www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

PAN ARCHIVES OF ACOUSTICS
Vol. 41, No. 2, pp. 189–201 (2016)

DOI: 10.1515/aoa-2016-0019

Research Papers

Roughness Prediction Based on a Model of Cochlear Hydrodynamics

Václav VENCOVSKÝ[1),2)]

[1)] *Musical Acoustics Research Center, Academy of Performing Arts in Prague*
Malostranké nám. 12, 118 00 Prague, Czech Republic; e-mail: vaclav.vencovsky@gmail.com

[2)] *Department of Radioelectronics, Czech Technical University in Prague*
Technická 2, 166 27 Prague, Czech Republic

(*received March 18, 2015; accepted December 18, 2015*)

The term roughness is used to describe a specific sound sensation which may occur when listening to stimuli with more than one spectral component within the same critical band. It is believed that the spectral components interact inside the cochlea, which leads to fluctuations in the neural signal and, in turn, to a sensation of roughness. This study presents a roughness model composed of two successive stages: peripheral and central. The peripheral stage models the function of the peripheral ear. The central stage predicts roughness from the temporal envelope of the signal processed by the peripheral stage. The roughness model was shown to account for the perceived roughness of various types of acoustic stimuli, including the stimuli with temporal envelopes that are not sinusoidal. It thus accounted for effects of the phase and the shape of the temporal envelope on roughness. The model performance was poor for unmodulated bandpass noise stimuli.

Keywords: roughness; prediction of roughness; auditory models; peripheral ear.

Notations

AM – amplitude-modulated,
AN – auditory nerve,
BM – basilar membrane,
CF – characteristic frequency,
ERB – equivalent rectangular bandwidth,
pAM – pseudo amplitude-modulated,
RMS – root mean square,
SAM – sinusoidally amplitude-modulated,
SFM – sinusoidally frequency-modulated,
SI – synchronization index,
SPL – sound pressure level.

1. Introduction

Roughness is the term first introduced by von Helmholtz to describe a harsh, buzzing or rattling sound sensation (Helmholtz, 1895). This sound quality usually accompanies stimuli with more than one spectral component within the same critical band. In the peripheral ear, interaction of these spectral components produces beats. For the frequencies up to about 20 Hz, the beats are perceived as fluctuations in loudness; for the frequencies between about 20 and 300 Hz, the beats are perceived as roughness (Mathes, Miller, 1947; Terhardt, 1974; Vassilakis, 2001).

The unit of roughness is the asper – 1 asper is the roughness of a 100% SAM 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). A number of studies have shown the dependencies of perceived roughness on physical parameters of synthetic stimuli, e.g., for two-tone stimuli (Plomp, Steeneken, 1968; Vassilakis, 2001; Miśkiewicz *et al.*, 2006; 2007), sinusoidally amplitude-modulated (SAM) tones (Terhardt, 1974; Aures, 1984), sinusoidally frequency-modulated (SFM) tones (Kemp, 1982), mixed modulated (SAM and SFM) tones (Hartmann, Hnath, 1982; Kin, Dobrucki, 1998), stimuli with envelopes that are not sinusoidal (Mathes, Miller, 1947; Pressnitzer, McAdams, 1999), and unmodulated bandpass noises (Aures, 1985), reviewed in (Fastl, Zwicker, 2007; Daniel, Weber, 1997). It has been shown that the dependence of roughness of SAM tones, broadband noises, or SFM tones on modulation frequency exhibits a bandpass characteristic – it increases with increasing modulation frequency, reaches its maximum, and then decreases, as the modulation frequency further increases

(Fastl, Zwicker, 2007). A similar bandpass characteristic exhibits the roughness of two simultaneous pure tones, shown as a function of their frequency difference (Miśkiewicz *et al.*, 2006).

Not only the frequency difference between the spectral components, but also their relative phase may affect roughness. Mathes and Miller (1947) showed the phase effects using quasi frequency-modulated (QFM) tones; Pressnitzer and McAdams (1999) used pseudo amplitude-modulated (pAM) tones and stimuli with asymmetrical temporal envelopes. Some of the stimuli had a similar temporal envelope but were perceived with a different roughness. Pressnitzer and McAdams (1999) processed these stimuli by a model of the peripheral ear and obtained signals with a different shape of the temporal envelope – the stimuli with a higher roughness had steeper rising parts of the temporal envelope.

Various roughness models have been developed in the past decades, for a review see (Vassilakis, 2001; Leman, 2000). So called "curve-mapping" models detect frequency components in the spectrum of the sound and map it onto a psychoacoustical curve of roughness (Leman, 2000). For example, Vassilakis (2000) designed a roughness model of this type. The curve-mapping models cannot predict roughness of signals with a continuous spectrum, e.g., noises, or roughness of pAM stimuli, or stimuli with asymmetrical temporal envelopes (Vassilakis, 2001; Leman, 2000). A more general approach to predict roughness is to employ auditory models (models of the peripheral ear). Daniel and Weber (1997) improved on a model designed by Aures (1985), see also (Wang *et al.*, 2013). The model predicts roughness from the estimated modulation depth of a signal filtered by a bank of critical band (Bark scale) filters. Daniel and Weber verified the roughness model and showed a very good agreement between predictions and subjective data for SAM and SFM tones and for unmodulated bandpass noises (Daniel, Weber, 1997). Leman (2000) introduced the synchronisation index (SI) roughness model. The SI model employs an auditory model which transforms the input acoustic stimulus into the simulated neural signal in auditory nerve (AN) fibers; it predicts roughness from the energy of the short-term spectrum of the simulated neural signal – synchronisation of the neural signal with beats increases the energy at low frequencies (up to 300 Hz).

The Daniel and Weber and SI roughness models cannot account for effects of the phase on roughness (Pressnitzer, McAdams, 1999; Leman, 2000). In order to cover these stimuli, Kohlrausch *et al.* (2005) adjusted the peripheral stages of the Daniel and Weber and SI roughness models. They included various algorithms simulating the adaptation of the neural signal in AN fibers. This modification made the Daniel and Weber roughness model sensitive to the shape of the waveform envelope. However, Kohlrausch *et al.* (2005) concluded that the predicted data did not agree well with the subjective roughness.

This study presents a roughness model which employs a previously designed model of cochlear hydrodynamics. The roughness model is an improved variant of the roughness models published in (Vencovský, 2014a; 2014b). It allows to take into account the shape of the temporal envelope of the signal obtained after processing by a model of the peripheral ear. The study verifies the model using results (reproduced from literature) of listening tests for SAM tones, pAM tones, stimuli with asymmetrical temporal envelopes, SFM tones, unmodulated narrowband noise, and two-tone stimuli composed of harmonic complex tones (harmonic intervals of the chromatic scale).

2. Roughness model

The roughness model is composed of two successive stages: peripheral and central. Figure 1 shows a diagram of the model. Algorithms used in the peripheral stage were adapted from literature; the central stage was designed in this study.

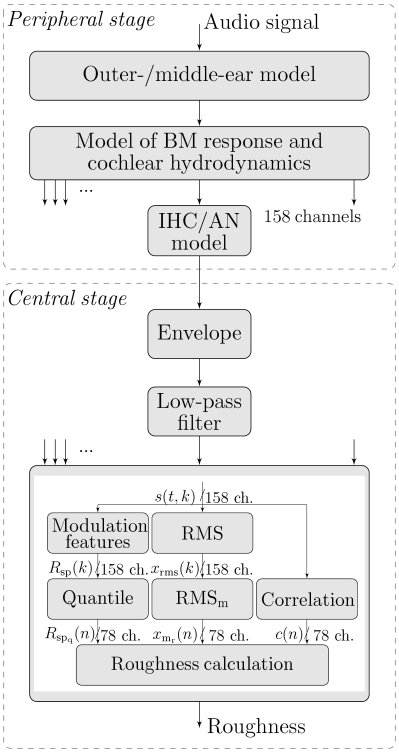

Fig. 1. Roughness model diagram.

