Deriving sound quality measures from a perceptual model

Lena Schell-Majoor[1], Jan Rennies[1], Stephan D. Ewert[2], Birger Kollmeier[2]
[1] *Fraunhofer IDMT, Project Group Hearing, Speech and Audio Technology, 26129 Oldenburg, E-Mail:Lena.Schell-Majoor@idmt.fraunhofer.de; Jan.Rennies@idmt.fraunhofer.de*
[2] *Medizinsche Physik und Cluster of Excellence Hearing4All, Universität Oldenburg, 26111Oldenburg, E-Mail: Stephan.Ewert@uni-oldenburg.de; Birger.Kollmeier@uni-oldenburg.de*

Introduction

Sound quality evaluation is of interest in many practical applications, e.g. automotive, household appliances and other technical devices. Reliable instrumental quality measures enable a quick and low-cost assessment of sound quality on the one hand, while they can also contribute to the understanding of the principles underlying human sound perception, on the other hand. Several studies dealing with instrumental evaluation of sound quality combine different technical or psychoacoustic measures (e.g., loudness, sharpness, tonality) to predict sound quality of certain products, e.g., [1, 2, 3]. Typically, these individual sub-measures used in combination to yield an overall quality measure, mimic the process of auditory perception with specific processing stages optimized for each measure. An alternative approach is to employ a single perception model that incorporates the main features of the human auditory system. In this case, only one frontend would be used to calculate overall sound quality, which would allow for easy modification and individualization of the model. Furthermore, it might be applicable to a large variety of signal features as they occur in product and real-life sounds from different application areas. The Perception Model PEMO [4, 5] is an existing model that has been validated in a number of basic psychoacoustic experiments. It accounts for mechanisms such as forward masking, modulation detection and masking, and spectral masking. Recently, it also has been applied to predict detection thresholds of real signals [6]. PEMO consists of several signal processing steps motivated by the effective human auditory processing and transforms the time-signal into a so-called internal representation (IntRep). In this study the aim is to derive the psychoacoustic measures sharpness and roughness from the IntRep for simple time-invariant sounds in a very straight-forward way. The results are compared to data from [7].

Methods

Model Description

PEMO was applied as described in [5]. One of the main changes compared to the original version from [4] was the inclusion of a modulation filter bank, which is important in this study especially for deriving roughness. The input for the model is the time signal of the sound under investigation. The input is transformed into an IntRep, which typically is a three-dimensional representation (time vs. auditory filters vs. modulation filters) of the input sound and is supposed to contain its perceptually relevant features. In this study stationary signals were used. Therefore, two modifications were applied on the original model output, so that all further analyses are based on a two-dimensional IntRep. This two-dimensional IntRep was obtained by neglecting the onset of the signal by discarding the first half of the internal representation and by averaging over the second half. All stimuli had an overall duration of 2 seconds. The effective IntRep thus contained the model output in model units averaged over the last second of the signals IntRep. Figure 1 shows the resulting two-dimensional IntRep of a 1-kHz pure tone with a sound pressure level of 60 dB SPL as an example. The auditory filters are given on the x-axis and the modulation filters on the y-axis. The 0-Hz modulation filter represents the DC-part of the signal. The upper part of the IntRep (modulation filters with center frequencies > 0 Hz) contains information on the temporal fluctuations of the envelope, hence, the energy for the different modulation frequencies.

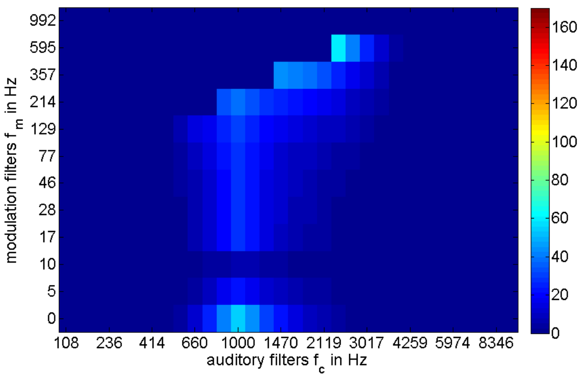

Figure 1: Example of the two-dimensional internal representation (IntRep) for a 1-kHz pure tone with a sound pressure level of 60 dB SPL.

