rating scales. *J. Speech. Lang. Hear. Res.*, **48**(2), 323–335.

Siebert, W.M. (1974). Ranke revisited – a simple short-wave cochlear model. *J. Acoust. Soc. Am.*, **56**(2), 594–600.

Sieroka, N., Dosch, H.G. & Rupp, A. (2006). Semirealistic models of the cochlea. *J. Acoust. Soc. Am.*, **120**(1), 297–304.

Smith, B.K., Sieben, U.K., Kohlrausch, A. & Schroeder, M.R. (1986). Phase effects in masking related to dispersion in the inner ear. *J. Acoust. Soc. Am.*, **80**(6), 1631–1637.

Steele, C.R. & Taber, L.A. (1981). Three-dimensional model calculations for guinnea pig cochlea. *J. Acoust. Soc. Am.*, **69**(4), 1107–1111.

Strube, H.W. (1985). A Computationally Efficient Basilar-Membrane Model. *Acustica*, **58**(4), 207–214.

Sumner, C.J., Lopez-Poveda, E.A., O'Mard, L.P. & Meddis, R. (2002). A revised model of the inner-hair cell and auditory-nerve complex. *J. Acoust. Soc. Am.*, **111**(5), 2178–2188.

Terhardt, E. (1968). Über akustische Rauhigkeit und Schwankungsstärke. *Acustica*, **20**(4), 215–224.

Terhardt, E. (1974). On the Perception of Periodic Sound fluctuations (Roughness). *Acustica*, **30**(4), 201–213.

Unoki, M., Irino, T., Glasberg, B., Moore, B.C.J. & Patterson, R.D. (2006). Comparison of the roex and gammachirp filters as representations of the auditory filter. *J. Acoust. Soc. Am.*, **120**(3), 1474–1492.

Vassilakis, P.N. (2001). Perceptual and physical properties of amplitude fluctuation and their musical significance. PhD thesis, University of California, Los Angeles.

Vassilakis, P.N. (2005). Auditory roughness as means of musical expression. In: R.A. Kendall & R.W.H. Savage (editors), *Selected Reports in Ethnomusicology*, volume 12. Los Angeles: UCLA Ethnomusicology Publications, 119–144.

Vencovský, V. (2008). Influence of ear model parameter setting on simulated tone-on-tone masking patterns. In: *Poster 2008*. Prague, CTU, 1–8.

Vencovský, V. (2009). A Physiological Auditory Model. In: *126th AES Convention Papers*. New York: Audio Engineering Society, 1–7.

Vencovský, V. (2010). Objective Audio Quality Assessment Using a Model of Auditory Perception. In: *Poster 2010*. Prague; CTU, 1–4.

Vencovský, V. (2014a). Modeling roughness perception for complex stimuli using a model of cochlear hydrodynamics. In: *Proceedings of International Symposium on Musical Acoustics (ISMA2014)*. Le Mans, France, 483–488.

Vencovský, V. (2014b). Modeling roughness perception for complex stimuli using a model of cochlear hydrodynamics. In: *Proceedings of Forum Acusticum*. Krakow, Poland, 376.

Vencovský, V. (2014c). Roughness prediction for complex acoustic stimuli. *Akustické listy*, **20**, 19–26.

Vencovský, V. (2015a). Prediction of masking thresholds for schroeder phase maskers: masker level effects. In: *Proceedings of DAGA, submitted in February 2015*. Nürnberg, Germany.

Vencovský, V. (2015b). Roughness prediction based on a model of the cochlear hydrodynamics. *to be submitted in March 2015*.

Vencovský, V. & Bouše, J. (2011). Binaural Processing Model Simulating the Lateral Position of Tones with Interaural Time Differences. In: *Poster 2011*. Prague; CTU, 1–6.

Vencovský, V. & Kadlec, F. (2009). Sound signal analyzer. In: *Technical Ccomputing Prague*. Prague; CTU, 1–8.

Vencovský, V. & Rund, F. (2012). Pure Tone Audiometer. In: *20th Annual Conference Proceeding's Technical Computing Bratislava*. 1–5.

Verhulst, S., Dau, T. & Shera, C.A. (2012). Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission. *J. Acoust. Soc. Am.*, **132**(6), 3842–3848.

Vogel, A. (1975). Über den Zusammenhang zwischen Rauhigkeit und Modulationsgrad. *Acustica*, **32**(5), 300–306.

Wang, Y. (2009). A Study on Sound Roughness Evaluation Based on an Auditory Synchronization Index Model. In: *Proceedings of the 2nd International Congress on Image and Signal Processing, IEEE*. 3612–3616.

Wang, Y.S., Shen, G.Q., Guo, H., Tang, X.L. & Hamade, T. (2013). Roughness modelling based on human auditory perception for sound quality evaluation of vehicle interior noise. *J. Sound and Vibration*, **332**(16), 3893–3904.

Westerman, L.A. & Smith, R.L. (1988). A diffusion model of the transient response of the cochlear inner hair cell synapse. *J. Acoust. Soc. Am.*, **83**(6), 2266–2276.

Wojtczak, M. & Oxenham, A.J. (2009). On- and Off-Frequency Forward Masking by Schroeder-Phase Complexes. *J. Assoc. Res. Otolaryngol.*, **10**(4), 595–607.