2.1. Peripheral stage

The peripheral stage is composed of three successive blocks: an outer- and middle-ear model, a model of the basilar membrane (BM) and cochlear hydrodynamics, and a model of the inner hair cells (IHCs) and

the auditory nerve (AN) synapse. Its input is an acoustic waveform at the entrance of the outer ear, and its output is a probability of neural discharge in AN fibers.

Outer- and middle-ear model

The outer- and middle-ear model was designed and implemented in (MAP, 2014). Its first part, the outer-ear model, is composed of two parallel 1-st order Butterworth bandpass filters which approximate resonances of the auditory canal. One of the filters has a gain of 10 dB, a lower cutoff frequency of 1 kHz, and a higher cutoff frequency of 4 kHz. The other one has a gain of 25 dB, a lower cutoff frequency of 2.5 kHz, and a higher cutoff frequency of 7 kHz. The output signal of the outer-ear model is a sum of the filtered signals and the sound pressure wave at the input of the outer-ear model. The middle-ear model transforms the output of the outer-ear model to the simulated vibrations of the stapes. It is composed of two cascaded 1st-order Butterworth filters: a lowpass filter with a cutoff frequency of 50 Hz, and a highpass filter with a cutoff frequency of 1 kHz. The filtered signal is multiplied by a scalar of 45×10^{-9}, which transforms the filtered signal into the stapes displacement in meters.

Model of the BM and cochlear hydrodynamics

The model of the BM and cochlear hydrodynamics was designed in (Mammano, Nobili, 1993; Nobili, Mammano, 1996; Nobili *et al.*, 2003). It approximates the BM by an array of damped oscillators coupled via surrounding fluid. The displacement, ξ_i, of the i-th oscillator is given by

$$m_i\ddot{\xi}_i(t) + h_i\dot{\xi}_i(t) + s_i[2\dot{\xi}_i(t) - \dot{\xi}_{i-1}(t) - \dot{\xi}_{i+1}(t)] + k_i\xi_i(t) = f_{H_i}(t) + f_{O_i}[\eta_i(t)], \quad (1)$$

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

$$f_{H_i}(t) = -G_{S_i}a_{S_i}(t) - \sum_{j=1}^{N} G_i^j\ddot{\xi}_j(t), \quad (2)$$

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 $f_{O_i}[\eta_i(t)]$ is a nonlinear sigmoidal function transforming outer hair cell (OHC) stereocilia displacement $\eta_i(t)$ into the OHC force. The stereocilia displacement is given by the second array of oscillators simulating the subsystem composed of the tectorial membrane (TM) and cells in the organ of Corti. For details, the reader is referred to (Nobili *et al.*, 2003).

In this study, the model is used with parameters given in (Nobili *et al.*, 2003). However, damping h_i was multiplied by a constant of 0.31 and feedback force $f_{O_i}[\eta_i(t)]$ of the oscillators by a constant of 0.43 in order to decrease the bandwidth of cochlear filters at higher levels. Table 1 shows the equivalent rectangular bandwidths (ERBs) of the simulated cochlear filters. $\mathrm{ERB}_{\mathrm{GM}}$ are the ERBs of auditory filters measured in (Glasberg, Moore, 1990). The model is composed of 300 oscillators with characteristic frequencies (CFs) ranging between 30 Hz and 17 kHz. The CF of each channel is the frequency of 10-dB SPL pure tone which causes the highest excitation in the given channel. Only the signals in 158 out of the 300 channels are fed into the subsequent IHC/AN model. The channels were chosen according to their CF, such that the density of the CF was four channels per each critical band with ERB provided in (Glasberg, Moore, 1990).

Table 1. Equivalent rectangular bandwidth (ERB) of the simulated cochlear filters.

Level [dB SPL]	Characteristic frequency [kHz]					
	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
$\mathrm{ERB}_{\mathrm{GM}}$	38	52	79	133	241	456

Inner hair cells and auditory-nerve synapse model

The IHC/AN model transforms the BM displacement in each channel of the cochlear model to a probability of neurotransmitter release into the synapse of AN fibers. This process is modelled by means of the algorithms designed in various studies and implemented in (MAP, 2014).

The BM displacement, $w(t)$, is first transferred into the displacement of the IHC stereocilia, $u(t)$, given by

$$\tau_c\frac{\mathrm{d}u(t)}{\mathrm{d}t} + u(t) = \tau_c C_{\mathrm{cilia}}\frac{\mathrm{d}w(t)}{\mathrm{d}t}, \quad (3)$$

where τ_c is a time constant and C_{cilia} is a gain factor (Shamma *et al.*, 1986). This equation ensures that the stereocilia move in phase with the BM velocity at low frequencies and in phase with the BM displacement at high frequencies (Shamma *et al.*, 1986; Sumner *et al.*, 2002). Bending of the stereocilia opens ion channels. This is, according to (Sumner *et al.*, 2002), modelled as a change of the apical conductance, $G(u)$, given by a Boltzmann function

$$G(u) = G_{\mathrm{cilia}}^{\max}\left[1 + \exp\left(-\frac{u(t) - u_0}{s_0}\right) \times \left[1 + \exp\left(-\frac{u(t) - u_1}{s_1}\right)\right]\right]^{-1} + G_a, \quad (4)$$

where $G_{\mathrm{cilia}}^{\max}$ is the maximal conductance with all channels open, s_0, u_0, s_1, and u_1 are constants determining

the shape of the nonlinear Boltzmann function and passive conductance

$$G_a = G_{\text{cilia}}^{\max}\left[1+\exp\left(\frac{u_0}{s_0}\right)\times\left[1+\exp\left(\frac{u_1}{s_1}\right)\right]\right]^{-1}+G_0, \tag{5}$$

where G_0 is the resting conductance.

Ions entering the IHC change the membrane potential. The process is modelled by

$$C_m\frac{\mathrm{d}V(t)}{\mathrm{d}t}+G(u)(V(t)-E_t)+G_k(V(t)-E_k')=0, \tag{6}$$

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: $E_k' = E_k + E_t R_p/(R_t+R_p)$ (Shamma *et al.*, 1986; Sumner *et al.*, 2002).

Changes in the membrane potential open ion channels for calcium ions, which then mediate the release of neurotransmitter into the synaptic cleft. Calcium current I_{Ca} is given by

$$I_{\text{Ca}}(t) = G_{\text{Ca}}^{\max} m_{I_{\text{Ca}}}^3(t)(V(t)-E_{\text{Ca}}), \tag{7}$$

where E_{Ca} is the reversal potential for calcium, $G_{\text{Ca}}^{\max}$ is the maximal calcium conductance with all the ion channels open, and $m_{I_{\text{Ca}}}$ is the fraction of open calcium channels (Sumner *et al.*, 2002). The steady state value of the fraction of open ion channels $m_{I_{\text{Ca}},\infty}$ is simulated by a Boltzmann function (Sumner *et al.*, 2002)

$$m_{I_{\text{Ca}},\infty} = [1+\beta_{\text{Ca}}^{-1}\exp(-\gamma_{\text{Ca}}V(t))]^{-1}, \tag{8}$$

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

$$\tau_{I_{\text{Ca}}}\frac{\mathrm{d}m_{I_{\text{Ca}}}(t)}{\mathrm{d}t}+m_{I_{\text{Ca}}}(t)=m_{I_{\text{Ca}},\infty}, \tag{9}$$

where $\tau_{I_{\text{Ca}}}$ is a time constant (Sumner *et al.*, 2002). The concentration of calcium ions in the cell $[\text{Ca}^{2+}]$ is calculated from the calcium current I_{Ca} by

$$\tau_{[\text{Ca}]}\frac{\mathrm{d}[\text{Ca}^{2+}](t)}{\mathrm{d}t}+[\text{Ca}^{2+}](t)=I_{\text{Ca}}(t), \tag{10}$$

where $\tau_{[\text{Ca}]}$ is a time constant (Sumner *et al.*, 2002).

