氏　　　　　　　名	CHAU THANH DUC
学　位　の　種　類	博士(情報科学)
学　位　記　番　号	博情第 307 号
学 位 授 与 年 月 日	平成 26 年 9 月 24 日
論　　文　　題　　目	A Study on Binaural Sound Source Localization based on Equalization-Cancellation Theory　（等化-キャンセル理論にもとづいた両耳聴音源方向推定に関する研究）

論 文 審 査 委 員	主査	赤木　正人	北陸先端科学技術大学院大学	教授
		党　　建武	同	教授
		鵜木　祐史	同	准教授
		田中　宏和	同	准教授
		水町　光徳	九州工業大学	准教授

論文の内容の要旨

Simulating the human auditory system to deal with problems in sound signal processing is an interesting topic that has attracted a huge number of research recently. One of the important problems is binaural sound source localization (SSL), which plays a crucial role in binaural speech enhancement, binaural source separation and humanoid robot. Although previous research has achieved many impressive results in sound localization, the problem of binaural SSL in the presence of noise and reverberation has not been completely solved. This thesis aims at an effective SSL method based on the human hearing mechanism, which is able to work on binaural systems in practical noisy reverberant environments.

Binaural SSL is an important task in binaural signal processing field as it provides the location of sound source, commonly the direction of arrival (DOA) of the target sound. In the past decades, a large number of DOA estimation methods have been introduced, in which each one differs from others by the way of exploiting two main localization cues: the interaural time difference (ITD) and the interaural level difference (ILD). The well-known conventional GCC-PHAT method is based on only ITD and does not account well for noise. Therefore, there has been many research showing that it is not effective for binaural SSL. Azimuth-dependent models of binaural cues, such as joint estimations of ITD and ILD and DOA classification, have been presented. Although these research showed relatively good results by combining both ITD and ILD, their applicability in adverse noisy reverberant environments is still limited since there has been lack of methods accounting for the effect of interference signals efficiently. Methods directly based on head-related transfer functions (HRTFs) have also been studied, such as the inverse HRTF filtering and the cross-channels HRTFs. However, these methods highly depend on the HRTFs and suffer from reverberation because the

HRTFs vary largely along the reverberation levels.

In psychoacoustic research field, binaural hearing has been studied for more than a century and several theoretical models of binaural processing have been developed. Among them, the equalization-cancellation (EC) model of Durlach has received a significant attention as its description is consistent with the human perception on binaural data. The EC model was originally proposed to explain the phenomenon of binaural masking-level differences (BMLDs) in binaural detection. Due to its well performance on BMLD prediction, the EC model was further extended to selective hearing in the 'cocktail party' scenarios. This suggested that the EC model has great potential for sound localization and segregation in the presence of multiple interference signals.

Inspired by the EC model, this thesis investigates a binaural DOA estimation method based on the EC mechanism. The principle idea is that the EC procedures are first utilized to eliminate the sound signal component at each interest direction; the direction of sound source is then determined as the direction at which the residual energy is minimal. In order to make this idea applicable in practice, two approaches are proposed to accommodate it with the problem of SSL under the effect of noise and reverberation, resulting in two improved algorithms namely Adaptive EC-BEAM and Weighted EC-BEAM. The Adaptive EC-BEAM algorithm improves SSL performance by adapting the EC model to the level of reverberation in room, using the direct-to-reverberant energy ratio (DRR). The Weighted EC-BEAM algorithm deals with the problem in a contradict way, in which two weighted functions are applied to reduce the negative effects from the observed signals, without modifying the localization model. Improvement of the suggested algorithms is verified by experimental results in various noisy reverberant conditions.

The proposed Weighted EC-BEAM algorithm is then selected to apply in two binaural applications, speech enhancement and source separation, as its assumption is easier to be satisfied in practice. In the first application, the proposed method is employed to localize the meaningful sound signals for an intelligent speech enhancement system, which is able to extract and present the meaningful signals together with the target speech. The second application applies the proposed SSL method to estimate the DOAs of all sound sources before extraction (separation), resulting in a new blind source separation method. Experimental results showed that the Weighted EC-BEAM localized the desired sound sources correctly in both applications, from which the effectiveness of the proposed SSL method is confirmed

Keywords: Binaural sound localization, binaural hearing, Equalization-Cancellation model, noisy reverberant environments, humanoid robot

論文審査の結果の要旨

本論文は，ヒトの両耳聴モデルである等化－キャンセル理論（Equalization-Cancellation

Theory: E-C モデル）を音源方向推定に応用した研究に関する報告である。

　音源方向推定は，多数の音源が存在する環境において，音声強調，音分離などの前処理法として重要であり，カクテルパーティ効果を持つヒューマノイドロボットの聴覚を実現する上で基礎となる研究分野である。現在，単一マイクロホン（1-ch）を用いた手法，あるいは複数マイクロホン（マイクロホンアレイ）を用いた手法が提案されているが，1-ch では精度が悪く，マイクロホンアレイでは精度は良いものの設置面積が大きくロボットへの実装は困難という問題がある。このため，ヒトの両耳聴知覚（Binaural hearing）の知見のもとづいた 2-ch の音処理システムが注目を集めている。これまでにも，両耳信号間の相互相関により両耳時間差を推定する方法などが提案されてきているが，実環境での雑音・残響に弱いという問題点を克服できていない。

　本研究では，両耳信号間の差異には時間だけではなく音圧，頭部反射などが含まれることから，これらを総合することで雑音・残響環境に頑健となる音源方向推定法を提案している。具体的には，

(1) 基本となる手法として，ヒトの両耳聴モデルである E-C モデルを音源方向推定へ応用した EC-BEAM を提案した。

(2) EC-BEAM を残響環境でも使用できるように，残響環境の物理指標である Direct-to-Reverberant Ratio (DRR)を用いて EC-BEAM 中の両耳間伝達関数を適応的に変形する手法，Adaptive EC-BEAM を提案した。

(3) EC-BEAM を雑音・残響環境でも使用できるように，EC-BEAM への入力である両耳信号を環境に適応して重みづけする手法である，Weighted EC-BEAM を提案した。

(4) これらを用いることにより，従来法に比べて，雑音・残響環境下での音源方向推定での推定角度誤差が著しく低下することを確認した。

　提案法の応用として，提案手法の一つである Weighted EC-BEAM を音声強調と音源分離の前処理法として使用したところ，両応用例とも，提案手法を従来法と置き換えることで従来法よりもより高精度の音声強調，音源分離が行えることが明らかとなった。

　以上のように，本研究は新しい概念のもとで，ヒトの両耳聴モデルである E-C 理論を音源方向推定に応用し，雑音・残響環境下での方向推定角度誤差を従来法に比べて著しく低下させる手法を実現したものであり，学術的に貢献するところが大きい。よって博士（情報科学）の学位論文として十分価値あるものと認めた。