

Title	聴覚的顕著性とスペクトル・時間変調情報の関係
Author(s)	木所, 晃利
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17092
Rights	
Description	Supervisor: 鷓木 祐史, 先端科学技術研究科, 修士 (情報科学)

The relationship between auditory saliency and
spectro-temporal modulation information

1910079 Akitoshi KIDOKORO

Sound, as a physical phenomenon, is a kind of wave. Humans perceive these waves through the auditory system, starting from the auricle, one of the auditory organs, and ending with the auditory nervous system and the auditory cortex of the cerebrum. The waves provide us with a great deal of information necessary for daily life. For example, the sirens that announce emergencies and the time signals on the radio that tell the time. In this way, our daily life can be said to be lived with sound. When listening to sounds, our ability to consciously select and listen to sounds is known as active auditory attention or the cocktail party effect.

On the other hand, there are sounds that we can recognize even if we are unconscious of their existence. The degree of conspicuousness of these sounds that can be recognized unconsciously is called auditory saliency, and it has been studied as an aspect of passive auditory attention by focusing on various acoustic features.

Auditory saliency The first model of auditory saliency was developed by focusing on Intensity, Temporal contrast, and Spectro contrast in the sound spectrogram. They reported that the model was able to explain the results of listening experiments, indicating that acoustic features related to intensity, temporal contrast, and frequency contrast were related to saliency. Subsequent studies of the model extended the first model to include Temporal Modulation (TM), which is a spectrum that varies along the time axis in the spectrogram, Spectro Modulation (SM), which is a spectrum that changes in the direction of the frequency axis, and Spectro-Temporal Modulation (STM), which is a spectrum in which both of them exist and change simultaneously. The results of the listening experiments showed that there was a significant correlation between the stimuli judged to be salient by the listeners and those judged to be salient by the model. This indicates that the acoustic features related to SM, TM, and some parts of STM are related to saliency. On the other hand, while investigating the relationship between auditory saliency and pupil diameter response, some studies have focused on the relationship between psychological quantity of sound and saliency, and investigated the relationship between loudness and saliency. The results showed that loudness and saliency are correlated, and it is thought that acoustic features related to loudness are also related to saliency. Following this study, other studies have focused on the relationship between acoustic features and saliency and investigated the relationship between loudness,

duration of acoustic features, and spectral structure and salient stimuli in environmental sounds. They suggested that spectral structure and duration of acoustic features also contribute to saliency. They also suggested that a single acoustic feature alone cannot account for saliency. However, even in the various studies presented so far, no consistently explanatory acoustic feature for auditory saliency has been identified.

The purpose of this study was to investigate what acoustic features appear in STM that can be observed, including the interaction between SM and TM, and to clarify how these features are related to auditory saliency. In this study, we examined the acoustic features that appear in STM including SM and TM. Therefore, this study focuses on STM analysis, which integrates SM and TM, and investigates the relationship between auditory saliency and acoustic features obtained from the results of STM analysis. In this study, the following steps are taken for this purpose: 1. prepare stimuli whose saliency is already known from previous studies. 2. perform STM analysis on the prepared stimuli to obtain STM. 3. analyze the acoustic features that are considered to be related to saliency based on the results of STM analysis and the findings of previous studies, such as average power, frequency spectrum spread, harmonicity, and temporal modulation. 4. Analyze to what extent the features obtained in 3 contribute to the saliency. 5. consider the contribution of the acoustic features obtained from the STM to the saliency based on the analysis results obtained in step 4.

The STM analysis used in this study is based on a method used in speech recognition and described in other review papers. The method is based on a two-dimensional Fourier transform of the spectrogram obtained from the filter bank. The advantage of this method is that it is easy to quantitatively measure the acoustic features from the final STM, and it is also possible to create quantitatively controlled stimuli by performing 2-D filtering and inverse transformation on the STM obtained by this method. The filter bank used for the STM analysis in this study is a constant-band gammatone filter bank, and adjacent filters intersect at a point of -3 dB. In this STM analysis, the filter bank and the frequency bandwidth per channel are related to the resolution of the SM. The narrower the filter, the higher the resolution of the SM, but the lower the resolution of the TM due to the effect on the amplitude envelope in the low frequency range. Based on the trade-off between these two factors and the characteristics of the stimuli used in this study, the frequency bandwidth of the filter per channel was set to 80 Hz. For the above reasons, the range of analysis was -40 Hz to $+40$ Hz on the TM axis and 0 cyc/kHz to 6.5 cyc/kHz on the SM axis.

In this study, the research strategy was to conduct STM analysis on the stimuli whose saliency was already known, and to further analyze the acous-

tic features that were thought to contribute to the saliency from the obtained STM. In this research, we need to understand the acoustic features that contribute to the STM. For this purpose, we created amplitude-modulated signals, harmonic complex tones, spectral structures, and frequency-modulated signals, and analyzed their STM. As a result, the information that appears in the STM analysis when the amplitude modulation signal, harmonic complex, spectral structure, and frequency modulation signal are input was found. Next, to investigate the relationship between auditory saliency and STM, 10 stimuli with known saliency at the Thurston scale were subjected to STM analysis. After that, we analyzed the results of the STM analysis and the findings of previous studies for mean power, frequency spectral spread, harmonicity, and temporal modulation, which were considered to be related to saliency. Correlation coefficients were calculated between the results obtained from the analysis and the prominence at the Thurston scale. As a result, the acoustic features obtained only from the SMTM did not have a significant correlation with the saliency. In order to investigate the correlation coefficient between the acoustic features on the STM plane and the saliency, we integrated the features from only SM and TM and calculated the correlation coefficient with the saliency scale. As a result, a weak correlation trend was observed. However, there were no acoustic features that were completely uncorrelated. These results indicate that the acoustic features related to SM and TM on the STM plane are related to auditory saliency. These results suggest that it is possible to examine auditory saliency using the STM analysis used in this study.