

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: 鶴木祐史, 先端科学技術研究科, 博士

氏名	磯山拓都		
学位の種類	博士(情報科学)		
学位記番号	博情第529号		
学位授与年月日	令和6年6月24日		
論文題目	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		
論文審査委員	鵜木 祐史	北陸先端科学技術大学院大学	教授
	長谷川 忍	同	教授
	岡田 将吾	同	准教授
	鈴木 陽一	東北文化学園大学	教授
	水町 光徳	九州工業大学	教授
	山内 勝也	九州大学	准教授

論文の内容の要旨

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

論文審査の結果の要旨

音の質感を客観的に評価する指標としてシャープネス、ラフネス、変動強度といった音質評価指標がある。これらの指標は、聴覚系の処理を模擬した Zwicker のラウドネス計算法 (ISO532-1) を基盤として求められる。これは、音評価の基本 3 要素の一つであるラウドネス (主観的な音の大きさ) を音の客観的な情報に基づいて高精度に推定するものである。しかし、この方法では、Bark 尺度が利用されているため、蝸牛内の周波数分解能とその場所情報に完全に対応しておらず、さらには、聴覚フィルタ形状の非対称性 (音圧依存による形状の非対称性) が考慮されていない。この方法に代わる新しい計算法として、Moore & Glasberg のラウドネス計算法 (ISO532-2) がある。この方法では、ERB 尺度が利用され、音圧に依存したフィルタ形状の非対称性も考慮されて、聴覚末梢特性をよく反映したものとして利用されている。しかし、この方法では、インパルス応答を持たない聴覚フィルタで構成されているため、時間領域のラウドネス計算法として、その実現性の課題が残っていた。さらに、これを基盤とする音質評価指標の開発にも未着手の状況であった。

本論文では、これらの問題を解決するために、時間領域で定義された Gammatone/Gammachirp 聴覚フィルタバンクを基盤としたラウドネス計算法の確立を目指した。また、時間領域モデルを用いることにより、快適な生活空間の快適さを評価する上で重要な音質評価値であるシャープネスやラフネス、変動強度の推定を時間領域で行うことも可能とした。ここでは、定常音の Moore-Glasberg モデルを元に、時間領域で定義される聴覚フィルタバンクを用いたラウドネスの計算モデルを構築し、これを基盤として構築された音質評価指標の計算モデルが音質評価指標に関する主観評価データを説明できるか検討した。その結果、(1) 提案法が、Moore-Glasberg のラウドネスの計算法の時間領域版と解釈できること、(2) ISO532 で定義されたものと同程度のラウドネスの推定精度を提案法が有すること、(3) 精緻化された聴覚末梢系モデルを使うことにより、シャープネスやラフネス、変動強度といった音質評価指標の推定誤差を減少させることができることを示した。

以上、本論文は、精緻化された聴覚末梢系モデルとして Gammatone/Gammachirp 聴覚フィルタバンクに基づき、ラウドネスの計算法ならびに音質評価指標の計算法を確立した。この着眼点は新規性と独創性を持ち、その成果が国際的に定評のある学術誌に掲載されるなど学術的水準も高い。本技術は、聴知覚メカニズムの解明の一助になり、応用範囲が広く、学術的に貢献するところも大きい。よって博士 (情報科学) の学位論文として十分価値あるものと認めた。