

機関番号：34315

研究種目：若手研究（B）

研究期間：2008～2010

課題番号：20700169

研究課題名（和文） サラウンドマイクロホンアレーを用いた音場トランスクリプションシステムの構築

研究課題名（英文） A design of acoustic sound field transcription system with surround microphone array

研究代表者

西浦 敬信（NISHIURA TAKANOBU）

立命館大学・情報理工学部・准教授

研究者番号：70343275

研究成果の概要（和文）：

サラウンドマイクロホンアレーを用いて音環境内のあらゆる音イベントを文字として記録するだけでなく、その音イベントの発生位置や時刻情報とあわせて構造化可能な音場トランスクリプションシステムの構築を試みた。その結果、音イベントの位置情報、時刻情報、擬音語表現を用いて、音環境全体をリアルタイムに構造化可能な音場トランスクリプションシステムの構築に成功した。

研究成果の概要（英文）：

We try to design an acoustic sound field transcription system with a surround microphone array. It consists of an acoustic sound dictation as recording any sound events with texts, a sound event localization, and a sound event active time estimation. As a result of our research, we could design the acoustic sound field transcription system that is possible to transcript any sound events based on the estimated positions, the estimated time-index and the estimated sound events with text in online.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,300,000	390,000	1,690,000
2009年度	1,100,000	330,000	1,430,000
2010年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：音情報処理

科研費の分科・細目：情報学 知覚情報処理・知能ロボティクス

キーワード：音声情報処理

1. 研究開始当初の背景

ここ数年、計算機を用いた音声認識装置は計算機の飛躍的な進歩だけでなく、隠れマルコフモデル（HMM）などの統計的手法の適用により、飛躍的に音声認識性能が改善され、すでに実用化されつつある。一方、音環境には音声以外の様々な音イベント（例えば、ドアを閉めた時の音やガラスが割れた時の音など）が存在する。これらの音イベントを人類は「ボタン」や「パリン」など擬音語によ

り言葉で表現することが可能である。人類が擬音語で表現できるということは計算機を用いた従来の音声認識技術の枠組みで音イベントも「文字として表現」もしくは「ある音イベントとして分類」できる可能性が大きい。そこで、音声だけでなく音環境全体の音イベントを構造化することができれば、我々の生活はより豊かになるのではないかという着想に基づき、本研究において音環境におけるあらゆる音イベントを文字として記録

するだけでなく、その音イベントの発生位置や時刻情報なども合わせて構造化する音場トランスクリプションシステムを提案し音環境のアーカイブ化を試みる。

2. 研究の目的

音場トランスクリプションシステムを構築する上で最も重要となるのが、音イベントの高音質な受音技術である。従来、音環境における高音質な音イベントの抽出ではマイクロホンアレーが注目されてきた。しかしながら従来のマイクロホンアレーは、発話者の声を受音する用途で設計されており、常に発話者がマイクロホンアレーに向かって発話しなければならないという大きな制約があった。この問題に対し本研究では、マイクロホンアレーを1箇所(1方向)集中的に配置する従来の受音器の枠組みを発展させたサラウンドマイクロホンアレーを提案する。サラウンドマイクロホンアレーでは、音環境の周囲全体にマイクロホンアレーを配置して音環境内のあらゆる位置において必ず直接音を受音できる環境を作り出すことで、高音質な音イベントの抽出を目指すものである。サラウンドマイクロホンアレーを用いて音環境内の音イベントの発生位置や時刻情報なども合わせて音場環境全体のトランスクリプションを行うことで、音響セキュリティシステムや防犯システムなど「安全・安心」を支える基礎技術としての確立を目指す。

3. 研究の方法

サラウンドマイクロホンアレーを用いて音環境内のあらゆる音イベントを文字として記録するだけでなく、その音イベントの発生位置や時刻情報と合わせて構造化可能な音場トランスクリプションシステムを構築するために以下の項目に対して研究を推進した。

(1) サラウンドマイクロホンアレーの最適配置問題の検討

音環境の周囲全体にマイクロホンアレーを配置して音環境内のあらゆる位置において必ず直接音を受音できる環境を構築することで、高品質な音イベントの抽出を目指す。しかしながら周囲にマイクロホンアレーを配置することは室内の物理的な制約も大きい。そこで本研究では音環境内の壁にマイクロホンアレーを設置する壁掛けサラウンドマイクロホンアレーによる高品質な音イベントの抽出を模索する。具体的には対象となる音環境において死角なく均一に音イベントを検出可能な壁掛けマイクロホンアレーの最適サラウンド配置アルゴリズム

の確立を目指す。

(2) サラウンドマイクロホンアレーによる音イベントの位置および発生検出

最適配置された壁掛けサラウンドマイクロホンアレーを用いて、従来のアレー信号処理の枠組みを拡張したサラウンド信号処理理論に基づく新しい音イベント検出(位置および発生検出)方法を模索する。