The concentration of calcium ions then defines the probability of release of the neurotransmitter into the synaptic cleft, $k(t)$. It is, according to (Meddis, 2006), calculated by

$$k(t) = z[\text{Ca}^{2+}]^3(t), \tag{11}$$

where z is a scalar constant converting the calcium concentration into the release rate.

Release of the neurotransmitter into the synaptic cleft causes firing of neural discharge into the synapse of AN fiber. Circulation of the neurotransmitter is, according to (Meddis, 1986), modelled by

$$\frac{\mathrm{d}q(t)}{\mathrm{d}t} = y(1-q(t))+xw(t)-k(t)q(t), \tag{12}$$

$$\frac{\mathrm{d}c(t)}{\mathrm{d}t} = k(t)q(t)-lc(t)-rc(t), \tag{13}$$

$$\frac{\mathrm{d}w(t)}{\mathrm{d}t} = rc(t)-xw(t). \tag{14}$$

The neurotransmitter is released from the immediate (q) store into the cleft (c) at rate $k(t)$. The release rate is mediated by the calcium concentration. Some of the transmitter in the cleft is lost at rate l. The remaining transmitter is taken back into reprocessing store (w) in the cell at rate r and then into (q) at rate x. Concentration of the neurotransmitter in the cleft (c) is the output of the IHC/AN model. It is proportional to the probability of spike firing into the AN fiber.

In this study, parameters of the algorithms were adjusted (see Table 2) in order to achieve a wide dy-

Table 2. 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
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
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

namic range. Figure 2 shows the input/output (I/O) functions obtained using pure tones with a frequency of 0.25, 1 and 4 kHz; the I/O functions were obtained in the model channels with CF equal to the frequency of the pure tones.

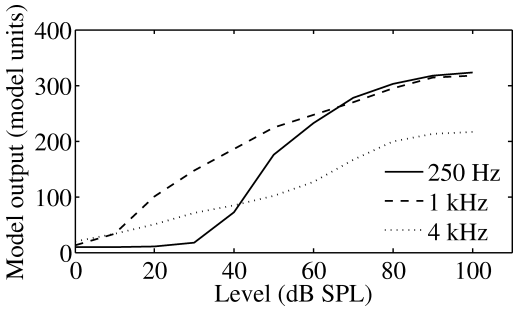

Fig. 2. Input/output (I/O) functions of the peripheral stage obtained using pure tones with a frequency of 0.25, 1 and 4 kHz. The functions were obtained in the channels of the peripheral stage with CF equal to the frequency of the pure tones.

2.2. Central stage

The central stage calculates the roughness from the signal processed by the peripheral stage (see Fig. 1). The parameters of the central stage were set in order to achieve an agreement with the subjective data concerning the roughness of 100% SAM tones. The subjective data – shown below in Subsec. 3.1.1 – were reproduced from Fig. 11.2 in (Fastl, Zwicker, 2007).

In the central stage, the first block called "envelope" estimates the envelope of the simulated neural signal in each k-th channel of the peripheral stage. The solid line in the top panel of Fig. 3 shows the simulated neural signal in the channel with CF = 1 kHz. The

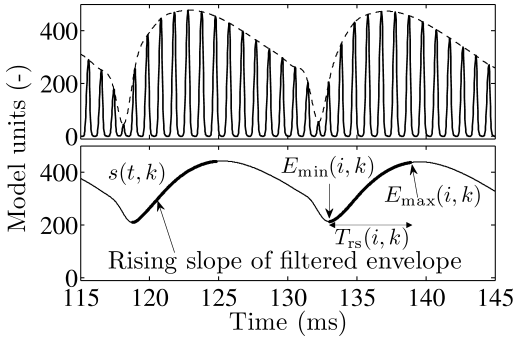

Fig. 3. Top panel: the auditory model output signal at CF of 1 kHz obtained in response to a 100% SAM tone with a frequency of 1 kHz, modulation frequency of 70 Hz, and level of 60 dB SPL. The dashed line shows the signal envelope. Bottom panel: the signal envelope smoothed by a 1st-order Butterworth filter with a cutoff frequency of 70 Hz. The tick lines show the rising parts of the smoothed envelope.

signal was 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 signal envelope – shown by the dashed line – was obtained by cubic spline interpolation of peaks in each half-wave of the signal's fine structure. The second block called "low-pass filter" processes the envelope estimated by the first block. It is a 1st-order Butterworth lowpass filter with a cutoff frequency of 70 Hz. This filter decreases the temporal resolution of the model, i.e., it decreases the envelope fluctuations at frequencies above 70 Hz. The bottom panel of Fig. 3 shows the filtered envelope of the neural signal plotted in the top panel.

The filtered envelope, $s(t,k)$, in each channel is then divided into successive time frames with duration, $T_{\rm fr}$, of 30 ms. Each time a frame is fed into three parallel branches: the first branch detects and processes the rising slopes of $s(t,k)$, the second branch calculates the root-mean-square (RMS) values of $s(t,k)$, and the third branch calculates the cross-correlation coefficients between the individual channels of $s(t,k)$. The outputs of the three branches are then used to calculate the overall roughness.

First branch: The block called "modulation features" detects the rising slopes of the filtered envelope, $s(t,k)$, (the tick lines in the bottom panel of Fig. 3). It then – separately for each i-th rising slope within the time frame – extracts its minimal, $E_{\min}(i,k)$, and maximal, $E_{\max}(i,k)$, value (see the bottom panel of Fig. 3). These values are used to calculate two modulation features of each rising slope: the modulation index, $M(i,k)$, given by

$$M(i,k) = \frac{E_{\max}(i,k) - E_{\min}(i,k)}{E_{\max}(i,k) + E_{\min}(i,k)}; \qquad (15)$$

and the duration of the rising slope, $T_{\rm rs}(i,k)$, given as the difference between the temporal positions of $E_{\max}(i,k)$ and $E_{\min}(i,k)$ in seconds. For 100% SAM tones, the dependence of roughness on the modulation frequency exhibits bandpass characteristic: for the SAM tone frequency $\geq$ 1 kHz, it is maximal at the modulation frequency of about 70 Hz; and for the SAM tone frequency $<$ 1 kHz, it is maximal at modulation frequencies $<$ 70 Hz (Fastl, Zwicker, 2007) (see Fig. 7). Parameter $T_{\rm rs}(i,k)$ allows to account for the lower slope of the roughness dependency, and modulation index $M(i,k)$ for the higher slope. The 70-Hz low-pass filter assures that the $M(i,k)$ decreases as the modulation frequency increases above 70 Hz. Moreover, the peripheral stage of the roughness model simulates the function of the cochlea, which can be modelled by a bank of bandpass filters. Since the bandwidth of the filters is small for low CF (Glasberg, Moore, 1990), the side components, $f_c - f_m$ and $f_c + f_m$, of a SAM tone with a frequency, f_c, and a modulation frequency, f_m, falls into the adjacent bands for

smaller values of f_m than in the case of SAM tones with higher f_c. The modulation index, $M(i,k)$, thus for SAM tones with a frequency < 1 kHz starts to decrease at modulation frequencies < 70 Hz.

