7 - 12 SEPTEMBER 2014, KRAKÓW
FORUM ACUSTICUM

Modeling roughness perception using a model of cochlear hydrodynamics

Václav Vencovský
Musical Acoustics Research Center, Academy of Performing Arts in Prague, Malostranské nám. 12, 118 00, Prague, Czech Republic

Summary
A term roughness is often used in conjunction with acoustic stimuli whose envelope or frequency rapidly fluctuates in time. It was first introduced by von Helmholtz to describe harsh, rattling and often unpleasant sound sensation usually accompanying these stimuli. This study introduces a computational roughness model which objectively measure an amount of the perceived roughness. The roughness model contains an auditory model transforming the input acoustic stimulus into the simulated neural signal. The perceived roughness is calculated from time envelope of the transformed signal. The predicted roughness is compared with subjective results published in literature. The model successfully predicts roughness of sinusoidaly amplitude modulated (SAM) and sinusoidaly frequency modulated (SFM) tones. It can predict the perceived roughness when relative phase between spectral components of sound stimuli is changed and the perceived roughness of unmodulated bandpass noise.

PACS no. 43.66.Lj, 43.66.Ba

1. Introduction

Spectral components of the acoustic stimuli may produce beats on the basilar membrane (BM) when they fall into the same critical band. The beats of a frequency between 20 and 300 Hz cause a harsh and rattling sound sensation which von Helmholtz described by a term roughness [1]. The dependence of roughness on the parameters of various types of acoustic stimuli (e.g. on the modulation frequency of sinusoidally amplitude modulated (SAM) tones) has been investigated in the last decades by means of listening tests [2, 3, 4, 5]. It was proposed that roughness is linked to envelope fluctuations after auditory filtering [1, 4, 6]. This proposal led to the development of roughness models estimating roughness from the signal after filtering [8, 9, 10, 11]. Pressnitzer and McAdams [4] estimated roughness of stimuli with temporally assymetrical envelopes. They concluded that not only rms value but also the shape of the envelope after auditory filtering have to be taken into account to predict roughness. An abrupt raise and slow decay of the signal envelope after auditory filtration was observed in the stimuli of higher roughness.

A model estimating roughness not just from the rms value of the envelope after auditory filtering as some of the roughness models do but also from duration of the rising part of the envelope is introduced in this study. Since the auditory filtering is essential for the roughness perception, a model of the basilar membrane (BM) response and cochlear hydrodynamics was used in the roughness model. Roughness predicted by the model is compared with the results of listening tests published in literature.

2. Roughness model

2.1. Peripheral stage

The auditory model is composed of an outer- and middle-ear model, a cochlear frequency selectivity model and an inner hair cells (IHCs) model (see Fig. 1)

2.1.1. Outer- and middle-ear model

The outer and middle-ear model simulates transduction of the sound pressure waveform to the stapes displacement. The used model was developed in the Essex Hearing Laboratory as a part of the Matlab Auditory Periphery (MAP) [12]. The model version MAP1.14 was used. The model is composed of three 1st-order Butterworth filters. The first filter has a lower and higher a cut-off frequency of 1 kHz and 4 kHz, respectively. It approximates the transfer function of the auditory canal. The middle-ear transfer function measured by Huber [13] is approximated by

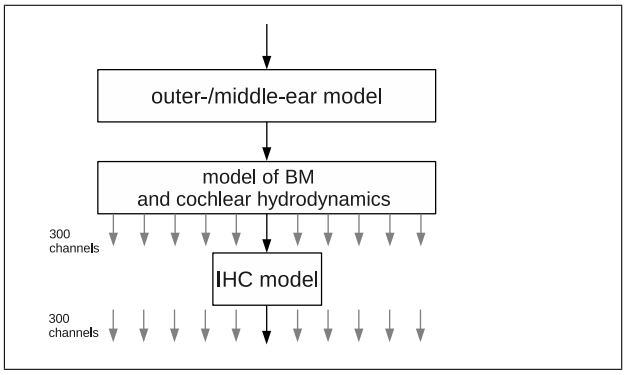

Figure 1. The auditory model diagram

a low-pass filter with a cut-off frequency of 50 Hz and a high-pass filter with a cut-off frequency of 1 kHz. A scalar of 45×10^{-9} is applied after the filters.

