Modeling roughness perception for complex stimuli using a model of cochlear hydrodynamics

V. Vencovsky
MARC, AMU in Prague, Malostranske nam. 12, 118 00 Prague, Czech Republic
vencovac@fel.cvut.cz

A roughness model composed of a physiological auditory model and an algorithm calculating roughness from the envelope of the auditory model output signal is described in the study. The roughness model is sensitive to phase changes between the spectral components and to shape of the temporal waveform of the analyzed stimuli which limits most of the state of the art roughness models. Synthetic and real complex stimuli were used in this study to test the model performance. Amplitude modulated harmonic complexes and intervals in the chromatic scale composed of harmonic complexes were among the synthetic stimuli. Voice samples of a vowel /a/ extracted from the signal recorded during the scale signing were used as the real stimuli. Some of the samples were dysphonic (with roughness). Listening tests were conducted to obtain roughness ratings of the stimuli. The subjective roughness ratings correlated with the ratings predicted by the presented roughness model.

1 Introduction

Two pure tones close in frequency which are added together create a signal with fluctuating envelope. The frequency of the fluctuations is equal to the frequency difference between the tones. The human ear perceives the envelope fluctuations as periodic changes of loudness when the frequency of the fluctuations is bellow approx. 30 Hz. Fluctuations of higher frequencies cause a jarring and rough sound sensation. A reason for the sensation of roughness seems to be in an inability of the ear to resolve spectral components of the sound stimuli which then interfere together and cause fluctuations of the neural signal in auditory nerve fibers [1, 2].

Roughness of sound stimuli can be quantified by listening tests [3, 4, 5, 6, 7]. A variety of models quantifying roughness have been proposed hand in hand with the listening tests [2, 6, 8, 9, 10]. One group of models, so called *curve-mapping* models, detects spectral components of the sound stimuli and map it into a psychoacoustical curve of roughness. The biggest drawback of the models is that they cannot process signals with continuous spectra, e.g. noises [10]. The second group of roughness models takes into account a critical band filtering in the cochlea [8, 9, 10]. Daniel and Weber [9] improved the model designed by Aures [8]. The model calculates the perceived roughness from a modulation depth of the signal in each critical band. The model can predict roughness of sinusoidally amplitude modulated and sinusoidally frequency modulated tones and not modulated narrow-band noises [9].

This study presents a roughness model whose peripheral stage contains a physiological auditory model. A central stage of the roughness model calculates roughness from the envelope of the auditory model output signal. It detects the raising parts of the envelope and calculates roughness from the modulation depth and duration of the raising parts. This processing allows to predict the effect of phase of the individual spectral components and the effect of shape of the time waveform on roughness which was shown in [5]. The roughness model is used in this study to predict roughness of synthetic and real complex stimuli. The predictions are compared with results of listening tests.

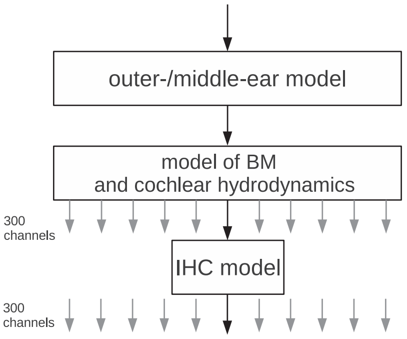

Figure 1: Block diagram of the auditory model

2 Roughness model

The roughness model is composed of two stages. The first stage is a computational model of the human auditory system (auditory model). The second stage, a central stage, processes the output signal of the auditory model and calculates the predicted roughness ratings.

2.1 Auditory Model

The auditory model is composed of an outer- and middle-ear model, a model of the basilar membrane (BM) response and cochlear hydrodynamics and a model of inner hair cells (IHCs) (see Fig. 1).

Outer- and middle-ear model: The outer- and middle-ear model transforms the acoustic signal entering the ear into the velocity of the oval window vibrations at the input of the cochlea. A transfer function of the outer- and middle-ear was measured in the human ears by a number of authors (e.g. [11, 12]). An author of this study designed its own transfer function which does not simulate the real transfer function of the outer- and middle-ear but works with the cochlear model and inner hair cells model described in the next paragraphs. The transfer function approximated by a 256-point FIR filter is plotted in Fig. 2.

