Title	Study on noise suppression based on the spectro-temporal modulation
Author(s)	Putri, Fanda Yuliana
Citation	
Issue Date	2023-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18749
Rights	
Description	Supervisor: 鵜木祐史, 先端科学技術研究科, 修士(情報科学)

Abstract

Noise suppression is an open-area research that addresses the challenge of obtaining speech from noisy speech as the result of environmental noise. The primary challenge of noise suppression is effectively attenuating the noise without substantially damaging the intricate balance between the quality and intelligibility of the noise-suppressed stimuli. Hence, achieving this delicate balance requires advanced consideration of the characteristic of speech and various noises in the respective features for noise suppression. An aggressive noise suppression may lead to the removal of important speech cues, affecting speech clarity and intelligibility in scenarios where speech and noise overlap in time; distinguishing between the two and suppressing noise while preserving speech components becomes more challenging.

In noise suppression algorithms that utilize spectral features, musical noise is a common problem. It refers to an undesirable artifact that occurs when certain noise components are mistakenly identified as speech and are suppressed inappropriately, resulting in a musical or tonal quality in the enhanced speech signal. This artifact often manifests as a periodic or rhythmic sound, resembling a musical note or a buzzing sound, which can be highly distracting and adversely affect speech intelligibility.

This study investigates a novel approach by integrating spectro-temporal modulation (STM) with a statistical noise suppression method, the minimum mean square error short-time spectral amplitude (MMSE-STSA) noise suppression algorithm, to achieve improvement in the noise suppression result. The research begins with a detailed exploration of the analysis-synthesis pipeline for STM feature extraction, highlighting the distinct spectro-temporal characteristics exhibited by speech and noise in the STM domain. Notably, the study delves into the averaged STM features of white, pink, and factory noises, revealing nuanced differences in their spectro-temporal properties, thus deepening an understanding of their impact on noise suppression.

The proposed analysis-modification-synthesis (AMS) pipeline is introduced, where the conventional noise suppression block is replaced by the implementation of the MMSE-STSA algorithm using the STM feature. This strategic integration leverages the joint spectro-temporal information provided by STM to adaptively obtain the estimation of speech, thereby enhancing the efficacy of the noise suppression algorithm. Moreover, based on empirical findings, the study uncovers the potential benefits of incorporating over-suppression in the STM domain, leading to further improvement in noise reduction results. As a consequence, the MMSE-STSA algorithm undergoes modification to accommodate the over-suppression, enhancing its noise reduction capabilities.

Parameter tuning is conducted to optimize the attenuation gain (β) and enhance noise reduction, particularly in the speech-dominant groups of the STM domain. The study reveals specific β values that lead to better noise suppression results.

An evaluation of the efficacy of the proposed noise suppression algorithm utilizing the STM feature was conducted in comparison to established statistical noise suppression methods based on spectral features, namely the Wiener filter and the MMSE-STSA noise suppression algorithm. Three objective evaluation metrics, Segmental Signal-to-Noise Ratio (SNR), Perceptual Evaluation of Speech Quality (PESQ), and Short-Time Objective Intelligibility (STOI), are utilized to assess the effectiveness of the proposed method.

The results indicate that the proposed algorithm achieves significant improvements in speech intelligibility, as demonstrated by higher STOI scores compared to other algorithms. However, the overall audio quality, as measured by PESQ, does not consistently surpass the benchmark methods. Despite this, the research identifies the potential of STM as an alternative feature for noise suppression, offering unique insights into the characterization of clean speech and various noises in the modulation domain.

Conclusively, the proposed STM-based noise suppression algorithm shows promise in enhancing speech intelligibility. While further research is needed to address certain limitations, such as the focus on single-channel uncorrelated noise and the non-linear nature of STM, the exploration of STM in noise suppression provides valuable contributions to this area of research. This study encourages future investigations to advance noise reduction algorithms and explore the potential of STM in various audio processing applications.

Keywords: noise suppression, temporal modulation, spectral modulation, spectro-temporal modulation, wiener filter, MMSE-STSA