2.1.2. The Cochlear model

The second block of the auditory model is a model of the BM response and cochlear hydrodynamics designed by Mammano and Nobili [14, 15]. You can download a time domain MATLAB (Mathworks) implementation of the model from this website [16]. It is a two dimensional model designed with the realistic dimensions of the human cochlea and BM. The BM is modeled as an array of 300 oscillators. Characteristic frequencies of the oscillators are distributed between 27 and 16 875 Hz. An author of this study decreased damping of the oscillators in order to increase the frequency selectivity of the cochlear model. The output signal of the model represents the BM displacement.

2.1.3. The inner hair cell model

The IHC model transforms the BM displacement in each channel of the cochlear model to the probability of spike generation in the auditory nerve fiber. This process consists of successive steps. Each step is modeled by algorithms implemented in MAP (MAP1.14) [12]. The BM displacement is transmitted to the displacement of the IHC cilia. This transduction is modeled by an algorithm designed by Shamma *et al.* [17]. The algorithm is followed by two algorithms of Sumner *et al.* [18]. The first algorithm simulates a change of the apical conductance of the IHC for ions. The second algorithm simulates a change of the membrane potential caused by a change of the IHC apical conductance. The membrane potential affects a fraction of open ion channels for calcium ions. The concentration of calcium ions in the cell controls the release of neurotransmitter into the synaptic cleft. Algorithms modeling the calcium concentration in the IHC were designed by Meddis [19]. The release of neurotransmitter into the synaptic cleft may lead to a generation of action potential in the auditory nerve fiber. This process is modeled by a probabilistic model designed by Meddis [20].