Model of the BM and cochlear hydrodynamics: The second block of the auditory model, a model of the BM was designed by Mammano and Nobili [13, 14, 15]. The BM is modeled as an array of oscillators with mass, damping and stiffness. An active function of outer hair cells is incorporated into the model as a force causing undamping

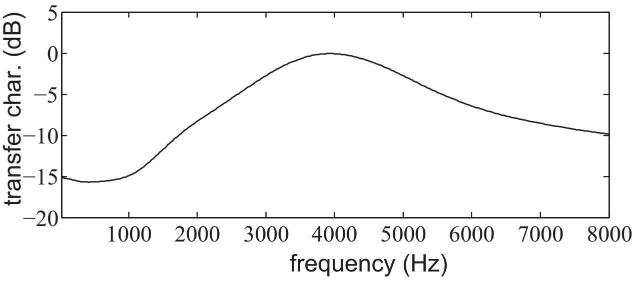

Figure 2: Amplitude transfer function of the experimentally designed outer- and middle-ear filter.

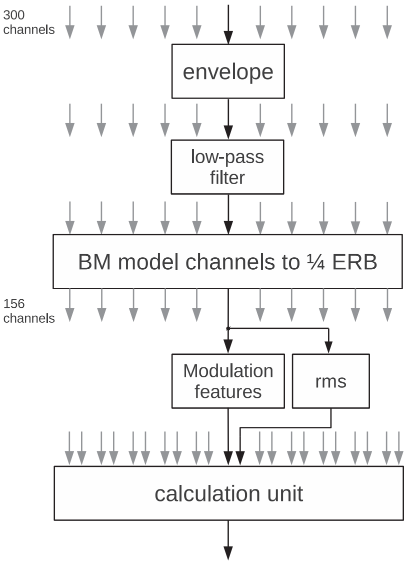

Figure 3: Block diagram of the central stage of the roughness model

of the array of oscillators. An author of the paper changed damping parameters of the model published in [15]. He decreased damping in order to increase the frequency selectivity of the model. Output of the cochlear model represents a displacement of the BM in discrete points along the BM (300 points distributed between the characteristic frequency (CF) of 27 and 16875 Hz). A time-domain Matlab (Mathworks) implementation of the model can be downloaded from the internet [16].

Model of inner hair cells: The IHC model is composed of a set of algorithms simulating the IHC physiology. The BM displacement is transformed to the displacement of the IHC cilia. This processes is modeled by the algorithm designed by Shamma *et al.* [17]. Bending of the cilia opens ion channels which entering the IHC cause its depolarization. This process is modeled by the algorithm designed by Sumner *et al.* [18]. The IHC depolarization in turn opens calcium ion channels which results in changes of concentration of calcium in the IHC. It is modeled by Meddis' algorithms [19]. The calcium concentration controls the release of neurotransmitter into the synaptic cleft which leads to increased neural activity in the auditory nerve fibers. The release and circulation of neurotransmitter in the synapse is modeled by the Meddis' probabilistic model [20]. The algorithms implemented in Matlab (Mathworks) can be downloaded from the internet [21].

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. 3.

The first block calculates envelope of the signal in each of the 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. 4 (upper part) shows the time course of the auditory model output signal and the calculated envelope in the channel of CF = 1 kHz in response to a 100% 1-kHz AM tone sinusoidally modulated with modulation frequency of 70 Hz. The envelope signal is then low-pass filtered by a 1st-order butterworth filter with cutoff frequency of 80 Hz. This filter assures decrease of roughness for modulation frequencies above approx. 70 Hz as was observed with various types of stimuli [6]. The filtered envelope of the envelope signal depicted in the upper part of Fig. 4 is plotted in the bottom part of the same Fig. The cochlear model simulates the BM response in 300 discrete points distributed in a frequency range between 27 and 16 875 Hz. Signals in channels of CF in a range of 1/4 of ERB are averaged into one channel. The ERB values represents the psychophysically measured auditory filter bandwidth given by the relationship ERB = 24.7(4.37f_c + 1) where f_c is an auditory filter center frequency in kHz [22]. This processing decrease the number of channels from 300 to 156.