Zeddies, D.G. & Siegel, J.H. (2004). A biophysical model of an inner hair cell. *J. Acoust. Soc. Am.*, **116**(1), 426–441.

Zhang, X., Heinz, M.G., Bruce, I.C. & Carney, L.H. (2001). A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J. Acoust. Soc. Am.*, **109**(2), 648–670.

Zilany, M.S.A. & Bruce, I.C. (2006). Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. *J. Acoust. Soc. Am.*, **120**(3), 1446–1466.

Zilany, M.S.A., Bruce, I.C., Nelson, P.C. & Carney, L.H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *J. Acoust. Soc. Am.*, **126**(5), 2390–2412.

Zinn, C., Maier, H., Zenner, H. & Gummer, A.W. (2000). Evidence for active, nonlinear, negative feedback in the vibration response of the apical region of the in-vivo guinea-pig cochlea. *Hear. Res.*, **142**(1-2), 159–183.

Zwislocki, J. (1950). Theory of the Acoustical Action of the Cochlea. *J. Acoust. Soc. Am.*, **22**(6), 778–784.

List of authors publications

Publications in the framework of the thesis

Journal articles

Vencovský, V. (2014c). Roughness prediction for complex acoustic stimuli. *Akustické listy*, **20**, 19–26

Submitted journal articles

Vencovský, V. (2015b). Roughness prediction based on a model of the cochlear hydrodynamics. *to be submitted in March 2015*

Conference papers

Vencovský, V. (2014b). Modeling roughness perception for complex stimuli using a model of cochlear hydrodynamics. In: *Proceedings of Forum Acusticum*. Krakow, Poland, 376

Vencovský, V. (2014a). Modeling roughness perception for complex stimuli using a model of cochlear hydrodynamics. In: *Proceedings of International Symposium on Musical Acoustics (ISMA2014)*. Le Mans, France, 483–488

Vencovský, V. (2010). Objective Audio Quality Assessment Using a Model of Auditory Perception. In: *Poster 2010*. Prague; CTU, 1–4

Vencovský, V. (2009). A Physiological Auditory Model. In: *126th AES Convention Papers*. New York: Audio Engineering Society, 1–7

Vencovský, V. (2008). Influence of ear model parameter setting on simulated tone-on-tone masking patterns. In: *Poster 2008*. Prague, CTU, 1–8

Submitted conference papers

Vencovský, V. (2015a). Prediction of masking thresholds for schroeder phase maskers: masker level effects. In: *Proceedings of DAGA, submitted in February 2015*. Nürnberg, Germany

Other publications

Conference papers

Bouše, J. & Vencovský, V. (2013). The Matlab Implementation of Binaural Processing Model Simulating Lateral Position of Tones with Interaural Time Differences. In: *21th Annual Conference Proceedings Technical Computing Prague 2011*. Prague, 1–6

Bouše, J. & Vencovský, V. (2012). Implementation of Binaural Processing Model. In: *Poster 2012*. Prague; CTU, 1–5

Vencovský, V. & Rund, F. (2012). Pure Tone Audiometer. In: *20th Annual Conference Proceeding's Technical Computing Bratislava*. 1–5

Bouše, J. & Vencovský, V. (2011). Matlab Implementation of the Count-comparison LSO Model. In: *19th Annual Conference Proceedings Technical Computing Prague 2011*. Prague, 1–7

Vencovský, V. & Bouše, J. (2011). Binaural Processing Model Simulating the Lateral Position of Tones with Interaural Time Differences. In: *Poster 2011*. Prague; CTU, 1–6

Vencovský, V. & Kadlec, F. (2009). Sound signal analyzer. In: *Technical Ccomputing Prague*. Prague; CTU, 1–8

Appendix A

Parameters of the IHC/AN model

Table A.1: Parameters of the IHC/AN model: IHC membrane potential.

τ_c, cilia/BM time constant (s)	1.3E-4
C_{cilia}, cilia/BM coupling gain (dB)	0.03
s_0, displacement sensitivity (m^{-1})	30E-9
u_0, displacement offset (m)	5E-9
s_1, displacement sensitivity (m^{-1})	1E-9
u_1, displacement offset (m)	1E-9
$G_{\text{cilia}}^{\max}$, max. mechanical conduct. (S)	6E-9
G_0, resting conductance (S)	8E-10
E_t, endocochlear potential (V)	0.1
E_k, potassium reversal potential (V)	-0.08
$R_p/(R_p + R_t)$, resting conductance (S)	2E-8
C_m, total capacitance (F)	4E-12

Table A.2: Parameters of the IHC/AN model: Presynaptic calcium level.

E_{Ca}, reversal potential (V)	0.066
β_{Ca}	400
γ_{Ca}	130
τ_m, calcium current time constant (s)	5E-5
τ_{Ca} calcium clearance time constant (s)	4E-5
z, converts from $[\text{Ca}^{2+}]^3$ to probability	2E42
$G_{\text{cilia}}^{\max}$,maximum Ca^{2+} conductance	1.4E-8

Table A.3: Parameters of the IHC/AN model: IHC transmitter release parameters.

y, replenishment rate (s^{-1})	6
l, loss rate (s^{-1})	250
x, reprocessing rate (s^{-1})	60
r, recovery rate (s^{-1})	500
M, maximum free transmitter quanta	12