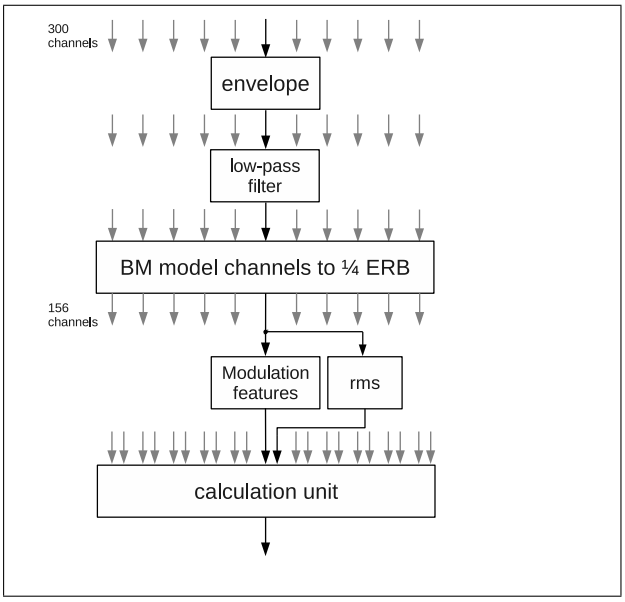

Figure 2. Block diagram of the central stage of the roughness model

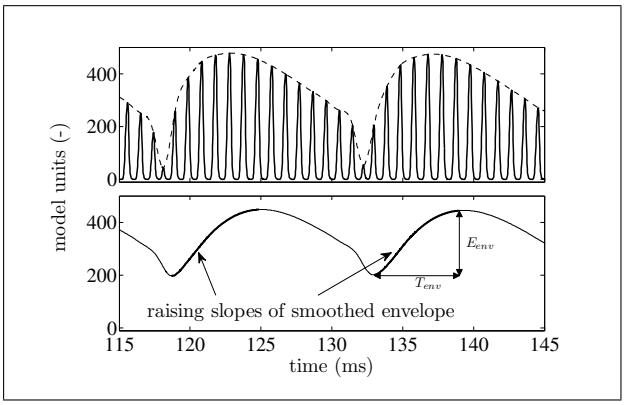

Figure 3. The top panel: The solid line is a time course of the auditory model output signal in a channel of CF = 1 kHz in response to a 100%, 1-kHz SAM tone modulated with a modulation frequency of 70 Hz. The dashed line is the signal envelope. The bottom panel: The envelope signal from the upper plot smoothed by a 2nd-order Butterworth filter. The tick lines are raised parts of the smoothed envelope from which are extracted the modulation features T_{env} and E_{env} used to estimate roughness.

2.2. Central stage

The central stage of the roughness model processes the output signal of the auditory model and predicts roughness of the analyzed sound stimuli. A block diagram of the central stage is in Fig. 2.

The first block calculates envelope of the signal in each of 300 channels of the auditory model. The algorithm detects peaks in the time course of the output signal and interpolates it by a cubic spline function. Fig. 3 (solid line in the top panel) shows a signal from the model output channel (CF = 1 kHz) obtained in response to a 100% AM tone of a frequency of 1 kHz, a level of 60 dB SPL and a modulation frequency of

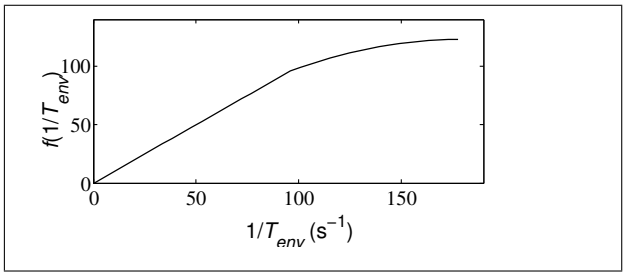

Figure 4. Transformation function $f()$ applied to the modulation feature T_{env}.

70 Hz. The dashed line in the same panel represents the signal envelope calculated as a cubic spline interpolation of peaks detected in the signal finestructure (peak of each half-wave of the model output signal). The bottom panel of Fig. 3 shows the signal envelope filtered by a 2nd-order Butterworth filter with a cut-off frequency of 90 Hz. This filter assures the decrease of roughness for the modulation frequencies above approx. 70 Hz as was observed in the listening tests [5]. The second block of the central stage decreases a number of channels from 300 to 156. The block averages the signals in adjacent channels with of CFs within a range of 1/4 of ERB given by ERB $= 24.7(4.37f_c+1)$ where f_c is the auditory filter center frequency in kHz [21]

The block called modulation features extracts two features from the envelope signal in each channel k. It is a time length of the raising part of the envelope, $T_{env}(k)$, and a difference between the minimal and maximal value of the raising part of the envelope, $E_{env}(k)$ (see the bottom panel of Fig. 3). Root mean square (RMS) values are calculated in each channel and only the channels with RMS values higher than half of the maximal RMS value across all channels are fed to the subsequent block called calculation unit estimating roughness. Narrow band stimuli, e.g. SAM tones, excites also the model channels of CF higher than the frequency of the highest spectral component of the stimuli. A modulation depth of the envelope fluctuations in these channels is very high even for the envelope fluctuations of higher frequencies. Removing these channels from the roughness estimation helped to predict the bandpass dependence of roughness on the modulation frequency for SAM tones of frequencies bellow 1 kHz.