To calculate roughness, $R_{\mathrm{sp}}(k)$, in each k-th channel, $M(i,k)$ and $T_{\mathrm{rs}}(i,k)$ are combined by[1, 2]

$$R_{\mathrm{sp}}(k) = \max_i \{M_{\mathrm{sat}}(i,k) \cdot F_{\mathrm{sat}}^{1.5}(i,k)\}, \tag{16}$$

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

$$M_{\mathrm{sat}}(i,k) = \begin{cases} M(i,k) & \text{if} \quad M(i,k) \le 0.5, \\ 0.5 & \text{if} \quad M(i,k) > 0.5, \end{cases} \tag{17}$$

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

$$F_{\mathrm{sat}}(i,k) = \begin{cases} \dfrac{1}{T_{\mathrm{rs}}(i,k)} & \text{if} \quad \dfrac{1}{T_{\mathrm{rs}}(i,k)} \le 96 \text{ s}^{-1}, \\ 119.5 & \text{if} \quad \dfrac{1}{T_{\mathrm{rs}}(i,k)} \ge 149 \text{ s}^{-1}, \\ \dfrac{3.7 \cdot 10^{-3}}{T_{\mathrm{rs}}^2(i,k)} + \dfrac{1.36}{T_{\mathrm{rs}}(i,k)} & \text{else.} \end{cases} \tag{18}$$

The values of the modulation index, $M(i,k)$, are often in a range between 0 and 0.5 in the model channels of CF close to the spectral components of a 100% SAM tone, and higher than 0.5 in the adjacent channels. Equation (17) thus decreases the predicted roughness in the channels which are not close to the CF of the spectral components of the 100% SAM tone. Equation (18) calculates the reciprocal of $T_{\mathrm{rs}}(i,k)$ and also shapes and limits its values (see Fig. 4) in order to predict the roughness of 100% SAM tones.

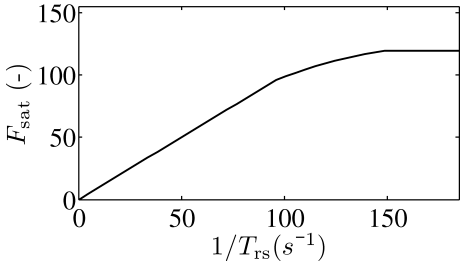

Fig. 4. Transformation function given by Eq. (18). The function transforms the duration, $T_{rs}(i,k)$, of the rising part of the filtered temporal envelope.

The calculated roughness, $R_{\mathrm{sp}}(k)$, in each channel k is then processed further. The quantiles (20%) of $R_{\mathrm{sp}}(k)$ over four adjacent channels are calculated as given by

$$R_{\mathrm{sp_q}}(n) = \underset{20\%}{\mathrm{quantile}}\{R_{\mathrm{sp}}(k), R_{\mathrm{sp}}(k+1), R_{\mathrm{sp}}(k+2), R_{\mathrm{sp}}(k+3)\}, \tag{19}$$

$$\forall k \in 1, 3, 5, ..., 158-3.$$

This gives $R_{\mathrm{sp_q}}(n)$ with n ranging from 1 to 78. The purpose of Eq. (19) is described on an example shown in Fig. 5, in which the solid line shows $R_{\mathrm{sp}}(k)$ of a 100% SAM tone with a frequency of 250 Hz, a level of 60 dB SPL, and a modulation frequency of 50 Hz. The dashed line in Fig. 5 shows $R_{\mathrm{sp_q}}(n)$ calculated from $R_{\mathrm{sp}}(k)$ using Eq. (19). The three dips in $R_{\mathrm{sp}}(k)$ (the solid line) corresponds to the three spectral components of the SAM tone, with a frequency of 200, 250, and 300 Hz. The roughness of a 250 Hz SAM tone is highest at a modulation frequency of about 43 Hz (see Fig. 7). If the modulation frequency of the SAM tone is increased above 43 Hz, $R_{\mathrm{sp}}(k)$ decreases mainly in the places of the dips; it may even increase in the adjacent channels. Equation (19) thus helps to ensure that the dependence of the predicted roughness of 100% SAM tones on the modulation frequency agrees with subjective data – mainly for SAM tones with a frequency < 1 kHz.

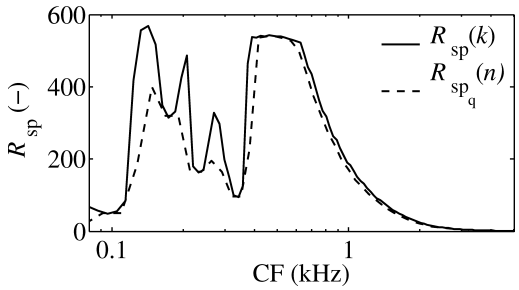

Fig. 5. Predicted roughness in individual channels, k or n, of the roughness model. The data are shown for a 100% SAM tone with a frequency of 250 Hz, a level of 60 dB SPL, and a modulation frequency of 50 Hz. The solid line shows the values of $R_{\mathrm{sp}}(k)$ given by Eq. (16). The dashed line shows the values of $R_{\mathrm{sp_q}}(n)$ given by Eq. (19).

Second branch: RMS values of the filtered envelope, $s(t,k)$, within one time frame are given by

$$x_{\mathrm{rms}}(k) = \sqrt{\frac{1}{T_{\mathrm{fr}}} \int_{T_{\mathrm{fr}}} s^2(t,k)\, \mathrm{d}t}, \tag{20}$$

where T_{fr} is duration of the time frame (30 ms). The RMS values are then averaged by

$$x_{\mathrm{m_r}}(n) = \mathrm{mean}\{x_{\mathrm{rms}}(k), x_{\mathrm{rms}}(k+1), x_{\mathrm{rms}}(k+2), x_{\mathrm{rms}}(k+3)\}, \tag{21}$$

$$\forall k \in 1, 3, 5, ..., 158-3.$$

[1] There is a mistake in (Vencovský, 2014a; 2014b). The exponent 1.5 should be placed above the term F_{sat} calculated from the duration of the rising slopes of the filtered envelope.

[2] The relation resembles the Fastl and Zwicker roughness model (Fastl, Zwicker, 2007, Eq. (11.1)).

The averaging decreases the number of channels to 78 (equal to the number of channels of $R_{\mathrm{sp_q}}(n)$). The channels of RMS below a specific threshold value are excluded from the roughness prediction by setting its RMS value to 0 as given by

$$x_{\mathrm{m_r}}(n) = \begin{cases} x_{\mathrm{m_r}}(n) & \text{if } x_{\mathrm{m_r}} \geq 0.1 x_{\max} \\ 0 & \text{else} \end{cases}$$
$$x_{\max} = \max_n \{x_{\mathrm{m_r}}(n)\}. \tag{22}$$

Third branch: The last parameter which is used to predict roughness are the cross-correlation coefficients, $c(n)$, between the filtered envelopes, $s(t,k)$. The cross-correlation coefficients are calculated between the first and the last channel of four adjacent channels in $s(t,k)$ as given by

$$c(n) = \mathrm{corr}\{s(t,k), s(t,k+3)\}$$
$$\forall k \in 1,3,5,...,158-3, \tag{23}$$
$$c(n) = \begin{cases} c(n) & \text{if } \quad c(n) \geq 0 \\ 0 & \text{else.} \end{cases}$$

The cross-correlation coefficients are employed in order to decrease the predicted roughness of noise stimuli. The roughness models designed in (Aures, 1985; Daniel, Weber, 1997) also employ cross-correlation.

Roughness calculation: The roughness in one time frame of $s(t,k)$ is calculated by

$$R = \sqrt{b} \sum_n \frac{h(n) R_{\mathrm{sp_q}}(n) c(n) x_{\mathrm{m_r}}(n)}{\sum_n x_{\mathrm{m_r}}(n)}, \tag{24}$$

where $R_{\mathrm{sp_q}}(n)$, $x_{\mathrm{m_r}}(n)$, and $c(n)$ are the aforementioned parameters, $h(n)$ is a weighting function, and L is the number of the channels where $x_{\mathrm{m_r}}(n) > 0$. The calculated RMS values, $x_{\mathrm{m_r}}(n)$, of $s(t,k)$ are used in Eq. (24) in order to emphasise the predicted roughness in the channels where $x_{\mathrm{m_r}}(n)$ is high. The term $\sqrt{b}$ and the weighting function, $h(n)$, ensure that the predicted roughness quantitatively fits the subjective roughness (in aspers) of 100% SAM tones (Fastl, Zwicker, 2007). Figure 6 shows the weighting function, $h(n)$. For the model channels with CF closest to the frequency of 0.125, 0.25, 0.5, 1, 2, 4, 8, and 16 kHz, the values of $h(n)$ were set to 1.35, 1.22, 1.01, 1, 0.72, 0.41, 0.33, and 0.25, respectively. The values of $h(n)$ in the remaining channels were then calculated by a cubic spline interpolation of the experimentally set values. Since the central stage predicts the roughness in successive 30-ms time frames, the overall predicted roughness is calculated as the median.

