Title	聴覚末梢モデルを利用した音質評価指標の計算モデルの構築
Author(s)	磯山, 拓都
Citation	
Issue Date	2024-06
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19332
Rights	
Description	Supervisor: 鵜木祐史, 先端科学技術研究科, 博士

Abstract

Sound quality metrics (SQMs), such as loudness, sharpness, roughness, and fluctuation strength, have been used for modeling sensory pleasantness and annoyance, for product sound design, and for soundscape analysis and as SQMs for these purposes. Moreover, the ability to convey appropriate SQMs is considered essential in creating more desirable products and environmental sounds. SQMs can be modeled computationally. In particular, those related to temporal properties i.e., roughness, and fluctuation strength, have been modeled on the basis of time-domain auditory filters. The classical models of SQMs use the temporal variation of specific loudness measured using a loudness meter on the basis of ISO 532B:1975.

The international standards ISO 532-1:2017 and ISO 532-2:2017, provide two loudness models called the Zwicker and Moore-Glasberg models, respectively. The Zwicker model calculates the loudness for stationary and time-varying sounds, while the Moore-Glasberg model calculates only the loudness for stationary sounds. Moreover, these models have differences in regard to the auditory filter; the Zwicker model uses the Bark frequency scale based on an auditory filter with a symmetric shape whereas the Moore-Glasberg model uses the equivalent rectangular bandwidth (ERB) scale based on an asymmetric filter.

In regard to the ERB scale, it was derived in psychoacoustics studies on auditory filters, it covers the appropriate bandwidth of the auditory filter by taking into account the asymmetrical shape of the auditory filter and off-frequency listening. On the other hand, the Bark scale does not take these auditory characteristics into account. A physiology study showed that 1 ERB_N-number corresponds to approximately 0.9 mm of the cochlea length, demonstrating the physiological and psychological importance of ERB scale. Since perceived sound is processed by the brain via the auditory filterbank, the SQMs models should incorporate an elaborate auditory filterbank based on the ERB scale.

The Moore-Glasberg model is a stationary loudness model (based on long-term spectra) and can be used directly only to define sharpness. The model uses a rounded-exponential (roex) auditory filterbank that is defined in the frequency domain as a frequency-specific gain without specifying the exact impulse response. It is, therefore, difficult to use this model as a basis for constructing other SQMs model, e.g., classical roughness and fluctuation-strength models, from the time variability of loudness.

In this study, SQMs models were derived that utilize a time-domain auditory filterbank and involve a two-part calculation of specific loudness and modeling of SQMs. In particular, to calculate specific loudness, a time-domain GTFB or GCFB is employed instead of the roex auditory filterbank used in the Moore-Glasberg model. Minute changes in specific loudness, which are essential for modeling SQMs, are captured by the time-domain filterbank. Furthermore, the impact of asymmetric auditory-filter shapes on SQMs estimations was investigated using the GCFB and the GTFB. The specific loudness is then used to model loudness, sharpness, roughness, and fluctuation strength.

To obtain loudness, the summation of specific loudness was calculated; to obtain sharpness, calculation of the specific-loudness centroid was calculated; and to obtain roughness and fluctuation strength, they were calculated from the difference between the peaks and dips in the time direction of the specific loudness.

The proposed loudness model and the Moore-Glasberg model were compared in terms of the root-mean-squared error (RMSE) of loudness, after which the RMSEs of the three other SQMs models were compared with perceived SQMs. It was found that the proposed loudness model can be regarded as a time-domain model for calculating loudness based on the Moore-Glasberg model, as indicated by its very small RMSEs. In addition, it was found that the proposed sharpness, roughness, and fluctuation-strength models could explain actual SQMs better than the previous SQMs model, as evidenced by their lower RMSEs.

Keywords: Sound quality metrics, timbre, specific loudness, gammatone auditory filterbank, gammachirp auditory filterbank