Title	Spectro-Temporal Modulation分析を用いた音質評価指標と聴覚的顕著性の対応関係の検討
Author(s)	田中, 聡一郎
Citation	
Issue Date	2022-12
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18169
Rights	
Description	Supervisor: 鵜木 祐史, 先端科学技術研究科, 修士(情報科学)

Investigation of the relationship between sound quality metrics and auditory saliency using Spectro-Temporal Modulation analysis

2010114　Soichiro TANAKA

Hearing has the widest range of information that can be acquired from any human sensory organ, and to this day it is indispensable to our rich life. In daily lives, humans rarely hear only one sound, instead, they hear multiple sounds coming from multiple sources simultaneously, i.e., they hear certain sounds and voices actively by directing their own attention to them. There is a phenomenon exists in which a particular sound is noticed, or the presence of a particular sound can be recognized simultaneously with the occurrence of the sound in a place where multiple sound sources are present. Thus, auditory saliency is introduced as the sounds that unintentionally attract the ear without actively directing attention to a particular sound.

The characteristics of sound that affect auditory saliency are important for the design of easily perceptible alarm sounds and for the design of comfortable sound environments in general. As mentioned above, the auditory system processes sound in a 360-degree range. Therefore, hearing plays an important role in situations that require a quick understanding of the surroundings, such as the detection of a car or sirens that warn of danger. Auditory saliency has been examined as passive auditory attention, focusing on various acoustic features. Nevertheless, the nature of auditory saliency has yet to be elucidated, as the correspondence between the two was investigated directly by hand. When people hear sounds, differences in timbre are also relevant. Therefore, differences in timbre may also play a central role in auditory saliency.

The relationship between acoustic features and auditory saliency has been investigated in a study by Kidokoro et al. based on Spectro-Temporal Modulation Spectrum (STMS) (2021). The results show that the degree of temporal variability of STMS is related to auditory saliency under conditions in which the loudness of the sound stimulus is controlled. Nevertheless, the direct investigation of the relationship between acoustic features and prominence only explains the one-to-one relationship. Another study by Arnal et al. (2015) used STM analysis to investigate the characteristics of screams with high auditory saliency. In this study, STM analysis was shown to be able to analyze acoustic features and sound quality metrics and to link acoustic features and sound quality metrics via STMS.

These results exhibit that features appearing in the STMS can be used to the explanation of the relationship between sound quality metrics and auditory saliency. The purpose of this study is to clarify the relationship between sound quality metrics (sharpness and roughness) and auditory saliency via features appearing in the STMS. To achieve the goal of this study, which is to determine the relationship between sound quality metrics and auditory saliency, 10 stimuli with known high and low auditory saliency are analyzed for STM. Additionally, sharpness and roughness were calculated using a computational model. Clarifying the relationship between sound quality metrics and auditory saliency via features appearing in STM would not only help to elucidate the mechanisms of human auditory perception, but it could also allow us to change auditory saliency by controlling the sound quality metrics.

Therefore, the following steps will be taken in this study.

1. STM analysis of stimuli of known auditory saliency

2. Analyze STMS using higher-order statistics

3. (1) Analyze the relationship between sound quality metrics and auditory saliency, (2) Analyze the relationship between features obtained from STMS and sound quality metrics, (3) Analyze the relationship between features derived from STMS and auditory saliency, (4) Consider the relationship between STMS-derived features and sound quality metrics and auditory saliency

In this study, STM analysis, which has been used in several studies, is used to examine this issue. The filter bank used in the study was a constant band gammatone filter bank, with a frequency bandwidth of 160 Hz per filter per channel based on previous studies and the definition of roughness, and a few divisions of 551 channels. The temporal modulation region of 80 Hz and the spectral modulation region of 0 to 3.5 cycl/kHz were defined as the analysis range in the STM. In this study, 10 stimuli, including natural and artificial sounds, used in a study examining the relationship between pupil response and auditory saliency were included in the analysis. This stimulus is the one for which the auditory saliency scale was calculated by Thurston's pairwise comparison method. In this study, the loudness of all stimuli was controlled for in the analysis based on the study by Kidokoro et al. For loudness control, Moore & Gracebarg's loudness model was used to control by adjusting the sound pressure level in all stimuli so that the maximum loudness of the stimuli was 65 Phon. Previous studies have shown that the features appearing in STM are related to auditory saliency. This study focused on the distribution of STMS as a feature related to spectral modulation and temporal modulation.

For the sake of obtaining distributions, mean and variance, skewness, and kurtosis of the higher-order statistics were obtained, and the relationship between the sound quality evaluation index and auditory saliency was examined. The statistics were calculated assuming a single normal distribution for the distribution of the values of the amplitude spectrum. The correlation coefficients between the analysis results and the salience scale were calculated. Auditory saliency was considered explainable when the number of relations was greater than r=0.5 and the significance level was less than 0.05.

The purpose of this study is to clarify the relationship between sound quality metrics (sharpness and roughness) and auditory saliency via features appearing in the STMS. Therefore, 10 stimuli with known auditory saliency were first subjected to STM analysis. Then, higher-order statistics (center of gravity, variance, skewness, and kurtosis) in the STMS distribution were obtained, and correlations between the sound quality metrics and the salience scale were determined. The results show the following about the relationship between STMS-derived features and sharpness, roughness, and auditory saliency, respectively.

・Skewness and kurtosis in the STMS distribution were found to be significantly positively correlated with auditory saliency

・Skewness and kurtosis in the STMS distribution were found to have a significant positive correlation with roughness, one of the sound quality metrics

・Sharpness and roughness, two measures of sound quality metrics, were not correlated with auditory saliency although showed a significant trend

These results indicate that it is possible to link the roughness of the sound quality metrics to auditory saliency via the higher-order statistics of the STMS.