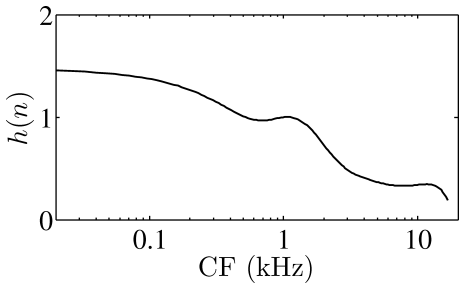

Fig. 6. Weighting function applied in Eq. (24) to predict quantitatively similar data of roughness in aspers as show the subjective data for 100% SAM tones (see Fig. 7).

3. Verification of the roughness model

The roughness model was verified using various types of acoustic stimuli. The predicted data were compared with subjective data reproduced from literature.

3.1. Roughness of sinusoidally amplitude-modulated tones

SAM tones are given by

$$p(t) = A\,[1 + m \cdot \cos(2\pi f_m t)] \cos(2\pi f_c t), \tag{25}$$

where A is the amplitude, m is the modulation index, f_m is the modulation frequency, and f_c is the tone frequency.

3.1.1. Dependence on the modulation frequency

Figure 7 shows the roughness of a 100% SAM tone as a function of the modulation frequency: the dashed

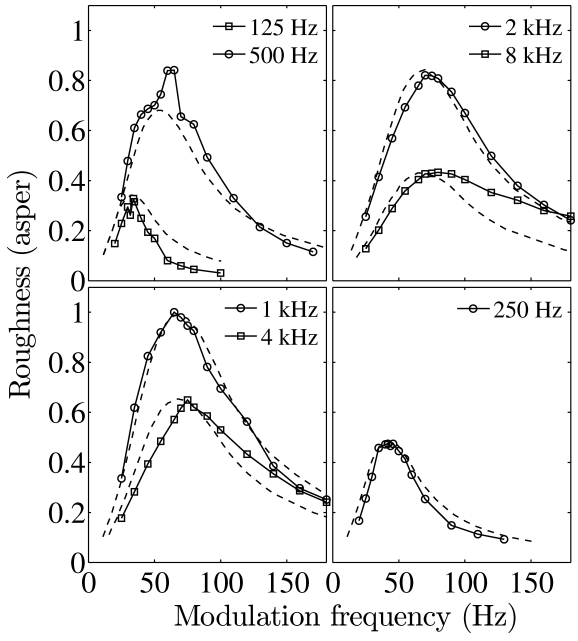

Fig. 7. Dependence of the roughness of a 100% SAM tone on the modulation frequency. The dashed lines show the results of listening tests reproduced from (Fastl, Zwicker, 2007). Markers connected by solid lines show the predicted roughness. The level of the SAM tone was 60 dB SPL and its frequency is given in the upper right corner of each panel.

lines show the perceived roughness, reproduced from (Fastl, Zwicker, 2007), measured by Aures (1984), and the markers connected by solid lines show the predicted data. Frequencies of the SAM tone are given in the upper right corner of each panel. The level of the SAM tone was 60 dB SPL. The central stage of the roughness model was designed and its parameters were set in order to achieve an agreement between the predicted and subjective data. Spearman's and Pearson's correlations between the data are significant; the lowest is the Spearman's correlation for the 4-kHz SAM tone (rho = 0.85, $p = 1.3 \cdot 10^{-4}$); however, for the remaining data: rho ≥ 0.96, $p \leq 7.72 \cdot 10^{-8}$.

3.1.2. Dependence on the modulation index

Figure 8 shows the roughness of a 1-kHz SAM tone as a function of the modulation index. The level of the tone was 70 dB SPL and the modulation frequency was 70 Hz. The dashed, dash-dotted, and dotted lines show the power-law relation used in (Daniel, Weber, 1997): $R = 1.36 \cdot m^p$, where m is the modulation index (0-1) and p is the exponent. The value of the exponent was estimated to be equal to 1.6 (Fastl, Zwicker, 2007), 2 (Terhardt, 1968), and 1.5 (Vogel, 1975). The predicted roughness is plotted as circles connected by solid lines. The best agreement between the model predictions and the power-law approximations is in the range of m between 0 and 0.8. Spearman's and Pearson's correlations calculated between the predicted and subjective data are significant: rho ≥ 0.98, $p \leq 1.12 \cdot 10^{-5}$.

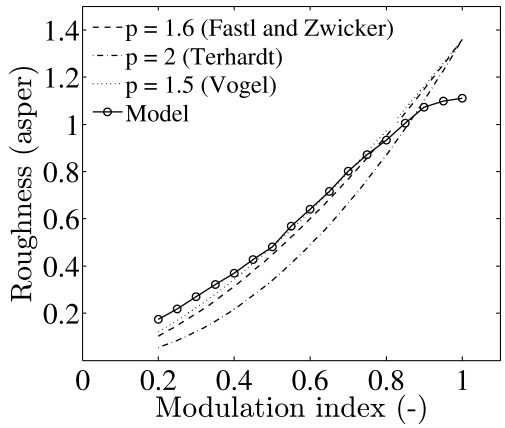

Fig. 8. Roughness of a SAM tone with a frequency of 1 kHz, a level of 70 dB SPL, and a modulation frequency of 70 Hz, plotted as a function of the modulation index. The dashed, dashed-dot, and dotted lines were obtained using relation $R = 1.36 \cdot m^p$, where m is the modulation index and p is an exponent equal to 1.6, 2, and 1.5, respectively. The values of p were estimated from the subjective experimental data: Fastl and Zwicker (2007) estimated $p = 1.6$, Terhardt (1968) $p = 2$, and Vogel (1975) $p = 1.5$.

3.2. Roughness of stimuli with envelopes that are not sinusoidal

3.2.1. Pseudo amplitude-modulated tones

The amplitude spectrum of a 100% SAM tone contains three spectral components. The central component with frequency f_c (frequency of the modulated tone) and side components with frequencies $f_c - f_m$ and $f_c + f_m$ (f_m is the modulation frequency), and amplitudes equal to half of the amplitude of the central component. Setting the starting phases of the side components to zero and altering the starting phase of the central component, ϕ, creates pseudo amplitude-modulated (pAM) tones. If ϕ is nonzero, the time waveform envelope of the pAM tone is not sinusoidal. pAM tones with opposite ϕ (e.g. $-\pi/6$ and $\pi/6$) have similar waveform envelopes but differ in their temporal fine structure (Pressnitzer, McAdams, 1999).

Figure 9 shows the roughness of pAM tones as a function of the starting phase absolute value, $|\phi|$, of the central component; panels in the top row show the perceived roughness reproduced from (Pressnitzer, McAdams, 1999), and panels in the bottom row show the predicted roughness; dashed lines show the data for negative values of ϕ, solid lines – for positive values. The level of the pAM tones was 60 dB SPL, the frequency of the central component was equal to f_c and the modulation frequency was equal to f_m, given in the upper right corner of each panel. Pressnitzer and McAdams (1999) used the method of pairwise comparison to measure the roughness of pAM tones; and then they used the Bradley-Terry-Luce (BTL) method to transform the comparison judgments into the normalised linear interval scale. The standard deviations of the roughness data were estimated using the bootstrap technique (Pressnitzer, McAdams, 1999). The predicted roughness was normalised by the maximal value of the data shown in each panel. This was done in order to better visualize the data. The left ordinate of each panel shows the normalised roughness and the right ordinate (gray) shows the predicted roughness in aspers. Since the subjective method gave the data on an interval scale, the subjective and predicted data should be compared as ranking data, which means that Spearman's correlations between the predicted and subjective data are the most important. However, Spearman's and also Pearson's correlations calculated between the data significant for all f_c: rho ≥ 0.79, $p \leq 0.048$.