The calculation unit estimates roughness from the maximal values of the detected modulation features multiplied by a 10 ms Bartlett window as is given by

$$R(t) = \sum_{k=1}^{156} max_\tau [w(n) F_{sat}(\tau,k) E_{sat}^{1.5}(\tau,k)] \qquad (1)$$
$$n \in [0,T], \tau \in [t, t+T]$$

where $F_{sat}(k) = f(2/T_{env}(k))$, the function $f()$ transforms the time length of the raising part of the envelope (see Fig. 4), E_{sat} is a value of $E_{env}(k)$ limited not to exceed a value of 220. The limitation of E_{env}decreased roughness at low modulation frequencies emphasized by the low-pass filter in the second block. The function $f()$ assured that the roughness did not increased for the modulation frequencies above 70 Hz. Parameters of the central stage were set experimentally to predict the bandpass dependence of roughness on the modulation frequency for SAM tones [5]. The maximal value of $R(t)$ across time is taken as the estimation of roughness.

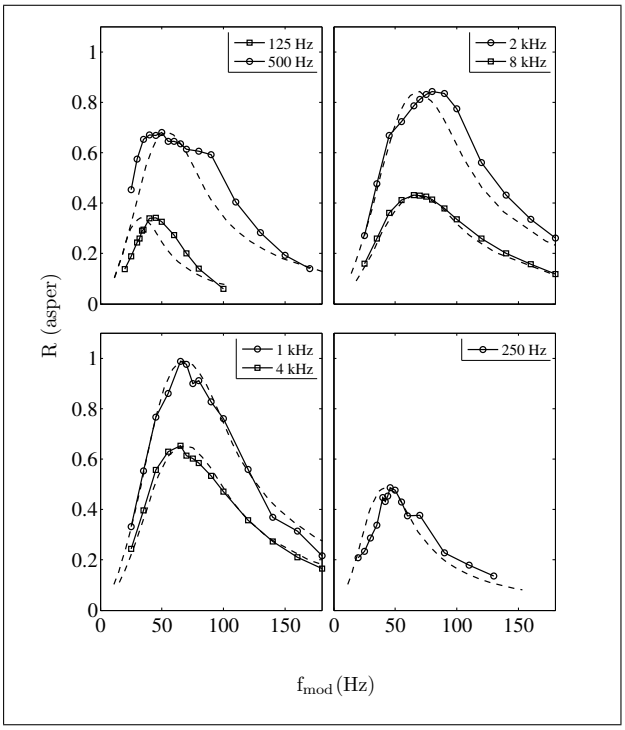

Figure 5. The dependence of roughness of SAM tones on the modulation frequency. The subjective data [5] are plotted as dashed lines. The model predicted data as markers connected by solid lines. The data were obtained with tones of a level of 60 dB SPL, a tone frequency given in the upper right corner of each panel and a modulation index of 1.

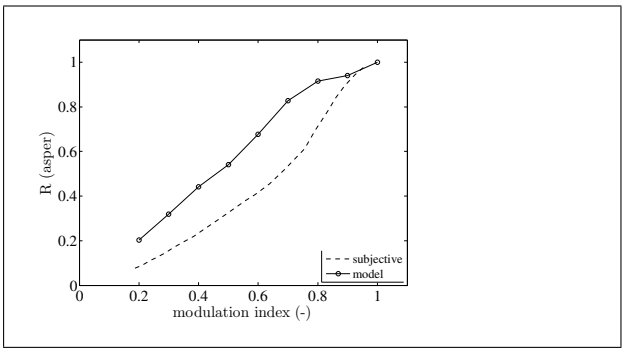

Figure 6. Roughness of a SAM tone of a frequency of 1 kHz and level of 60 dB SPL plotted as a function of the modulation index. The dashed line represents the subjective data taken from [5]. The circles connected by solid lines represents the predicted roughness data.