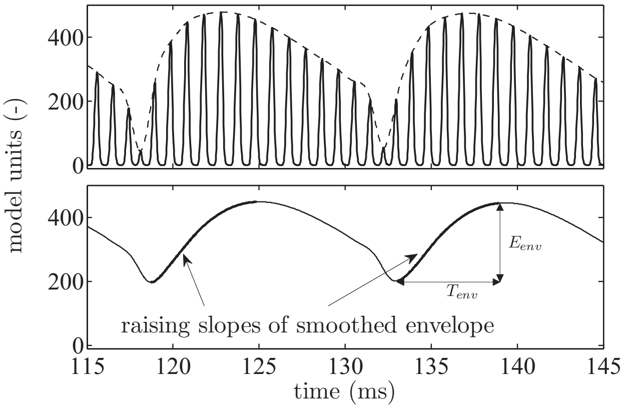

Figure 4: Upper plot: The solid line is a time course of the auditory model output signal in a channel of CF = 1 kHz in response to a 100%, 1-kHz sinusoidally amplitude modulated tone modulated with a modulation frequency of 70 Hz. The dashed line is an envelope calculated as a cubic spline interpolation of peaks in the time course of the auditory model output signal. Bottom plot: The envelope signal from the upper plot smoothed by a 1-st order Butterworth filter. The tick lines are raised slopes of the envelope used to extract the modulation features for roughness calculation, T_{env} and E_{env}.

The block called modulation features extracts two features from the envelope signal in each channel k. It is a time length of the raising slope of the envelope, $T_{env}(k)$, and a difference between the minimal and maximal value of the raising slope of the envelope, $E_{env}(k)$ (see the bottom part of Fig. 4). Calculation unit predicts roughness from the maximal values of the detected modulation features and the root mean square values of the envelope signal $RMS(k)$ in 40 ms long time windows by the algorithm

$$R(t) = \sum_{k=1}^{156} \frac{RMS(k)F_{sat}(k)E_{sat}^{1.5}(k)}{\sum_{k=1}^{156} RMS(k)} \qquad (1)$$

where $F_{sat}(k) = f(2/T_{env}(k))$, the function $f()$ transforms the time length of the raising slope of the envelope (see Fig. 5), E_{sat} is a saturated depth of the amplitude modulation of the signal envelope calculated from $E_{env}(k)$ divided by $RMS(k)$. It is hardly limited to values between 0 and 0.4. The predicted roughness is a maximal value of $R(t)$ across the time windows.

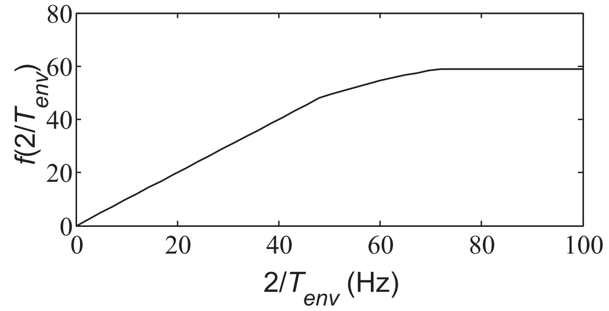

Figure 5: Transformation function $f()$ applied to the modulation feature T_{env}.

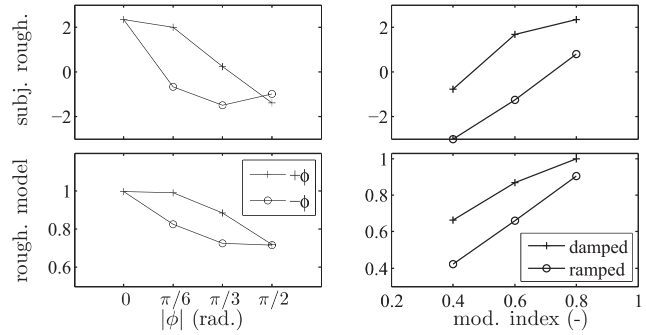

Figure 6: Subjective and predicted roughness of pseudo AM stimuli (left column) and damped/ramped stimuli (right column).

The roughness algorithm allows to predict the effect of phase of the spectral components and temporal assymetry of signal waveform on roughness which was psychophysically measured by Pressnitzer and McAdams [5]. Their subjective data are shown in the upper row of Fig. 6. The bottom row shows the model predictions. The left column is for pseudo AM tones of 500 Hz, the right column for damped and ramped tones of 2.5 kHz. More results with these stimuli will be published in a different study.