The roughness model accounts for the subjective data because of the algorithms used in the central stage – after processing the pAM tones by the peripheral stage, the rising slopes of the signal envelopes are for the pAM tones with $+\phi$ values steeper than for the pAM tones with $-\phi$ values.

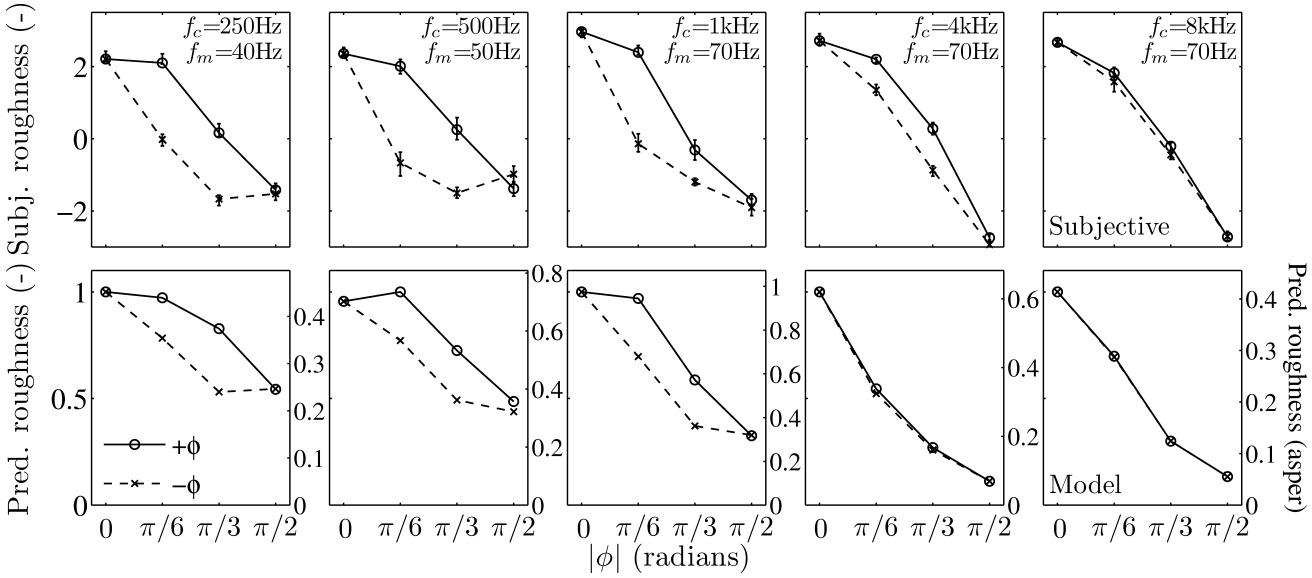

Fig. 9. Roughness of pAM tones as a function of the starting phase absolute value, $|\phi|$ – the dashed and solid lines show the data for negative and positive values of ϕ, respectively. The top panels show the mean values and standard errors of the mean of subjective data reproduced from (Pressnitzer, McAdams, 1999). The bottom panels show the predicted roughness normalised by the maximal predicted value in each panel. The right ordinates show the predicted roughness in aspers. The frequency, f_c, and the modulation frequency, f_m, of the pAM tones are given in the panels.

3.2.2. Sawtooth and reversed stimuli

To study effects of the waveform envelope asymmetry on roughness, Pressnitzer and McAdams (1999) used "sawtooth" and "reversed" stimuli (also called "ramped" and "damped", respectively). The "sawtooth" stimuli are amplitude-modulated tones given by

$$p_{st}(t) = \left(1 + \frac{m \cdot E_{st}(t)}{\max\{E_{st}(t)\}}\right) \cos\left(2\pi f_c t - \frac{\pi}{2}\right), \quad (26)$$

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

$$E_{st}(t) = \sum_{n=1}^{N} \frac{1}{n} \cos\left(2\pi n f_m t - \frac{\pi}{2}\right), \quad (27)$$

where f_m is the modulation frequency and N is a number of harmonics. The value of N is set to fulfill the condition $N \cdot f_m \leq 0.5\mathrm{ERB}(f_c)$. The condition ensures that the spectral components of $x_{st}(t)$ are within one critical band of bandwidth equal to the ERB of auditory filters, which was estimated in (Glasberg, Moore, 1990). The "reversed" stimuli – time reversals of the "sawtooth" stimuli – can be generated by inverting the sign (to $+\pi/2$) in the argument of cosine functions in Eq. (26) and Eq. (27).

Figure 10 shows the roughness of the "sawtooth" (crosses connected by dashed lines) and "reversed" (circles connected by solid lines) stimuli: panels in the top row show the perceived roughness, as reproduced

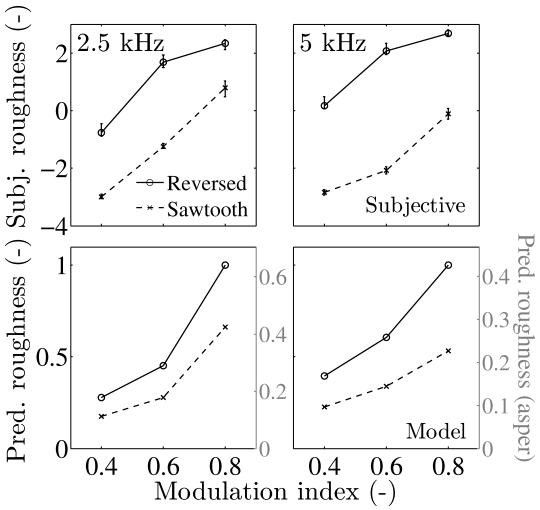

Fig. 10. Roughness of tones with asymmetrical temporal envelope: crosses connected by dashed lines show the roughness of "sawtooth" stimuli, and circles connected by solid lines show the roughness of "reversed"stimuli. The top panels show the mean and standard errors from the mean of subjective data reproduced from (Pressnitzer, McAdams, 1999). The bottom panels show the predicted roughness normalised by the corresponding maximal value for each panel, which gave the maximal predicted roughness of 1. The frequency, f_c, of the stimuli was 2.5 and 5 kHz, the modulation frequency, f_m, was 70 Hz and the level was 60 dB SPL.

from (Pressnitzer, McAdams, 1999), obtained using the same method as for the pAM tones, and pan-

els in the bottom row show the predicted roughness. The data are plotted as a function of the modulation index. The level of the stimuli was 60 dB SPL, the modulation frequency was 70 Hz, the tone frequency was 2.5 and 5 kHz, and the number of harmonics in the modulation signal, N, was 2 and 4, respectively, for the panels in the left and right column. The predicted roughness, which is shown by the panels in the bottom row of Fig. 10, was normalised to its maximal value, separately for the data in each panel. The left ordinates show the normalised roughness, and the right ordinates (gray) show the predicted roughness in aspers. The roughness model predicted – in agreement with the subjective data – more roughness for the "reversed" stimuli. However, the contrast between both stimuli is not as high as in the subjective data for $f_c = 5$ kHz, where the "sawtooth" stimulus with a modulation index of 0.8 was judged less rough than the "reversed" stimulus with a modulation index of 0.4. A similar discrepancy is in the data for $f_c = 2.5$ kHz – for the "sawtooth" stimulus with a modulation index of 0.8 and the "reversed" stimulus with a modulation index of 0.6. However, Spearman's and Pearson's correlations calculated between the predicted and subjective data are significant: rho ≥ 0.85, $p \leq 0.032$.