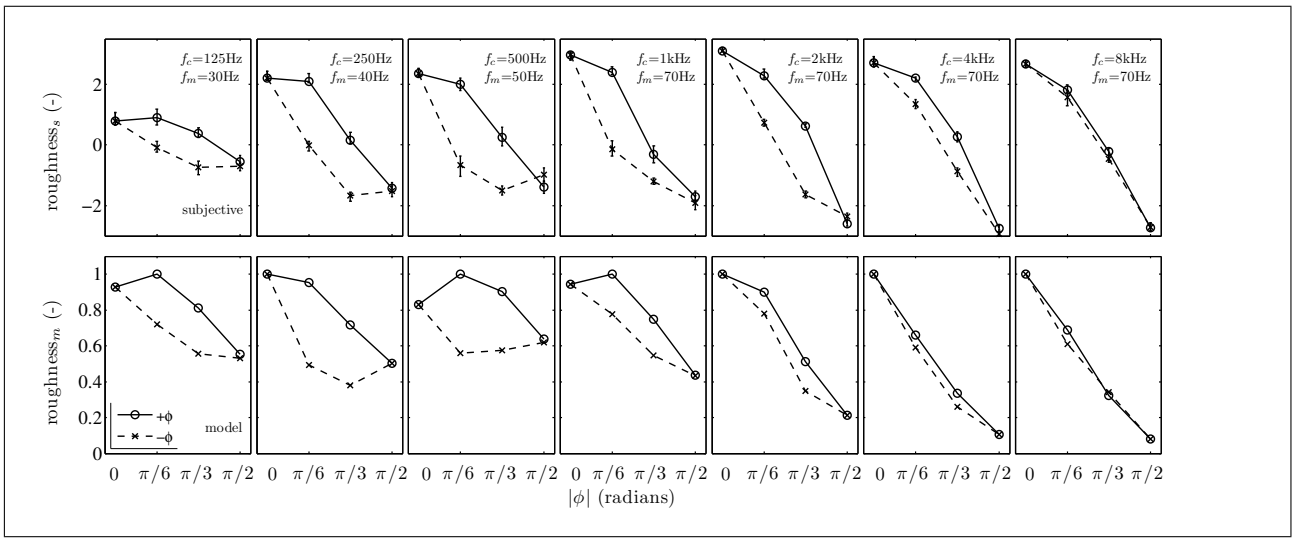

Figure 7. Roughness of pAM tones. The top panels represent the subjective data taken from [4]. The bottom panels represent the data predicted by the roughness model.

3. Model predictions of roughness

3.1. Roughness of sinusoidally amplitude modulated tones

3.1.1. Dependence on the modulation frequency

Subjective data of roughness of SAM tones were taken from the Fastl and Zwicker's book [5] (the data were measured by Aures [22]). The roughness of the SAM tones in apers is plotted as a function of the modulation frequency (dashed lines in Fig. 5). Roughness of 1 asper is defined as the roughness of a 100% SAM tone of a frequency of 1 kHz, a level of 60 dB SPL and a modulation frequency of 70 Hz. The markers connected by solid lines represent the predicted roughness data. The data for each tone frequency were normalized by its maximal value and multiplied by the maximal value of the subjective roughness for the corresponding tone frequency. The scaling was done for each tone frequency separately because the predicted roughness data did not fit quantitatively with the subjective data.

3.1.2. Dependence on the modulation index

The dependence of roughness of a SAM tone on the modulation index is depicted in Fig. 6. The dashed line shows the subjectively measured roughness of a 1 kHz, 60 dB SPL SAM tone modulated with a modulation frequency of 70 Hz. The data in aspers were taken from the book of Fastl and Zwicker [5]. The markers conected by solid lines represent the data predicted by the roughness model. The data were normalized by its maximal value and multiplied by the maximal subjective value.

3.2. Roughness of pseudo amplitude modulated tones

When the starting phase of the carrier of SAM tones is manipulated and the starting phases of the side band components are set to 0, pseudo amplitude modulated (pAM) tones are created. Pressnitzer and McAdams [4] observed the dependence of roughness of pAM tones on the starting phase of the carrier (top panels of Fig. 7, the circles are for the positive values of ϕ, crosses are for the negative values). They used the Bradly-Terry-Luce (BTL) method to construct a psychophysical scale from the paired comparison judgments. The method does not allow to quantitatively compare the subjective and predicted data. The model predictions are plotted in the bottom panels. The predicted data were normalized by the maximal roughness for each carrier frequency. A frequency of the stimuli is indicated in the upper left corner of each panel. A level of the stimuli was 60 dB SPL.

