

Title	音質評価指標を用いた音の快・不快の予測モデルの構築
Author(s)	松尾, 洋杜
Citation	
Issue Date	2026-03
Type	Thesis or Dissertation
Text version	author
URL	https://hdl.handle.net/10119/20528
Rights	
Description	Supervisor: 鷗木 祐史, 先端科学技術研究科, 修士(情報科学)

Study on constructing a model for predicting of
the pleasantness/unpleasantness of sounds
using sound quality metrics

2410173 Hiroto Matsuo

Perceived impressions of sounds are determined not only by physical magnitude but also by sensory attributes such as pleasantness and unpleasantness. Even when sound levels are similar, impressions may differ due to temporal fluctuation, spectral balance, and tonal components. For practical applications including product sound quality design and assessment of acoustic environments, it is therefore important to predict pleasantness/unpleasantness from interpretable descriptors. Although data-driven approaches can yield accurate predictions, the relationship between acoustic factors and perceptual outcomes may be difficult to interpret, which limits their direct use for sound modification or design decisions. In contrast, psychoacoustical motivated sound quality metrics (SQMs) provide interpretable features that are closely related to auditory perception and can be linked to concrete modification strategies. The objective of this thesis is to construct an interpretable prediction model for sensory pleasantness and unpleasantness by integrating representative SQMs, and to evaluate whether the proposed framework is valid not only for controlled stimuli but also for unknown real environmental sounds.

This study was conducted as a continuation of prior work and was organized around four research questions. First, it was examined whether sensory pleasantness and sensory unpleasantness can be represented on a single perceptual axis, rather than being treated as independent dimensions. Second, it was examined whether the stimulus conditions can be designed to cover a sufficiently wide perceptual range for discussing pleasantness/unpleasantness in a stable manner. Third, it was examined whether an SQM-based model can explain subjective evaluation results in an interpretable way. Fourth, it was examined whether the proposed model remains effective for unknown sounds, including real environmental sounds that are generally nonstationary and acoustically complex.

To address these questions, the study proceeded in three compact steps. First, previously obtained subjective data were analyzed to clarify the relationship between pleasantness and unpleasantness and to motivate unified modeling. In this analysis, the subjective data were used to examine whether the paired relationship can be observed consistently across stimulus conditions and whether a common perceptual axis can summarize evaluation results in a compact way. This perspective is important for a unified treatment because practical prediction requires a common evaluation target that can be interpreted in terms of specific psychoacoustic attributes, rather than separate dimensions that may lead to ambiguous interpretation for sound modification and design decisions. Second, a listening experiment was conducted using systematically designed artificial stimuli.

Sensory pleasantness was evaluated by a paired-comparison procedure based on a classical psychometric framework, and continuous perceptual scale values were derived from aggregated paired judgments. The paired-comparison procedure was selected because it supports stable evaluation when perceptual impressions are subtle and when participants may differ in their use of absolute rating scales, which is often the case for

pleasantness/unpleasantness evaluation. The stimulus set was constructed so that impressions would span from pleasant to unpleasant, including intermediate conditions, which is necessary for stable modeling and meaningful interpretation. This design also supports a consistent discussion of how psychoacoustical motivated descriptors relate to perceptual outcomes across a sufficiently wide perceptual range, including conditions where temporal fluctuation, spectral balance, and tonal components change while sound levels are similar. Third, an interpretable prediction model was formulated and evaluated.

Five SQMs widely used in psychoacoustic modeling were integrated as explanatory variables: roughness, sharpness, tonality, loudness, and fluctuation strength. These SQMs were selected because they represent complementary aspects of auditory perception and provide interpretable descriptors that can be linked to concrete modification strategies in product sound quality design and assessment of acoustic environments. Following a conventional psychoacoustic modeling approach, an exponential-form representation was adopted so that the contribution of each metric is explicitly expressed through model parameters while allowing nonlinear sensitivity when required. In this formulation, model parameters play an important role not only for prediction but also for interpretation, because the fitted parameters indicate how each SQM contributes to sensory pleasantness and unpleasantness in the present framework. Model parameters were estimated so that the model output reproduces the tendency of the subjective scale values across the designed stimulus conditions.

Model validity was examined from two viewpoints. For controlled stimuli, it was confirmed whether the proposed model can explain the subjective evaluation trend over the full range of designed conditions, and whether the fitted parameters provide a consistent interpretation of how each SQM contributes to pleasantness/unpleasantness. In addition, the controlled stimuli evaluation supports discussion of whether the model can reproduce systematic changes in perceptual outcomes when temporal fluctuation, spectral balance, and tonal components are manipulated in a designed and interpretable manner.

For generalization, the proposed model was applied to real environmental sounds that were not used during model construction, and model predictions were compared with subjective ratings obtained in an additional listening test. This generalization viewpoint is essential for practical applications, because real environmental sounds are typically nonstationary and acoustically complex, and impressions can be influenced by time-varying components and mixtures of sources even when the overall physical magnitude is similar. In addition, representative conventional models designed for separate targets, such as pleasantness-oriented or annoyance-oriented prediction, were used as references to examine the benefit of a unified SQM-based formulation on a common evaluation target. This comparison clarifies whether a unified prediction model can provide meaningful predictions for unknown sounds while retaining interpretability and while avoiding target-dependent formulations that may be difficult to interpret for sound modification or design decisions. Furthermore, robustness was examined by computing

SQMs under an alternative auditory filtering framework based on an ERB-related filterbank, because the numerical values of SQMs can depend on the auditory preprocessing model. This robustness examination is practically relevant because the same psychoacoustical motivated metric may produce different numerical values depending on the auditory preprocessing model, and an effective prediction model should remain valid under such alternative SQM computation when the objective is to provide interpretable descriptors that are closely related to auditory perception.

Overall, the investigations supported a unified treatment of sensory pleasantness and sensory unpleasantness and demonstrated that an interpretable SQM-based model can be constructed to explain subjective evaluation results and to provide meaningful predictions for unknown sounds. The proposed model is expected to be useful as a basis for future work toward practical sound modification and design, where perceptual improvements can be discussed in terms of specific psychoacoustic attributes. The results showed that,

(1) Sensory pleasantness and sensory unpleasantness were supported to form a paired relationship on a common perceptual axis.

(2) The stimulus conditions were designed to provide a sufficiently wide range for pleasantness/unpleasantness evaluation.

(3) An interpretable prediction model was constructed by integrating five SQMs and was shown to explain subjective results.

(4) The proposed model was shown to remain effective for unknown sounds, including real environmental sounds, and to be robust to alternative SQM computation.

Based on these results, the sound pleasantness/unpleasantness prediction model proposed in this thesis was shown to be an effective new prediction model that can evaluate sensory pleasantness and sensory unpleasantness within a single unified model while retaining interpretability through SQMs.