3.3. Roughness of sinusoidally frequency-modulated tones

SFM tones are given by

$$p(t) = A \cdot \sin\left[2\pi f_c t - \frac{\Delta f}{f_m}\cos(2\pi f_m t)\right], \qquad (28)$$

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) used the method of magnitude estimation to measure the roughness of SFM tones. He presented the listeners with a pair of stimuli, standard and comparison (stimulus under test), and asked them to assign a number reflecting the roughness of the comparison relative to the roughness of the standard. A 100% SAM tone with a frequency of 1.6 kHz, a level of 60 dB SPL, and a modulation frequency of 70 Hz was used as the standard.

3.3.1. Dependence on the modulation frequency

Figure 11 shows the roughness of SFM tones as a function of the modulation frequency, f_m: crosses connected by dashed lines show medians and quartiles of the subjective data reproduced from (Kemp, 1982), and circles connected by solid lines show the predicted roughness. The SFM tones had a frequency of 1.6 kHz, a level of 60 dB SPL and a frequency deviation of ±800 Hz. The data are plotted as relative roughness to the roughness of a SFM tone with a modulation frequency of 70 Hz. The right ordinate shows the predicted roughness in aspers. The subjective and predicted data agree within a range of quartiles. Spearman's and Pearson's correlations are significant: rho ≥ 0.94, $p < 1 \cdot 10^{-4}$.

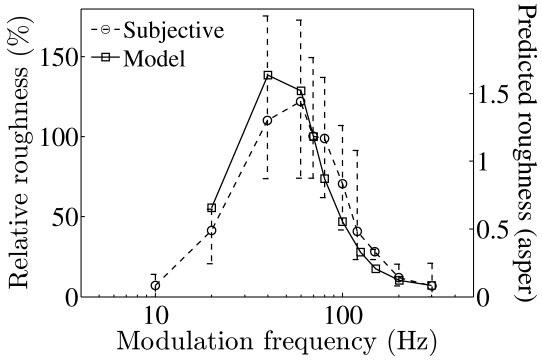

Fig. 11. Dependence of roughness of SFM tones on the modulation frequency, f_m. Crosses connected by dashed lines show medians and quartiles of the subjective data reproduced from (Kemp, 1982). Circles connected by solid lines show the predicted roughness. The frequency of the SFM tones was 1.6 kHz, the modulation index, Δf, was 800 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 modulation frequency of 70 Hz. The right ordinate shows the predicted roughness in aspers.

3.3.2. Dependence on the frequency deviation

Figure 12 shows the roughness of SFM tones as a function of the frequency deviation: crosses con-

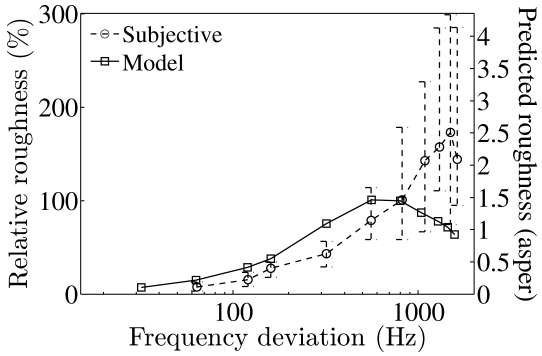

Fig. 12. Dependence of roughness of SFM tones on the frequency deviation (modulation index), Δf. Crosses connected by dashed lines show medians and quartiles of the subjective data reproduced from (Kemp, 1982). Circles connected by solid lines show the predicted roughness. The frequency of the SFM tones 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. The right ordinate shows the predicted roughness in aspers.

nected by dashed lines show the medians and quartiles of subjective data reproduced from (Kemp, 1982), and circles connected by solid lines show the predicted roughness. The data are plotted as relative roughness to the roughness of a SFM tone with a frequency deviation of 800 Hz. The frequency of the SFM tones was 1.6 kHz, the modulation frequency was 70 Hz, and the level was 60 dB SPL. For the frequency deviation up to 800 Hz, the predicted data qualitatively agree with the subjective data. However, above 800 Hz the predicted roughness starts to decrease. As a result, Spearman's and Pearson's correlations between the predicted and subjective data are not significant at the 0.05 level: Spearman's correlation is rho = 0.49, $p = 0.15$ and Pearson's correlation is rho = 0.63, $p = 0.53$.

3.4. Roughness of unmodulated bandpass noise

Aures (1985) measured the roughness of unmodulated bandpass noises. During the measurements, he asked the listeners 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, to be perceived with the same roughness as the bandpass noise. Figure 13 shows the roughness of unmodulated bandpass noise as a function of its bandwidth:

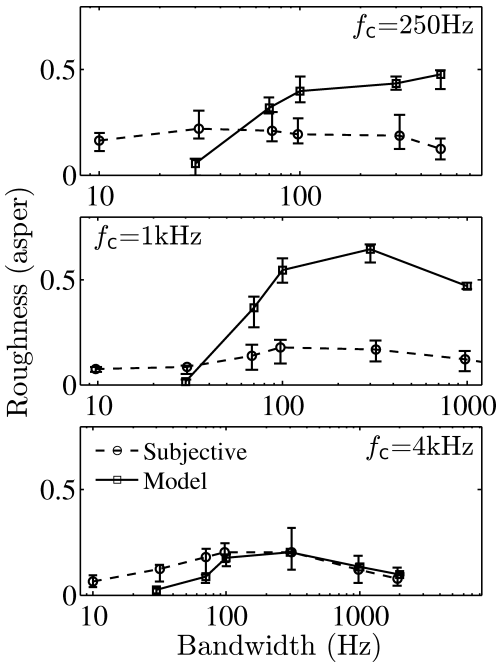

Fig. 13. Roughness of unmodulated bandpass noises. Circles connected by dashed lines show the medians and quartiles of subjective data adapted from (Aures, 1985). Squares connected by solid lines show the medians and quartiles of the predicted roughness obtained for ten realisations of the stimuli. Abscissa shows the noise bandwidth. Each panel is for a bandpass noise with centre frequency f_c and a level of 70 dB SPL.

circles connected by dashed lines show the medians and quartiles of the subjective data, and squares connected by solid lines show the medians and quartiles of the predicted roughness. The subjective data measured by Aures were transformed into aspers using the relation $R = 1.36 \cdot m^{1.6}$. Each panel shows the data for unmodulated bandpass noise with frequency equal to f_c and level of 70 dB SPL. The predicted roughness agrees with the subjective data only for $f_c = 4$ kHz; for the remaining stimuli, the predicted roughness is much larger than the subjective data. Since the disagreement is very obvious, Spearman's and Pearson's correlations between the predicted and subjective data were not calculated.

3.5. Roughness of harmonic intervals of the chromatic scale

Vassilakis (2005) measured the roughness of two tone stimuli (dyads) composed of harmonic complex tones. Each tone had six harmonics with amplitude A_n of the n-th harmonic given by $A_n = A_1/n$. The fundamental frequency of the lower tone in the dyads was set to middle C (C4, fundamental frequency 256 Hz, equal temperament[3]). The level of the stimuli was 75 dB SPL. Listeners rated the roughness on a continuous scale ranging from 0 (not rough) to 42 (rough) (Vassilakis, 2005).

Figure 14 shows the roughness of the dyads as a function the fundamental frequency of the higher tone: squares connected by dashed lines show the mean values and standard deviations of the mean of the subjective data reproduced from (Vassilakis, 2005),

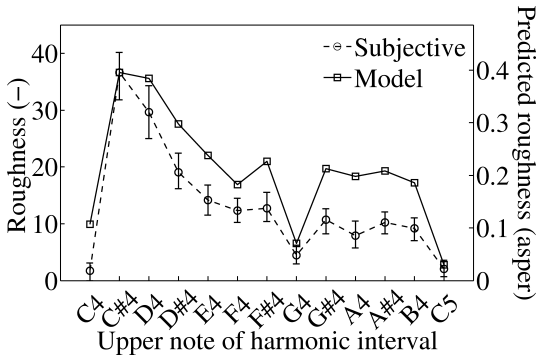

Fig. 14. Roughness ratings of two tone stimuli composed of harmonic complex tones. Circles connected by dashed lines show the mean values of the subjective data, reproduced from (Vassilakis, 2005), across ten listeners. Squares connected by solid lines show the predicted roughness normalised to its maximal value and scaled by the maximal value of the subjective roughness, to be in the range of the used rating scale.

