Noise reduction is one of the most important elements in speech communication and recognition systems. Recently, many researchers have paid attention to the multi-microphone techniques which can exploit the spatial diversity to discriminate desired signal from undesired noises. A variety of beamforming or multi-channel filtering techniques have been successfully applied to the multi-microphone noise reduction.

This dissertation focuses on the multi-channel Wiener filter (MWF) which is shown to have better performance than the standard beamforming techniques. In the MWF, the estimation of noise statistics is an important issue as in most noise reduction techniques. In the conventional MWF, the noise statistics are recursively estimated with a forgetting factor during noise-only periods and kept fixed during speech-present periods based on a voice activity detector (VAD). For better noise estimation, two approaches are proposed in this dissertation. Firstly, a time varying forgetting factor is proposed for the noise estimation instead of the conventional scheme that needs an explicit VAD. The time varying forgetting factor is parameterized from the normalized cross correlation (NCC) between microphone pairs. Secondly, the frequency domain multi-channel Wiener filtering and subspace decomposition based techniques are investigated. In speech-present periods, the multi-channel input signals are decomposed into speech and noise spatial subspaces. The noise eigenvalues are modified in order to update the noise statistics not only in the noise-only period but also in the speech-present period. Three methods are proposed for the noise eigenvalue modification, which are based on the rank-1 property of the speech spatial spectral matrix for the single speech source.

VAD plays an important role for noise reduction techniques which estimate noise statistics and adjust filter coefficients during speech-absent intervals to avoid speech cancellation. Most single channel techniques are unreliable in the presence of non-stationary or broadband speech-like noise. Two techniques are proposed for robust VAD using multi-microphones. First, Cross Power Spectrum Phase (CPSP) based technique is investigated in the presence of coherent interference using two microphones. Under the assumption that the direction of desired speech signal is known and the time delay between microphones is compensated, the Averaged CPSP (A-CPSP) can be utilized as a VAD measure. In order to improve the VAD performance in the presence of strong coherent interference from other direction, a Maximum Partially Averaged Real CPSP (MPA-RCPSP) method is proposed, which detects the cophased frequency region with high Signal-to-Interference Ratio (SIR). Another multi-channel VAD technique is proposed based on the phase vector. This technique exploits the spatial information obtained by the eigendecomposition of multi-channel correlation matrix in the frequency domain. The phase vector is used as a measure for VAD, which is derived from the principal eigenvector. Voice activity is detected by the log likelihood ratio test under the assumption that phase vectors of speech-absent and speech-present signals have complex Gaussian distributions.

* Note: The text above is the abstract of the thesis. You can use the full-text by clicking the “Link” (URL) of item record and the viewer program will be installed automatically. In case of having problem with the viewer program installation, please download it manually using this url http://pdfdrm.snu.ac.kr/snudrm/download/PDFReaderSetup.exe