

SUMMARY OF Ph.D. DISSERTATION

School Integrated Design Engineering	Student Identification Number	SURNAME, First name HIOKA, Yusuke
Title A Research on Voice Activity Detection and Speaker Direction Estimation Using Microphone Array		
<p>Abstract</p> <p>Due to the recent progress of digital processors and the rapid spread of broadband data transmitting channels, the application of digital signal processing to multimedia signals is now part of our daily lives. Moreover, the utilization of speech for human-machine interface is in the spotlight, because it simplifies the complicated procedures in the operation of systems. In various applications using speech, the information about the speaker is useful. In particular, the voice activity detection (VAD) identifies “When the speaker utters?” and the speaker direction estimation identifies “Where the speaker’s location is?”. Both of these are often used as preliminary data. This dissertation describes the research results achieved using specific methods to acquire this data using microphone array signal processing.</p> <p>1. Voice activity detection using speech features in multiple signal domains (Chapter 3)</p> <p>Conventional VAD methods suffer from the performance degradation due to nonstationary interference that often occurs in practical environments. Added to this, the discrimination capability of VAD is reduced when the speech and interference arrive from close directions. This research proposes a VAD method achieved by array signal processing in the wavelet domain. Since the method used the speech signal features in temporal, spectral, and spatial domains, it succeeded correctly in discriminating the segments of speech from interference.</p> <p>2. Speaker direction estimation with uniform accuracy for omni-direction (Chapter 4)</p> <p>The methods proposed for speaker direction estimation provided uniform accuracy for omni-direction. This method employed equilateral-triangular microphone array and achieved uniform estimation accuracy for omni-direction. This was achieved through the integrated use of the data extracted from microphone pairs on each of the three sides. The estimation accuracy was successfully improved by selecting the harmonics of voiced sound consisting of major speech components. In addition, a method is proposed to estimate both azimuth and elevation angles using tetrahedral microphone array.</p> <p>3. Tracking of moving speaker direction (Chapter 5)</p> <p>The proposed speaker direction tracking method is a refinement of 2. above. In this method, the performance index was changed during the adaptation to achieve fast and accurate global convergence to cope with abrupt movement of speaker direction.</p> <p>In each situation, the effectiveness of the methods through experiments employed was verified in real acoustic environment.</p>		