3.3. Roughness of Sawtooth and Reversed stimuli

Roughness of AM tones with the "sawtooth" envelope was measured by Pressnitzer and McAdams [4]. The psychophysical scale was again constructed by means of the BTL method. Stimuli with an abrupt increase and slow decay of the envelope were rougher (the circles in the top panels of Fig. 8) than the same temporally reversed stimuli (slow increase and abrupt decay of the envelope, the crosses in the top panels of Fig. 8). The bottom panels show roughness predicted by the roughness model. The data were normalized by the maximal roughness for each carrier frequency. A frequency of the stimuli is indicated in the upper right corner of each panel. The level was 60 dB SPL.

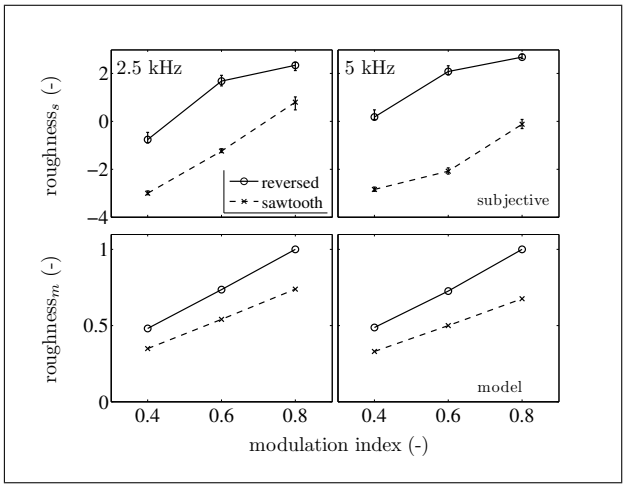

Figure 8. Roughness of AM tones with sawtooth and reversed envelope. The top panels represents the subjective data taken from [4]. The bottom panels represents the data predicted by the roughness model.

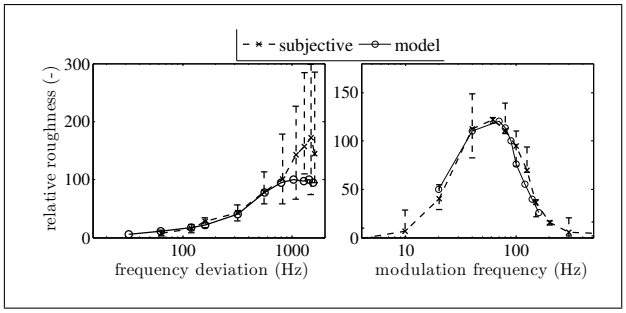

Figure 9. Roughness of FM tones. The crosses connected by dashed lines represent the subjective roughness taken from [7]. The circles connected by solid lines are the roughness model predictions. The left panel shows the roughness dependence on the frequency deviation (modulation index). The right panel shows the dependence on the modulation frequency.

3.4. Roughness of frequency modulated tones

Kemp [7] used the method of magnitude estimation to measure roughness of FM tones. Fig. 9 shows medians and quartiles of the Kemp's data (the crosses connected by dashed lines). The left panel shows the dependence of roughness of a 1.6 kHz, 60 dB SPL SFM tone modulated with a modulation frequency of 70 Hz on the frequency deviation (modulation index). The right panel shows the dependence of roughness of the same SFM tone on the modulation frequency. The tone was modulated with the frequency deviation of 800 Hz. The data predicted by the roughness model are plotted as the circles connected by solid lines. Each predicted data set was normalized and multiplied by a value of 100 to give roughness equal to 100 for the frequency of 800 Hz (data in the left panel) and the modulation frequency of 70 Hz (data in the right panel).