Sharpness

The psychoacoustic parameter sharpness is related to the spectral content of the signal and has the unit acum [7]. Its calculation has been standardized in [8] and is based on weighting the specific loudness with a function where high frequencies are given increasingly more weight. The function and the entire formula are given in the standard. The stimuli for this part of the study were taken from [7], which also includes corresponding reference data. Three different noises were used. The first set of stimuli were narrowband noises with a bandwidth of one critical band and different center frequencies from $f_{c,min}$ = 1Bark to $f_{c,max}$ = 24 Bark. The second set included bandpass noises with a fixed upper cut-off frequency f_u = 10 kHz and lower cut-off frequencies varying from 2 to 23 Bark. The third set also included bandpass noises, but with a fixed lower cut-off frequency $f_l = 200$ Hz and upper cut-off frequencies from 2 to 23 Bark. To derive sharpness in acum from the IntRep, the standard had to be transferred to the IntRep. In the first step the DC-part was extracted from the IntRep and the center frequencies of the auditory channels were transformed into critical band rates. Then the original weighting function from the standard, which delivers weighting factors for the different critical bands, was applied. All of these weighted values were summed up and normalized to the sum of the unweighted values of the DC-part of the IntRep. The resulting value represents the sharpness of the signal with regards to the model units, *s*. To yield sharpness values in acum, a polynomial (2nd order) was fitted to the data for the narrowband noises with a bandwidth of one critical band and different center frequencies. This led to Eq. 1, which allows for the calculation of sharpness *S* in acum from the sharpness estimate, *s*, of PEMO (in model units):

$$S = -0.0015s^2 + 0.17s - 0.43 \tag{1}$$

It should be noted that this fit was derived from one condition only and then kept fix for the other conditions.

Roughness

Roughness is a psychoacoustic measure which is related to the modulation of a signal and the unit is asper. According to [7] it is defined on the basis of an amplitude-modulated (AM) pure tone with a carrier frequency $f_{ca} = 1$ kHz, a modulation frequency $f_{mod} = 70$ Hz, a modulation depth $m = 1$ and a sound pressure level $L = 60$ dB SPL. The roughness of this signal is defined as $R = 1$ asper. The roughness depends highly on f_{mod}. The perception of roughness starts at $f_{mod} \approx 15$ Hz, then the roughness increases up to $f_{mod} = 70$ Hz and decreases until the signal is not perceived as rough anymore at $f_{mod} \approx 300$ Hz [7]. According to this dependency, roughness was derived from the IntRep by summing over all values in the modulation channels from 17 Hz to 214 Hz. Just as for sharpness, Eq. 2 was generated to calculate roughness in asper from the summed model output (roughness estimate in model units), *r*, by fitting a polynomial (3rd order) to the data from the AM-tone with different modulation depths.

$$R = 1.84 \cdot 10^{-11} r^3 - 1.61 \cdot 10^{-7} r^2 + 6.16 \cdot 10^{-4} r - 0.55 \tag{2}$$

As for sharpness this fit was derived from this one condition only and was then kept fix for all other conditions.

The set of stimuli to test the prediction of roughness was based on data provided in [7]. Different AM-tones were used. First, the standard signal as described above ($f_{ca} = 1$ kHz, $f_{mod} = 70$, $L = 60$ dB SPL) was generated with different modulation depths from 0.2 to 1. Additionally, AM-tones with different carrier frequencies (250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz) and different modulation frequencies (from 10 Hz to 400 Hz), but a with fixed modulation depth $m = 1$ and a fixed loudness of $N = 4$ sone were part of this study.

Results

Figure 2 displays the sharpness calculated from PEMO. The corresponding reference data are shown in Fig. 9.1 of [7]. The solid line with square symbols represents data for the narrowband noise with different center frequencies f_c and a bandwidth of one critical band, which was also used to derive Eq. 1. The dotted line with the circles represents the bandpass noise with a fixed lower cut-off frequency, the dashed line with the diamonds those with a fixed upper cut-off frequency. There is quite a good agreement between the absolute values and the shape of the different curves. Only at the left and right edges of the plot some differences are obvious.

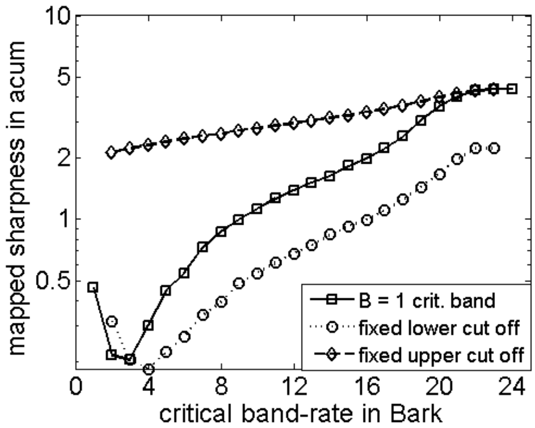

Figure 2: Sharpness for different narrowband and bandpass-filtered noises derived from PEMO's IntRep.

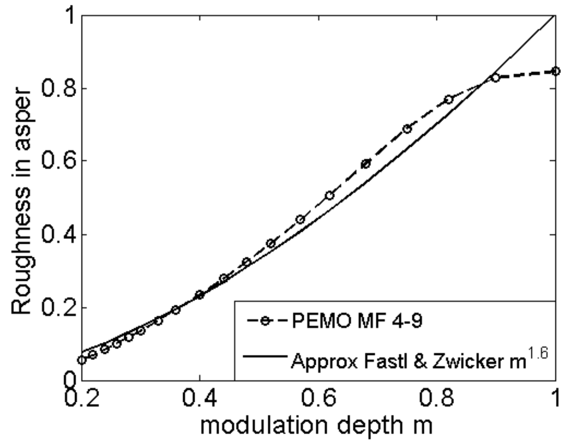

Figure 3: Roughness of an amplitude-modulated pure tone with a carrier frequency of 1 kHz, a modulation frequency of 70 Hz and a sound pressure level of 60 dB SPL for different modulation depth.