3 Experiments

3.1 Experiment 1: Roughness of amplitude-modulated harmonic complexes

Synthetic amplitude modulated harmonic complexes were used in the Experiment 1. Roughness of the stimuli was measured by means of the rating listening test with a 7-point rating scale and by means of two roughness models.

Stimuli: The stimuli were harmonic complexes composed of the first three harmonics at frequencies of 300, 600 and 900 Hz, respectively. The spectral components were sinusoidally amplitude modulated in order to control its roughness. The modulation frequencies were 30, 40, 50, 60 and 70 Hz and the modulation depth was 0, -3, -6, -9 and -12 dB calculated as $20\log_{10} m$, where m is the modulation index ranging from 0 to 1. Amplitude of the first, second and third spectral component was 0, -10 and -20 dB, respectively. Duration of the stimuli was 600 ms and they were ramped on and off with 30 ms raised-cosine ramps. A level of the stimuli was 75 dB SPL. Combination of the modulation frequencies and the modulation depths led to 25 different stimuli.

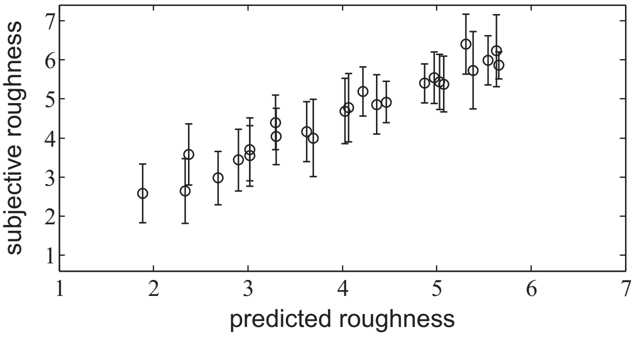

Figure 7: Subjective roughness ratings of the amplitude modulated harmonic complexes plotted as a function of the roughness model ratings.

Listeners: Six listeners (one woman, five men, age ranged from 25 to 44 years) participated in the experiment. The listeners had normal hearing (pure-tone thresholds below 20 dB HL for frequencies between 250 Hz and 8 kHz). One of the listeners was an author of the study.

Procedure and equipment: The procedure was inspired by Patel *et al* [7]. Roughness of the stimuli was rated on a discrete scale from 1 to 7 in steps of 1, where 1 was for the lowest and 7 for the highest roughness. The listeners rated roughness of 25 different stimuli presented in random order. Each stimulus was rated separately. The listeners could listen to it as many times as they desired and after assigning the roughness rating, they could listen to the next stimulus. The listening test was composed of 10 sets of randomly ordered 25 stimuli. It means that each stimulus was rated 10 times giving overall number of 250 stimuli in the test. The test was conducted on a computer. The stimuli were presented to the listeners via Sennheisser HD-600 headphones (same signal in both ears).

Results: Mean values and standard deviations of the roughness ratings across all listeners measured by the listening test are plotted in Fig. 7 as a function of the roughness model predictions. The predicted roughness data were scaled to the 7-point scale. Pearsons's correlation between the subjective and predicted roughness is $r = 0.97$ with $p = 5.4 \times 10^{-16}$, Spearman's correlation is $r = 0.97$ with $p = 7.8 \times 10^{-7}$. The roughness model ratings correlates with the subjective ratings.

The same stimuli were processed by means of the roughness model designed by Daniel and Weber [9]. The model implementation in the sound analysis software PsySound3 [23] was used. The Daniel and Weber's model gave results which correlated with the subjective ratings (Pearson's correlation: $r = 0.90$, $p = 1.4 \times 10^{-9}$, Spearman's correlation: $r = 0.98$, $p = 3.5 \times 10^{-18}$).

3.2 Experiment 2: Roughness of real voice samples

Experiment 2 used real voice samples containing different amount of roughness. The roughness of the stimuli

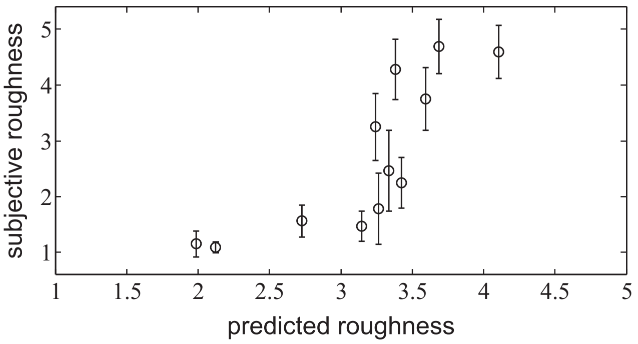

Figure 8: Subjective roughness ratings of the real voice samples plotted as a function of the roughness model ratings.

was again measured by means of a listening test and by means of the roughness models.

