Basilar membrane excitation patterns of rough sounds simulated by means of an auditory model

Václav Vencovský, Marek Frič

Musical Acoustics Research Centre, Musical and Dance Faculty, Academy of Performing Arts in Prague, Malostranské nám. 13, 118 00, Praha 1, Czech Republic,
e-mail: vencovac@fel.cvut.cz

Annotation

Roughness is a term used to describe a specific sound sensation usually accompanying cases when the sound signal spectrum contains more than one harmonic component within one critical band. It is the case of amplitude or frequency modulated signals but also the case of aperiodic voiced speech signals during dysphonia. It is believed that interaction of the spectral components on the basilar membrane causes perception of roughness. Sound stimuli with roughness were processed by means of the hydrodynamical model of the cochlea. The model allows better prediction of masking patterns of harmonic complexes than widely used filterbank models. Amplitude modulations in the model output signal were observed when rough sounds were processed by the auditory model which is in accordance with theory [Terhardt(1974),Acustica,30,201-213]. Three parameters were taken from the model output envelope. It was the modulation frequency, depth and steepness of the raising part of the envelope. Simple algorithm calculating amount of roughness from these three parameters was developed. Predicted roughness was compared with results of the rating listening test conducted with synthetic vowels /a/. Spearman correlation between the predicted and subjective results is very high which indicates that the model can be successfully used for automatic detection of roughness.

Results

Output signal of the auditory model represents the probability of spike firing (firing of the neuronal pulse) in auditory nerves. It is synchronized with time fine structure of the signal on the basilar membrane up to approx. 4 kHz. Fig. 1, 2 and 3 shows envelope of this neural signal (values of probability taken in time of maximal excitation). Ordinate of the graphs represents characteristic frequencies of different sections on the basilar membrane, abscissa is time.

Two tones:

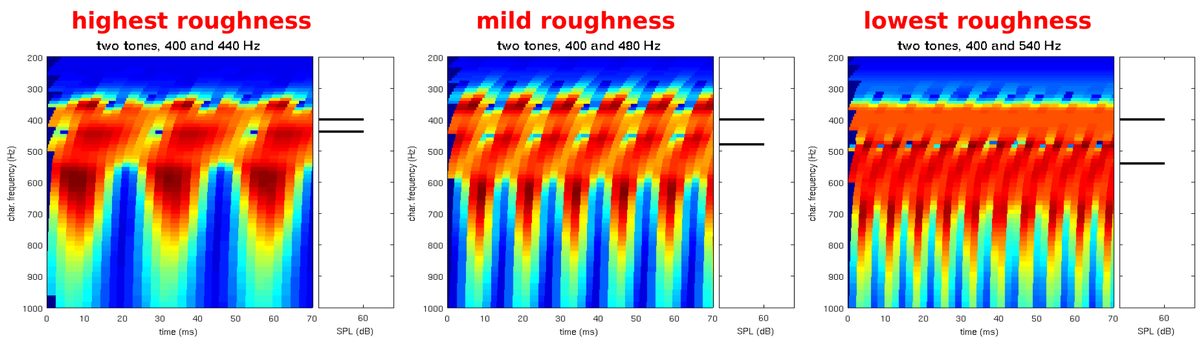

Fig.1 – Envelope of the model output signal during two tone excitation

Fig 1 shows the envelope of the simulated neural signal in response to two 60 dB SPL tones of different frequencies (amplitude spectra of the stimuli are on the right side of each output signal). The graph on the left corresponds to the signal with highest roughness. The lowest roughness is in the case of the right graph where relatively smooth areas corresponds to individual spectral components. Modulation frequency is 140 Hz which also leads to lower perception of roughness [Zwicker, Fastl Psychoacoustics: Facts and Models, Sringer].

Real sustained vowels:

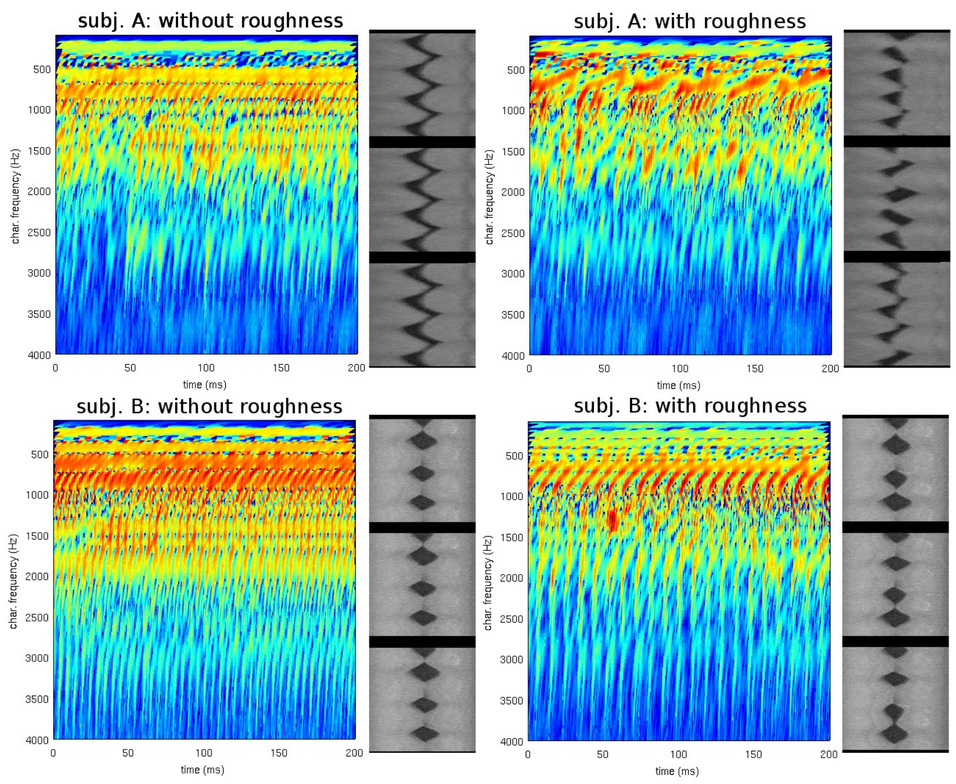

Fig.2 – Envelope of the model output signal in response to sustained vowel recorded during videokymography. Each row is for one subject. Left column for the smooth presentation, right with dysphonia

Envelopes of the model output signal in response to sustained vowels recorded during videokymography are shown in Fig. 2. The left and right column corresponds to the signal without roughness and with roughness, respectively. Relatively smooth areas in the graphs represent individual harmonic components of the sound stimuli. Envelopes of the model output signals are again much more amplitude modulated and the modulation frequencies are lower for rough stimuli (graphs in the right column) than for the smooth vowels.

Synthesized sustained vowels:

Synthesized sustained vowels /a/ were generated by Klatt synthesizer [Klatt (1980) JASA, 67, 971-995]. The synthesizer calculates a glottal signal from unit impulses. Amplitude and period of these impulses was manipulated in order to increase roughness of the vowels. Glottal Jitter and Shimmer values ($Jitt_{synth}$ and $Shim_{synth}$) in Tab.1 shows this manipulation. Fig.3 shows the model output envelopes in response to some of the synthesized speech stimuli. Different graphs corresponds to different amounts of Jitter and Shimmer or fundamental frequency. It is also visible that Jitter and Shimmer causes different patterns in the envelope of the model output signal.