The results for roughness of signals with different modulation depths are shown in Figure 3 and Figure 4. Figure 3 shows data for roughness of the AM-tone ($f_{ca} = 1$ kHz, $f_{mod} = 70$, $L = 60$ dB SPL) as a function of modulation depth. The solid line in Figure 3 represents the approximation to subjective data from [7], the dashed line represents the results derived from the IntRep of PEMO. This is also the basis used for deriving Eq. 2. Close correspondence was achieved for lower modulation depths, while differences occurred at larger $m \geq 0.5$.

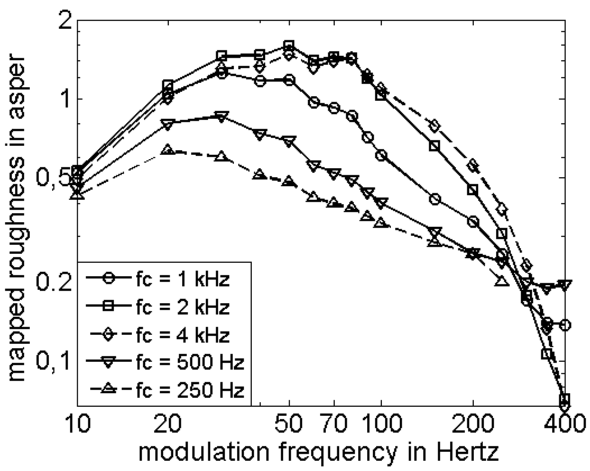

Figure 4: Roughness of amplitude-modulated pure tones with different carrier frequencies with a modulation depth of 1 and a loudness of 4 sone for different modulation frequencies.

In Figure 4 roughness derived from PEMO's IntRep is plotted for AM-tones ($m = 1$) with different carrier frequencies f_{ca} as a function of modulation frequency f_{mod}. The corresponding reference data are shown in Fig. 11.2 of [7]. The different curves represent the different carriers as denoted in figure legend. The general characteristics of the curves derived from PEMO and the reference data are similar, indicating low roughness for low f_{mod}, then increasing roughness up to a certain modulation frequency and then a decrease in roughness towards higher f_{mod}. In agreement with the reference data, the modulation frequency yielding maximum roughness changes with varying carrier frequency. However, while this effect is not found for $f_{ca} \geq 1$ kHz in the reference data, data from this study also shows a shifted maximum for $f_{ca} \geq 1$ kHz. Further differences between the datasets can be found regarding the absolute roughness values as well as the slope of the curves, which is shallower for data of this study than in the reference data.

Discussion

Using very simple means to derive the psychoacoustic parameters sharpness and roughness from the output of a perceptual model led to good results for sharpness when compared to the reference data. This was expected because the procedure to calculate sharpness from a time signal is described clearly in a standard and it has been validated well, especially for the set of artificial stimuli used in this study. For roughness the results from the model output did not match the reference data as well as for sharpness. Some general trends could be obtained by the output of the model, but also distinct differences were found. Partly, this could be due to the very simple approach, which did not include any sophisticated processing steps or transformations, but only a simple sum over a certain part of the IntRep. Modifying this procedure, e.g. by weighting the modulation channels differently, could improve the agreement between the model predictions and the empirical reference data. The results from this study support the idea of employing a perceptual model as a frontend for sound quality evaluation, but the transformation of the model's internal representation to sharpness and roughness has to be further developed and has to be validated on a larger set of stimuli.

References

[1] Ellermeier, W., Kattner, F., Kurtze, L., Bös, J..: Psychoacoustic characterization of the noise produced by photovoltaic inverters. Acta Acustica united with Acustica 100 (2014), 1120-1128

[2] Kim, E.Y., Shin, T.J., Lee, S.K.: Sound quality index for assessment of sound quality of laser printers based on a combination of sound metrics. Noise Control Engr. J. 61 (6) (2013), 534-546

[3] Kuwano, S., Seiichiro, N.: Subjective impression of copy machine noises: An examination of physical metrics for the evaluation of sound quality. Proceedings of Inter-noise 2009 (2009)

[4] Dau, T., Püschel, D., Kohlrausch, A.: A quantitative model of the „effective“ signal processing in the auditory system. I. Model structure. J. Acoust. Soc. Am. 99 (1996), 3615-3622

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

[6] Schell-Majoor, L., Rennies, J., Ewert, S.D., Kollmeier, B..: Application of psychophysical models for audibility prediction of technical signals in real-world background noise. Applied Acoustics 88 (2015), 44-51

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

[8] Deutsches Institut für Normung e.V., DIN 45692-2009-08: Messtechnische Simulation der Hörempfindung Schärfe. Beuth Verlag GmbH, Berlin, 2009