[3]Equal tempered C4 with a fundamental frequency of 256 Hz corresponds to A4 with a fundamental frequency of 430.5 Hz.

and circles connected by solid lines show the predicted roughness. In order to compare the data visually, the predicted roughness was normalised to its maximal value and scaled by the maximal value of the subjective roughness. The predicted data agree with the subjective data – Spearman's correlation is rho $= 0.92$, $p = 0$, and Pearson's correlation is rho $= 0.94$, $p = 1.5 \cdot 10^{-6}$.

4. Conclusion

The study presented a new roughness model and verified it using various types of acoustic stimuli: sinusoidally amplitude-modulated (SAM) tones, stimuli with envelopes that are not sinusoidal, sinusoidally frequency-modulated (SFM) tones, unmodulated bandpass noise stimuli, and two tone stimuli composed of harmonic complex tones. The perceived roughness of the stimuli was reproduced from literature. The model performed well for most of the used stimuli. The performance was poor for unmodulated bandpass noise where the model predicted more roughness than was shown by the subjective data.

The central stage of the roughness model predicts roughness from the envelope – its rising parts – of the simulated neural signal at the output of the peripheral stage. This way of prediction allowed the model to account for the roughness of stimuli with envelope that is not sinusoidal. After processing by the peripheral stage, the roughness differences between these stimuli are reflected in the shape of the temporal envelopes – for the stimuli with more roughness, the rising parts of the temporal envelope are steeper. These effects are poorly accounted for by the roughness models published in literature (e.g. Daniel, Weber, 1997; Leman, 2000).

The inability of the roughness model to predict the roughness of unmodulated bandpass noises may indicate another result of this study. Roughness models (e.g. Daniel, Weber, 1997; Leman, 2000), which employ models of the peripheral ear often contain a set of weighting filters allowing the models to account for the dependence of roughness on modulation frequency, which for SAM and SFM tones exhibits bandpass characteristic. The designed roughness model does not contain such a set of weighting filters. Instead, it was designed to rely on its peripheral stage with the physical cochlear model. Since the limited frequency resolution of the peripheral ear is suggested to contribute to roughness perception, this approach may be more reasonable. Despite the strong amplitude fluctuations at the output of the peripheral stage in response to unmodulated bandpass noises, its perceived roughness is small. This may imply that roughness of these stimuli is accounted for at higher stages of the hearing system. This would mean that the presented modelling approach allows the separation of stimuli whose roughness is more determined by the function of the peripheral ear from those whose roughness is more determined at higher stages of the hearing system.

Acknowledgments

I thank Aleš Vetešník for sending parameters of the cochlear model, Jesko L. Verhey, Arne Oetjen and two anonymous reviewers for useful comments on earlier versions of the paper. This research was supported by Grant Agency of the Czech Technical University in Prague, grant No. SGS14/204/OHK3/3T/13 and 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.

References

1. Aures W. (1984), *Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse* [in German], Ph.D. Thesis, TU München.
2. Aures W. (1985), *Ein Berechnungsverfahren der Rauhigkeit* [in German], Acustica, **58**, 268–281.
3. Daniel P., Weber R. (1997), *Psychoacoustical Roughness: Implementation of an Optimized Model*, Acustica, **83**, 113–123.
4. Fastl H., Zwicker E. (2007), *Psychoacoustics: Facts and Models*, 3rd. Ed., Springer, Berlin, Heidelberg, New York.
5. Glasberg B.R., Moore B.C.J. (1990), *Derivation of auditory filter shapes from notched-noise data*, Hearing Research, **47**, 103–138.
6. Hartmann W.M., Hnath G.M. (1982), *Detection of Mixed Modulation*, Acustica, **50**, 297–312.
7. Helmholtz H.L.F. von (1895), *Sensation of tone as a physiological basis for the theory of music*, 3rd Ed., Longmans, Green, and Co., London, New York.
8. Kin M.J., Dobrucki A.B. (1998), *Perception of mixed modulation for single components in harmonic complex for high modulating frequencies*, Archives of Acoustics, **23**, 379–390.
9. Kohlrausch A., Hermes D., Duisters R. (2005), *Modeling roughness perception for sounds with ramped and damped temporal envelopes*, Proceedings of Forum Acusticum, pp. 1719–1724, Budapest.
10. Leman M. (2000), *Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (SIM)*, Proceedings of the COST G-6 Conference on Digital Audi Effects (DAFX–00), pp. DAFX 1–DAFX 6, Verona.
11. Kemp S. (1982), *Roughness of Frequency-Modulated Tones*, Acustica, **50**, 126–133.
12. Mammano F., Nobili R. (1993), *Biophysics of the cochlea. I: Linear approximation*, Journal of the Acoustical Society of America, **93**, 3320–3332.

13. Mathes R.C., Miller R.L. (1947), *Phase Effects in Monaural Perception*, Journal of the Acoustical Society of America, **19**, 780–797.

14. Matlab Auditory Periphery (MAP1.14), University of Essex, Hearing Research Laboratory: http://www.essex.ac.uk/psychology/department/hearinglab/modelling.html.

15. Meddis R. (2006), *Auditory-nerve first-spike latency and auditory absolute threshold: a computer model*, Journal of the Acoustical Society of America, **119**, 406–417.

16. Meddis R. (1986), *Simulation of mechanical to neural transduction in the auditory receptor*, Journal of the Acoustical Society of America, **79**, 702–711.

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

18. Miśkiewicz A., Rogala T., Szczepańska-Antosik J. (2007), *Perceived roughness of two simultaneous harmonic complex tones*, Archives of Acoustics, **32**, 737–748.

19. Nobili R., Mammano F. (1996), *Biophysics of the cochlea. II: Stationary nonlinear phenomenology*, Journal of the Acoustical Society of America, **99**, 2244–2255.

20. Nobili R., Vetešník A., Turicchia L., Mammano F. (2003), *Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model*, Journal of the Association for Research in Otolaryngology, **4**, 478–494.

21. Plomp R., Steeneken H.J.M. (1968), *Interference between two simple tones*, Journal of the Acoustical Society of America, **43**, 883–884.

22. Pressnitzer D., McAdams S. (1999), *Two phase effects in roughness perception*, Journal of the Acoustical Society of America, **105**, 2773–2782.

23. 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*, Journal of the Acoustical Society of America, **80**, 133–145.

24. 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*, Journal of the Acoustical Society of America, **111**, 2178–2188.

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

26. Terhardt E. (1968), *Über akustische Rauhigkeit und Schwankungsstärke* [in German], Acustica, **20** 215–224.

27. Vassilakis P.N. (2005), *Auditory roughness as means of musical expression*, [in:] *Selected Reports in Ethnomusicology, UCLA Ethnomusicology Publications*, pp. 119–144, Los Angeles.

28. Vassilakis P.N. (2001), *Perceptual and physical properties of amplitude fluctuation and their musical significance*, Ph.D. Thesis, University of California.

29. Vencovský V. (2014), *Modeling roughness perception using a model of cochlear hydrodynamics*, Proceedings of Forum Acusticum, Krakow.

30. Vencovský V. (2014), *Modeling roughness perception for complex stimuli using a model of cochlear hydrodynamics*, Proceedings of International Symposium of Musical Acoustics (ISMA 2014), pp. 483–488, Le Mans.

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

32. 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*, Journal of Sound and Vibration, **332**, 3893–3904.