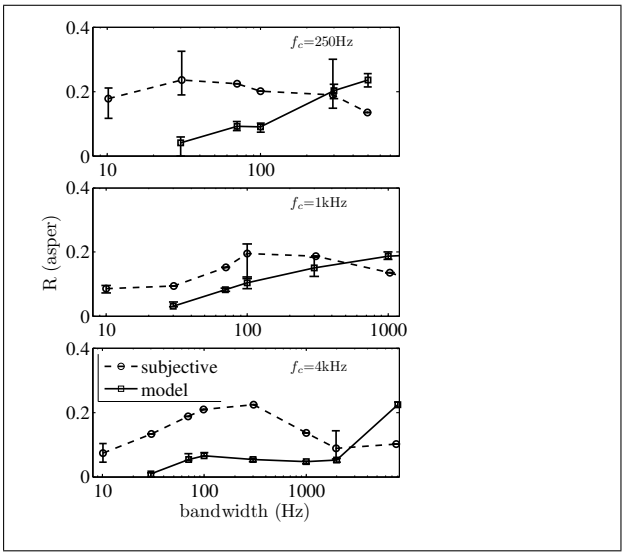

Figure 10. Roughness of unmodulated bandpass noise. The circles connected by dashed lines are medians and quartiles of subjective data taken from [8]. The squares connected by solid lines are medians and quartiles from ten model predictions of roughness.

3.5. Roughness of unmodulated narrowband noise

Subjectively measured roughness data of unmodulated noise were taken from Aures [8]. The circles connected by dashed lines represents medians and quartiles of the subjective data. The squares connected by solid lines represent medians and quartiles calculated from ten repeated roughness estimations conducted with the roughness model. The predicted data were normalized by its maximal values of roughness for each data set and multiplied by the corresponding maximal value of the subjective roughness.

The model failed to qualitatively predict roughness of the noise stimuli. Moreover the predicted roughness for the noise centered at 1 kHz expressed in aspers was higher than the subjective data (approx. 0.7 aspers for the bandwidth of 70 Hz). This can be partly connected with saturation observed in the roughness predictions for the roughness dependence on the modulation index of a SAM tone (see Fig. 6. It can also mean that the features used to predict roughness do not give all necessary information to calculate roughness. Daniel and Weber's model predicted roughness of the bandpass noise stimuli. The Leman's model predicted increasing roughness with increasing bandwidth of the stimuli as well as the presented roughness model.

4. Effect of cochlear model

The roughness model was designed with the Nobili's cochlear model. Two other types of models which are more effective in terms of its computational demands were also tested. Both were filterbank models, one

was composed of a bank of gammatone filters [23], the other was composed of a bank of dual resonance nonlinear filters [24]. Roughness model with all of the tested models predicted the bandpass characteristics of the roughness of SAM and SFM tones, but fit with the subjective data was best for the Nobili's cochlear model. The model with a gammatone filterbank predicted smaller roughness difference for the positive and negative starting phases of pAM stimuli than were the subjective results. On the other hand, the roughness model with the Nobili's cochlear model was in some cases oversensitive to phase differences (e.g. pAM tone $f_c = 500\,\mathrm{Hz}$ in Fig. 7).

5. CONCLUSIONS

The presented roughness model calculates roughness from the envelope of the simulated neural signal in auditory nerve fibers. Beside the amplitude changes of the fluctuating envelope, the model takes into account also a duration of the raising part of the envelope. This allowed to predict the effect of phase of the individual spectral components on roughness. The model quantitatively predicted roughness for a broad range of stimuli.

The peripheral stage of the roughness model is composed of a Nobili's model of the BM response and cochlear hydrodynamics. Alternative cochlear models were also tested (gammatone and DRNL filterbanks) but the best fit with the subjective roughness data was reached with the Nobili's cochlear model.