Stimuli: The stimuli were real voice samples of a vowel /a/. The samples (12 stimuli) were extracted from the recordings of the scale signing. The voices were recorded on eleven different subjects. The subjects had a pathology on larynx which caused some of the recorded samples to be dysphonic (with roughness). Pitch of the stimuli was not the same. Duration of the stimuli was 300 ms and they were ramped on and off by 30 ms raised-cosine ramps. A level of the stimuli was 75 dB SPL.

Listeners: Six listeners (men of age ranging between 25 and 36 years) 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). One of the listeners was an author of the paper.

Procedure and equipment: Roughness was rated on a discrete 5-point scale from 1 to 5 in steps of 1 (1 for the lowest and 5 for the highest roughness). The procedure and equipment was the same as in Experiment 1. Randomly ordered 12 stimuli were rated 10 times giving 120 stimuli.

Results: The listening test results (mean values and standard deviations) are plotted in Fig. 8 as a function of the roughness model predictions. The predicted results were scaled to span the subjective scale. Pearson's correlation between the subjective and predicted roughness is $r = 0.81$ with $p = 1.5 \times 10^{-3}$, Spearman's correlation is $r = 0.9$ with 5.9×10^{-6}. The roughness model successfully rated the stimuli with high and low roughness but its performance was worse in the middle of the subjective roughness scale.

The Daniel and Weber's model failed to predict roughness for the real voice samples (Pearson's correlation: $r = -0.26$, $p = 0.41$, Spearman's correlation: $r = -0.21$, $p = 0.51$).

3.3 Experiment 3: Roughness of intervals in the chromatic scale

Roughness of thirteen intervals in the chromatic scale was investigated in Experiment 3. The subjective roughness ratings taken from Vassilakis [2] were compared with the predictions of the roughness models.

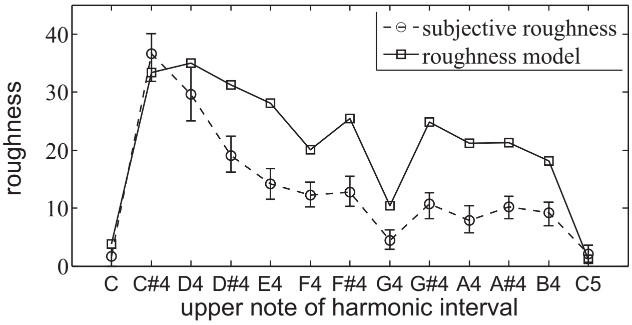

Figure 9: Roughness ratings of the harmonic intervals constructed from synthetic complex tones. Circles represents mean values of the subjective data across ten listeners taken from [2]. Squares are the roughness model ratings.

Stimuli: Synthetic complex tones were used to construct harmonic intervals. The complexes were composed of six harmonics with amplitudes $A_n = A_1/n$, where n is the number of the harmonics and A_n is amplitude of the nth harmonic. The intervals started on middle C(C4, fundamental frequency 256 Hz, equal temperament) [2].

Listeners and procedure: Roughness was rated by 10 listeners on a continuous scale between 0 (not rough) and 42 (rough). The stimuli were presented to them by earphones (same signal in both ears). The listeners were asked to set a position of a scroll to the perceived amount of roughness.

Results: The roughness model results were scaled to fit the scale used in the listening test. The model data are depicted in Fig. 9 as squares connected with the solid line. The subjective data taken from Vassilakis [2] are plotted as circles connected with the dashed line. Fit between the roughness data is not quantitative. Pearson correlation coefficient is $r = 0.85$ with $p = 2.7 \times 10^{-4}$. The model successfully predicts the lowest roughness for the intervals of octave and also reflects the dip for the interval G4. Spearman's correlation coefficient is $r = 0.95$ with $p = 0$.