(3) サラウンドマイクロホンアレーによる音イベントの高音質受音

壁掛けサラウンドマイクロホンアレーとサラウンド信号処理理論に基づく新しいビームフォーミングの開発を行う。

(4) 上記に基づく音場トランスクリプションシステムの構築

サラウンドマイクロホンアレーにより高音質に抽出した音イベントを文字として認識することに加えて、音イベントの位置および発生検出結果を合わせて記録可能な音場トランスクリプションシステムの構築を試みる。特に音イベントを文字として認識する手法については、従来の擬音語としての認識(例えばドアが閉まる音なら「バタン」)手法をさらに拡張し、音イベント内容にあわせて分類(例えばドアが閉まる音なら「ドアの開閉」)することで、アーカイブ価値の高い音場のトランスクリプション手法を試みる計画である。(図1参照)



[音場トランスクリプション結果]

時刻	場所	音イベント内容
14:16	A地点	ドアの開閉
14:18	B地点	テレビスイッチ ON
14:19	C地点	メール到着

図1: 音場トランスクリプションのイメージ

4. 研究成果

(1) サラウンドマイクロホンアレーの最適配置問題の検討
マイクロホンアレーをサラウンド配置

した際の、音源位置推定精度を評価した。その結果、中央一極配置と比較して、音源位置推定精度の飛躍的な向上を確認した。単一のマイクロホンアレーでは、音源がマイクロホンアレーから離れるにつれて音源位置推定精度が劣化する傾向があるが、サウンドマイクロホンアレーでは、各マイクロホンアレーから等距離に位置する音源の位置推定精度が最も高いことが分かった。この傾向を基に(2)の音源発生検出も含めた研究へと拡張を行った。

(2) サウンドマイクロホンアレーによる音イベントの位置および発生検出

(1)の最適配置検討結果を基に、室内の音イベントの位置推定および発生時刻の検出を行った。特に発生時刻の高精度な検出を実現するため、VAD(Voice Activity Detection)の考え方を拡張することで、サウンドマイクロホンアレーを用いて様々な音源の発生時刻と終了時刻の検出を試みた。その結果、複合環境音については改善の余地はあるものの、その他の音源に関しては概ね高精度で音源の発生時刻および終了時刻を検出できることを確認した。そこで、この結果を基に(3)の音イベントの高音質受音も含めた研究へと拡張を行った。

(3) サウンドマイクロホンアレーによる音イベントの高音質受音

これまで、音イベントの高音質な受音を実現するためにはアレー信号処理と単一チャンネルによる信号処理の2つのアプローチがあったが、本研究ではこの2つを統合してさらなる高音質受音を試みた。特にアレー信号処理に関してはAMNOR(Adaptive Microphone array for Noise Reduction)を利用し、単一チャンネル処理ではSS(Spectral Subtraction)法を改良した、最適フロアリング係数を用いたIterative-SS法を提案し、サウンドマイクロホンアレーに統合することで性能改善を試みた。評価実験の結果、これまで従来法において問題とされていたミュージカルノイズを大幅に軽減することに成功し、音イベントの高音質な受音を実現した。そこでこの結果を基に(4)の音場トランスクリプションシステムの構築へと研究を拡張した。

(4) 上記に基づく音場トランスクリプションシステムの構築

サウンドマイクロホンアレーにより高音質に抽出した音イベントを文字としてディクテーションする(書き取る)ことで、音イベントの位置情報、発生時刻情報、

およびテキストによる音イベント内容をリアルタイムで記録可能な音場トランスクリプションシステムの構築を試みた。特に音イベントを文字として記録する手法については、従来の擬音語としての音源認識(例えばドアが閉まる音ならボタン)手法をさらに拡張し、音イベント内容にあわせて分類(例えばドアが閉まる音なら「ドアの開閉」)することで、アーカイブ価値の高い音場のトランスクリプション手法を検討した。その結果、リアルタイムで音場を構造化可能な音場トランスクリプションシステムの構築に成功した。今後はさらに音場の映像情報も忠実にテキストすることで、さらに実用性の高い音場トランスクリプション手法を模索する計画である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計4件)

福森隆寛, 森勢将雅, 西浦敬信, 山下洋一, "室内音響指標を用いた残響指標RSR-Dnに基づく残響下音声認識性能の予測," 電子情報通信学会論文誌(D), Vol. J94-D, No. 4, pp. 712-720, Apr. 2011.