The roughness model failed to predict roughness of not modulated bandpass noise stimuli. Further studies are necessary to explain these results.

Acknowledgement

This paper was written at the Academy of Performing Arts in Prague as part of the project "Sound quality" with the support of the Institutional Endowment for the Long Term Conceptual Development of Research Institutes, as provided by the Ministry of Education of the Czech Republic in the year 2014.

References

[1] H. L. F. von Helmholtz: On the Sensations of Tone as the Physiological Basis for the Theory of Music. Translated by A. J. Ellis (1885), 4th ed., Dover, New York, 1954.

[2] R. C. Mathes, R. L. Miller: Phase Effects in Monaural Perception. J. Acoust. Soc. Am. **19** (1947) 780-797.

[3] E. Terhardt: On the perception of periodic sound fluctuations (Roughness). Acustica **30**,(1974) 201-213.

[4] D. Pressnitzer, S. McAdams: Two phase effects in roughness perception. J. Acoust. Soc. Am. **105** (1999) 2773-2782.

[5] H. Fastl, E. Zwicker: Psychoacoustics: Facts and Models. Springer, Berlin, Heidelberg, 2007.

[6] P. N. Vassilakis: Auditory roughness as a means of musical expression. Selected Reports in Ethnomusicology **12** (Perspectives in Systematic Musicology) (2005) 119-144.

[7] S. Kemp: Roughness of frequency modulated tones. Acustica **50** (1982) 126-133.

[8] W. Aures: Ein Berechnungsverfahren der Rauhigkeit. Acustica **58** (1985) 268-281.

[9] P. Daniel, R. Weber: Psychoacoustical roughness: Implementation of an optimized model. Acustica **83** (1997) 113-123.

[10] M. Leman: 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 Audio Effects (DAFX-00) (2000) DAFX 1-DAFX 6.

[11] A. Kohlrausch, D. Hermes, R. Duisters: Modeling roughness perception for sounds with ramped and damped temporal envelopes. Forum Acusticum (2005) 1719-1724.

[12] Essex Hearing Research Lab. Matlbab Auditory Periphery (MAP) `http://www.essex.ac.uk/psychology/department/hearinglab/modelling.html`

[13] A. Huber, T. Linder, M. Ferrazzini, S. Schmid, N. Dillier, S. Stoeckli, U. Fisch: Intraoperative assessment of stapes movement. Ann Otol Rhinol Laryngol **110** (2001) 31-35.

[14] F. Mammano, R. Nobili: Biophysics of the cochlea: Linear approximation. J. Acoust. Soc. Am. **93** (1993) 3320–3332.

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

[16] Cochlea Modeling `http://www.pd.infn.it/~rnobili/cochmodels/cochmodels.html`

[17] S. A. Shamma, R. S. Chadwick, W. J. Wilbur, K. A. Morrish, J. Rinzel: A biophysical model of cochlear processing: Intensity dependence of pure tone responses. J. Acoust. Soc. Am. **80** (1986) 133-145.

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

[19] R. Meddis: Auditory-nerve first-spike latency and auditory absolute threshold: A computer model. J. Acoust. Soc. Am. **119** (2006) 406-417.)

[20] R. Meddis: Simulation of Mechanical to Neural Transduction in the Auditory Receptor. J. Acoust. Soc. Am. **79** (1986) 702-711.

[21] B. C. J. Moore, B. R. Glasberg: A revision of Zwicker's loudness model. Acta Acustica united with Acustica **82** (1996) 335-345.

[22] W. Aures: Berechnungsverfahren für den Wohlklang beliebiger Schallsignale, ein Beitrag zur gehörbezogenen Schallanalyse. PhD thesis, TU München, 1984.

[23] M. Slaney: An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank. Apple Computer Technical Report 35, 1993.

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