The Daniel and Weber's model ratings were less correlated with the subjective data than in the case of the presented roughness model (Pearson's correlation:

Table 1: Pearson's and Spearman's correlation coefficients between the subjective and predicted roughness ratings. R.M. stands for the presented roughness model, D.W. stands for the Daniel and Weber's roughness model.

			Exp. 1	Exp. 2	Exp. 3
R.M.	Pearson	r	0.97	0.81	0.85
		p	5.4×10^{-16}	1.5×10^{-3}	2.7×10^{-4}
	Spearm.	r	0.97	0.90	0.95
		p	7.8×10^{-7}	5.9×10^{-6}	0
D.W.	Pearson	r	0.90	-0.26	0.71
		p	1.4×10^{-9}	1.4×10^{-3}	2.7×10^{-4}
	Spearm.	r	0.98	-0.21	0.41
		p	7.8×10^{-7}	0.41	6.5×10^{-3}

$r = 0.71$, 2.7×10^{-4}, Spearman's correlation: $r = 0.41$, $p = 6.5 \times 10^{-3}$.).

4 Conclusion

The presented roughness model uses a physiological auditory model together with an algorithm calculating roughness from the envelope of the auditory model output signal. The roughness model was used in this study to predict roughness of synthetic and real stimuli. The predicted ratings were compared with roughness ratings obtained by means of listening tests. Both, Pearsons's and Spearman's correlations between the predicted and subjective ratings were high for all types of the tested stimuli.

For a comparison with the roughness model which also employs the human physiology (simulates the critical band filtering in the human cochlea), the subjective roughness ratings were compared with the results of the Daniel and Weber's roughness model. The model ratings correlated with the subjective ratings just for one type of the synthetic stimuli. There was no correlation for the real voice samples.

Acknowledgments

I would like to thank to Aleš Vetešník for sending the Matlab scripts of the cochlear model. The research was supported 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.

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

[3] R. C. Mathes, R. L. Miller, ”Phase Effects in Monaural Perception”, *J. Acoust. Soc. Am.* **19**(5), 780-797 (1947)

[4] E. Terhardt, ”On the perception of periodic sound fluctuations (Roughness)”, *Acustica* **30**(4), 201-213 (1974)

[5] D. Pressnitzer, S. McAdams, ”Two phase effects in roughness perception”, *J. Acoust. Soc. Am.* **105**(5), 2773-2782 (1999)

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

[7] S. A. Patel, R. Shrivastav, D. A. Eddins, ”Identifying a Comparison for Matching Rough Voice Quality”, *J. Speech Lang. Hear. Res.* **55**, 1407-1422 (2012)

[8] W. Aures, ”Ein Berechnungsverfahren der Rauhigkeit”, *Acustica*, **58**(5), 268-281 (1985)

[9] P. Daniel, R. Weber, ”Psychoacoustical roughness: Implementation of an optimized model”, *Acustica* **83**(1), 113-123 (1997)

[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)*, Verona, Italy, DAFX 1-DAFX 6 (2000)

[11] D. Hammershøi, H. Møller, ”Sound transmission to and within the human ear canal”, *J. Acoust. Soc. Am.* **100**(1), 408-427 (1996)

[12] W. Chien, J. J. Rosowski, M. ,E. Ravicz, S. D. Rauch, J. Smullen, S. .M. Merchant, ”Measurement of stapes velocity in live human ears”, *Hear. Res.* **249**, 54-61 (2009)

[13] F. Mammano, R. Nobili, ”Biophysics of the cochlea: Linear approximation”, *J. Acoust. Soc. Am.* **93**(6), 3320–3332 (1993)

[14] R. Nobili, F. Mammano, ”Biophysics of the cochlea II: Stationary nonlinear phenomenology”, *J. Acoust. Soc. Am.* **99**(4), 2244-2255 (1996)

[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**, 478-494 (2003)

[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**(1), 133-145 (1986)

[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**(5), 2178-2188 (2002)

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

[20] R. Meddis, ”Simulation of Mechanical to Neural Transduction in the Auditory Receptor”, *J. Acoust. Soc. Am.* **79**(3), 702-711, (1986)

[21] Matlab Auditory Periphery (MAP) `http://www.essex.ac.uk/psychology/department/hearinglab/modelling.html`

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

[23] PsySound3 `http://psysound.wikidot.com/`