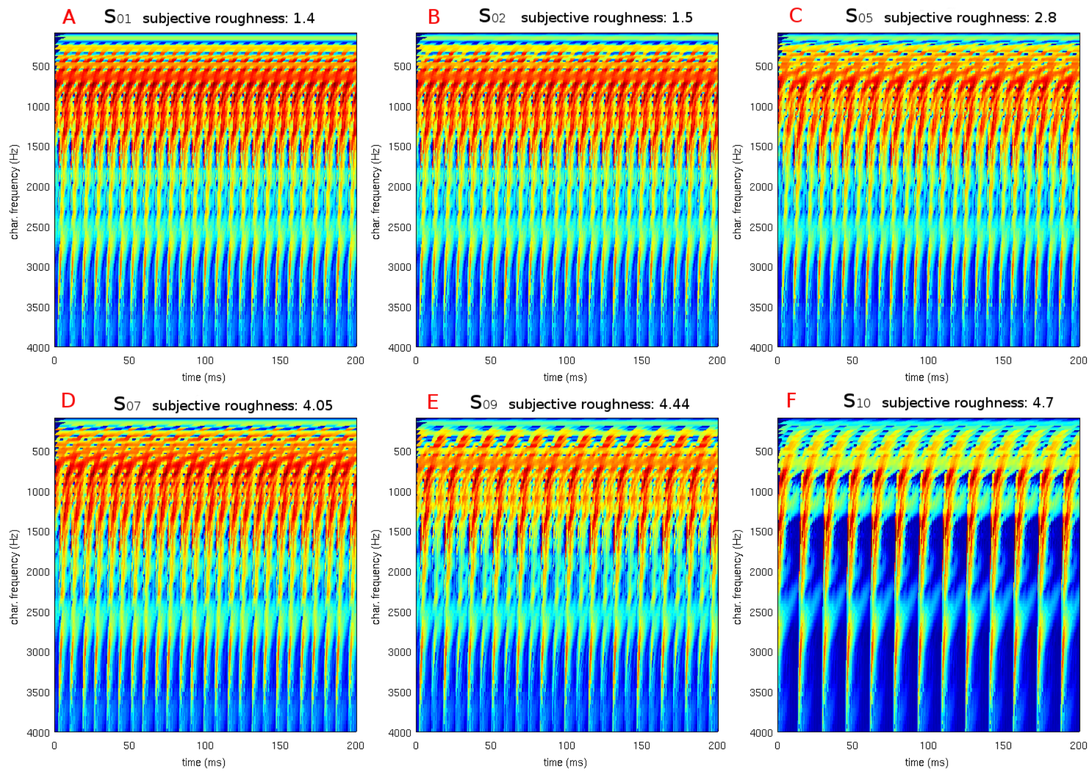

Fig.3 – Envelope of the model output signal in response to the synthesized sustained vowels /a/. Perceived roughness increases from 1 to 5.

Listening test and automatic detection of roughness:

Simple algorithm detecting amount of roughness from the envelope of the neural signal was developed. The model predictions were compared with results of the listening test conducted with 10 stimuli (synthesized wovels /a/) listed in Tab.1. Four participants rated amount of perceived roughness of each stimulus separately on a 5-point scale with 1 corresponding to minimal and 5 to maximal roughness. The stimuli were presented in random order. Intensity of the stimuli was 75 dB SPL. Each stimulus was rated 10 times. Intrasubject correlation (Cronbach alpha) was in all cases higher than 0.75. Subjective and model results are listed in Tab.1 as $R_{subj.rating}$ and $R_{model\ rating}$, respectively. Fig. 4 shows mean results and standard deviation from the mean across 4 listeners. Fig. 5 shows the model results vs. subjective results. Spearman correlation is high which indicates that the model can predict an amount of roughness of complex speech stimuli.

stimuli	s_{01}	s_{02}	s_{03}	s_{04}	s_{05}	s_{06}	s_{07}	s_{08}	s_{09}	s_{10}
$f_{0,synth}$(Hz)	125	125	125	125	125	125	125	125	125	63
$Jitt_{synth}$(%)	0	0	5	0	0	5.1	12.5	9.7	0	0
$Shim_{synth}$(%)	0	9.7	0	20	33.3	33.3	0	33.3	80	0
$R_{subj.\ rating}$	1.4	1.5	1.88	1.91	2.8	3.75	4.05	4.05	4.44	4.7
$R_{model\ rating}$	1.6	2.05	2.2	2.3	2.63	3.49	2.9	2.97	6.6	14

Tab.1 – synthetic sustained vowels and its parameters

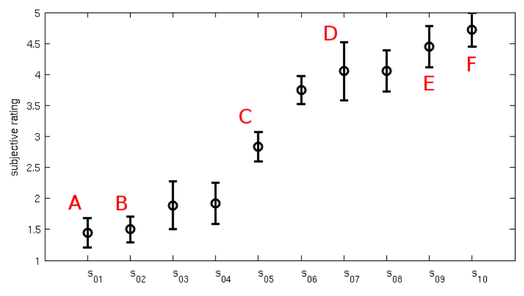

Fig. 4 – Listening test results. Amount of roughness on 5 point rating scale. 1 – lowest roughness, 5 – highest roughness

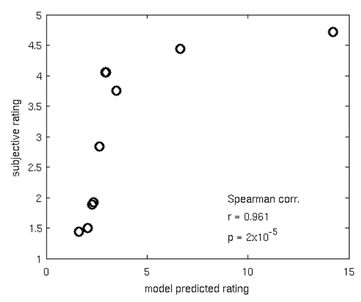

Fig. 5 – Predicted vs subjective roughness

Conclusion

A hydrodynamical auditory model was used to process stimuli with roughness. It was in accordance with the theory [Terhardt(1974),Acustica,30,201-213] observed that interaction of spectral components within one critical band causes amplitude modulations in the envelope of neural signal.

Simple algorithm to calculate amount of perceived roughness from three parameters (the modulation frequency, depth and steepness of the raising part of the envelope) was developed.

The rating listening test with synthetic sustained vowels /a/ was conducted to obtain amount of perceived roughness and its results were compared with the amount of roughness calculated by means of the auditory model and detecting algorithm.

Spearman correlation between predicted and subjective results is very high which indicates that the auditory model plus the detection algorithm can reliable calculate amount of roughness in case of complex speech stimuli.

Acknowledgements: This work was supported by *CESNET Development Fund No. 423/2011 and 424/2011.* Videokymographical recordings were recorded by doc. Šram as a part of project EUREKA E! 2614

PEVOC 2013