(査読有)

Kook CHO, Hajime OKUMURA, Takanobu NISHIURA, and Yoichi YAMASHITA, "Multiple Sound Source Localization Based on Inter-Channel Correlation Using a Distributed Microphone System in a Real Environment," IEICE Trans. INF. & SYST., Vol. E93-D, No. 9, pp. 2463-2471, Sep. 2010. (査読有)

中山雅人, 西浦敬信, 山下洋一, "母音/子音特徴量に基づく適応形マイクロホンアレーを用いた雑音下音声認識," 電子情報通信学会論文誌(D), Vol. J92-D, No. 9, pp. 1568-1578, Sep. 2009. (査読有)

傳田 遊亀, 田中 貴雅, 溝口 遊, 中山 雅人, 西浦 敬信, 山下 洋一, "話者方位推定を利用した動的な時間領域処理に基づく遠隔発話区間検出," 電子情報通信学会論文誌(D), Vol. J92-D, No. 1, pp. 112-122, Jan. 2009. (査読有)

[学会発表](計14件)

堀井圭祐, 福森隆寛, 森勢将雅, 西浦敬信, 山下洋一, "スペクトル減算を用いた音質改善のための減算係数最適化の検討," 電子情報通信学会 音声研究会, Vol. 110, no. 401, SP2010-113, pp. 59-64, 2011年1月28日, 国際電気通信

基礎技術研究所(京都府) (査読無)

Kohei Hayashida, Masanori Morise, Takanobu Nishiura, "Near Field Sound Source Localization Based on Cross-power Spectrum Phase Analysis with Multiple Channel Microphones," Proc. INTERSPEECH 2010, pp.2758-2761, 2010年9月30日, 幕張メッセ(千葉県) (査読有)

Takahiro Fukumori, Masanori Morise, Takanobu Nishiura, "Performance Estimation of Reverberant Speech Recognition Based on Reverberant Criteria RSR-Dn with Acoustic Parameters," Proc. INTERSPEECH 2010, pp.562-565, 2010年9月27日, 幕張メッセ(千葉県) (査読有)

堀井圭祐, 福森隆寛, 森勢将雅, 西浦敬信, 南條浩輝, "スペクトル減算のための等ラウドネス曲線に基づく減算係数最適化の検討," 日本音響学会2010年秋季研究発表会, pp. 711-712, 2010年9月15日, 関西大学(大阪府) (査読無)

Kohei Hayashida, Yu Mizoguchi, Junpei Ogawa, Masanori Morise, Takanobu Nishiura, "The Acoustic Sound Field Dictation with Hidden Markov Model Based on an Onomatopoeia," ICA2010, PaperID:171, 2010年8月24日(オーストラリア シドニー) (査読有)

Takahiro Fukumori, Masanori Morise, Takanobu Nishiura, "Performance Estimation of Speech Recognition Based on Acoustic Parameters under Reverberation Environments with CENSREC-4," ICA2010, PaperID:166, 2010年8月24日(オーストラリア シドニー) (査読有)

林田亘平, 溝口 遊, 森勢将雅, 西浦敬信, "擬音語HMMに基づく音場ディクテーションの検討," 信学技報, S Vol.110, No. 56, SP2010-9, pp. 49-54, 2010年5月27日, 甲南大学(兵庫県) (査読無)

林田亘平, 森勢将雅, 西浦敬信, "近接音源位置推定のための2D-CSP法の評価," 信学技報, Vol.110, No. 56, SP2010-10, pp. 55-60, 2010年5月26日, 甲南大学(兵庫県) (査読無)

Kohei Hayashida, Masanori Morise, Takanobu Nishiura, "Near Field Sound Source Localization Based on Cross-power Spectrum Phase Analysis with Paired Microphones," NCSP2010, pp. 273-276, 2010年3月5日(米国 ハワイ) (査読有)

Yu Mizoguchi, Kohei Hayashida, Masanori Morise, Takanobu Nishiura,

Yoichi Yamashita, "Hand-free Voice Activity Detection Based on Direction of Arrival Estimation and Sound Source Identification," WESPAC2009, PaperID:123, 2009年9月21日(中国 北京) (査読有)

Takahiro Fukumori, Yoshiki Hirano, Masanori Morise, Takanobu Nishiura, "Performance Estimation of Speech Recognition Based on Acoustic Parameters under Reverberation Environments," WESPAC2009, PaperID:128, 2009年9月21日(中国 北京) (査読有)

溝口 遊, 川野 弘, 西浦 敬信, 山下 洋一, "音源方位推定と音源識別に基づくハンズフリー発話区間検出法の検討," 日本音響学会 2009年春季研究発表会, pp.121-122, 2009年3月17日, 東京工業大学(東京都) (査読無)

Cho Kook, Hajime Okumura, Takanobu Nishiura, Yoichi Yamashita, "Localization of Multiple Sound Sources Based on Inter-Channel Correlation Using a Distributed Microphone System," INTERSPEECH2008, pp.443-446, 2008年9月24日(オーストラリア ブリスベン) (査読有)

西浦敬信, 傳田遊亀, 中山雅人, "マイクロホンアレーを用いた遠隔発話音声処理", 第52回システム制御情報学会研究発表講演会, pp.353-354, 2008年5月17日, 京都情報大学院大学(京都府) (査読無)(招待講演)

〔図書〕(計0件)

〔産業財産権〕
出願状況(計0件)

取得状況(計0件)

〔その他〕
ホームページ等
<http://www.aspl.is.ritsumei.ac.jp/>

6. 研究組織

(1) 研究代表者

西浦 敬信 (NISHIURA TAKANOBU)
立命館大学・情報理工学部・准教授
研究者番号: 70343275