報告番号	※甲　　　第　　　号

主　論　文　の　要　旨

論文題目　Blind source separation with the low computational costs for the mobile and portable speech equipment
（小型音声処理装置のための低演算量で実現可能な音声音源分離に関する研究）

氏　　名　近藤多伸

論　文　内　容　の　要　旨

In this dissertation, frequency bin selection is proposed as a method to reduce the computational cost of blind source separation (BSS) based on frequency domain independent component analysis (FDICA). Clear voice quality is expected by users when communicating through speech processing equipment such as teleconferencing equipment and mobile telephones. Speech processing equipment is usually manufactured with embedded processors, especially digital signal processors (DSP), because of their low electric power consumption and the need for real-time processing. BSS has been widely investigated for use in speech enhancement applications for the purpose of obtaining higher voice quality, and FDICA is one of the most popular methods used by researchers to perform BSS. This is because acoustic conditions are usually reverberant, and FDICA is an inverse filter method, and thus minimizes reverberation influence. However, the computational cost of FDICA is quite high because FDICA estimates inverse filter coefficients in each frequency bin using an optimization scheme employing higher-order statistics, usually an iterative update algorithm. Current DSPs can achieve high levels of performance, however the appropriate choice of a DSP strongly depends on the specifications of the speech processing equipment in which it is to be used. For mobile devices, due to battery life considerations, low-powered DSPs are normally used. ICA stores long-term observed signals because its optimization scheme employing higher-order statistics, so that the external memory usually consists of dynamic random access memory (DRAM). However, the required wait states of DRAM is another issue, because of the waste of electric power this involves. In addition, the amount of power wasted is not negligible. Therefore, implementing FDICA using embedded processors is difficult, despite the high performance of current DSPs. The proposed method aims at reducing computational cost by reducing the number of frequency bins in which the ICA algorithm is performed, thus reducing computational cost and increasing the feasibility of implementing FDICA in embedded processors. The proposed frequency bin selection method utilizes spatial correlation; the determinant of the spatial covariance matrix or the magnitude squared coherence (MSC) between two microphones.

Two types of frequency bin selection are proposed for FDICA, one for mobile telephone devices and one for portable speech processing equipment, such as portable teleconferencing systems. For mobile devices, the determinant of the spatial covariance matrix is used to select the frequency bins. The determinant is theoretically analyzed, since its characteristics simultaneously reflect both directional information as well as the power of the source signals. In other words, signal separability can be evaluated using the determinant. In the unselected frequency bins, a Wiener filter is obtained using the tentatively separated signals, which are the output of the null-beamformers. Use of the separated signals results in improved performance, because this cancels out the signal distortion caused for acoustic reasons by microphone array signal processing. Performance, as measured by the segmental signal-to-noise ratio, is experimentally evaluated, and significant improvement is achieved. Cepstral distortion is also employed to evaluate performance, but results show deterioration in performance instead of improvement as the number of frequency bins selected is reduced. The trade-off between these two measures of distortion is used as a criterion to determine the number of frequency bins selected, and it is determined that 64 bins should be selected. Compared to conventional FDICA methods, the proposed method achieves a more than 80 percent reduction in computational cost. Despite this large reduction in computational cost, the segmental signal-to-noise ratio is improved by about 2 dB, while deterioration in cepstral distortion is restrained to only about 1 dB.

For portable speech equipment, a dodecahedral microphone array (DHMA) is used as a small, agglomerative sound capture system. DHMAs have ten faces, with six microphones installed on each face. When performing BSS using DHMAs, FDICA is executed under overdetermined conditions, since the number of the microphones exceeds the number of source signals. The size of the

FDICA separation matrix when performing BSS with a DHMA is quite large, making this approach extremely computationally expensive. The permutation solution requires especially large computational resources because its clustering method involves a large number of similarity calculations. Magnitude squared coherence is utilized to select frequency bins for analysis. The shape of a DHMA is similar to that of a spherical microphone array, but its acoustic characteristics are somewhat different due to its many flat faces. Two arbitrarily chosen microphones are used to calculate magnitude squared coherence, and frequency regions are evaluated on this basis, with lower MSC values indicating the separable frequency region. Frequency bin selection is not uniformly spaced when using the proposed method. The resulting reduction in computational cost exceeds 80 percent, even with only a 68 percent reduction in the number of frequency bins selected. Separation performance, as measured by the signal-to-interference ratio, deteriorates, however, compared to FDICA which frequency bins are uniformly selected, the deterioration is restrained by about 1 dB. As for distortion, the segmental signal-to-noise ratio deteriorates slightly, although cepstral distortion is improved. For these distortion measures, compared to methods in which the frequency bins are uniformly selected, both measures are improved when using the proposed method.

Both frequency bin selection methods calculate the spatial correlation between only two microphones, and these spatial correlations are categorized as second-order statistics. This implies that BSS based on FDICA can become more computationally efficient, while maintaining the signal distortion level, through the combination of second-order statistics. When speech processing equipment is manufactured with DSPs, the required wait states of the DRAM external memory is an important issue to consider when implementing FDICA. FDICA must store long-term observed signals in DRAMs because the high-speed internal memory in the DSPs should not be taken up just to store these signals. By reducing the number of frequency bins selected, the proposed method also reduces required memory consumption, which restrains the amount of memory access required. The computational costs of conventional FDICA are clarified, and a target number of operations for the use of FDICA with internal DSPs are estimated: 200 mega-operations for mobile devices and 160 giga-operations for portable equipment. Calculations are made which show that the proposed method does not exceed the target computational cost. The function responsible for the dominant computational expense changes according to the number of frequency bins selected, indicating to engineers which software function should be optimized. Taking into consideration objective measures used to evaluate popular speech enhancement applications, about 10 to 20 dB is the allowable value for the absolute signal-to-interference ratio. Even though separation performance deteriorated when using the proposed methods, performance remained in the range allowable for practical commercial applications. On the other hand, measures of distortion remained at equivalent levels, while still achieving a more than 80 percent reduction in computational cost. As a consequence, the proposed method involves a practical trade-off between separation performance and computational cost on the one hand, and a quite advantageous trade-off between signal distortion and computational cost on the other. These findings indicate that the proposed BSS methods are practical, and that they are acceptable for use in speech processing equipment with embedded processors. Future work includes more investigation into the separation method used for the unselected frequency bins, as well as into on-line implementation.