

2入力2出力系の逆同定への入力信号特性の影響

Influence of Input Signal Characteristics on the Inverse Identification of a Two-input Two-output System

○加茂 毅一郎 (岩手大・院), 西山 清 (岩手大), 佐藤克昌(ARI)

○Kiichiro Kamo, Kiyoshi Nishiyama, Katsuaki Satoh

岩手大学, Iwate University, (株)ARI, Amenity Research Institute

キーワード: 逆システム同定 (inverse system identification), トランスオーラルシステム (transaural system),
バイノーラル録音 (binaural recording)

連絡先: 〒020-8551 盛岡市上田4-3-5 岩手大学 工学部 電気電子・情報システム工学科 西山研究室
西山清, Tel.: (019)621-6475, Fax.: (019)621-6475, E-mail: nishiyama@cis.iwate-u.ac.jp

1. はじめに

聴覚的な臨場感の再現、すなわち立体音響に関する試みとして、古くから様々な研究が成されており、再生方式としてもモノラル、ステレオ、サラウンドと進化してきた。特に、サラウンドは映画館などで臨場感を得るために古くから使われているほか、家庭用でも比較的安価に5.1ch、7.1chサラウンドが実現されている。このように、擬似的な音場を再現する技術は広く用いられているが、一方でコンサートや野外録音などの臨場感を録音、再現したいというニーズもある。それに応えることのできる技術としてバイノーラル録音¹⁾がある。人工的に臨場感を再現する方法としては、距離による音量の減衰や両耳間強度差により音像定位を再現する音量差、音波が到達する時間の差によって音像定位を再現する時間差などの方法に加え、周波数特性、位相、残響の変化によって再現する方法があるが、バイノーラル録音は実際に人間の耳で聞くように録音することで実際の臨場感も含

め録音する。しかし、聴取する際にはイヤフォンやヘッドフォンを用いる必要がある。このバイノーラル音源再生を2チャンネルのスピーカを用いて実現するための手法としてトランスオーラル再生がある。トランスオーラル再生を実現するためには、伝達系のクロストーク成分をキャンセルするための補正フィルタとして未知の伝達系の逆システムが必要になる。

本研究では、4経路同時逆同定法で求めた2入力2出力系の逆フィルタで構成されたトランスオーラルシステムが、左右で異なったステレオ音声をどの程度分離再生できるかシミュレーションを行う。逆同定には、高速 H_∞ フィルタとNLMSアルゴリズムを用いた適応フィルタを使用して、左右無相関の白色雑音、2次、16次のAR信号および実際の音声信号を未知系の入力信号としたときの同定精度をそれぞれ比較することで、ロバスト性や収束性が高いことで知られる高速 H_∞ フィルタ^{2),3)}の有効性を示す。

2. バイノーラル録音とトランスオーラル再生

2.1 バイノーラル録音

実際の人間の耳にマイクを装着、または人間の頭部形状を模したダミーヘッド(Fig.1)と呼ばれる人形の両耳部分にマイクを埋め込んで録音する手法をバイノーラル録音という。この音源をヘッドホンやイヤホンにより聴取した場合、録音時の情報がそのまま再現されるため高臨場感が得られる。

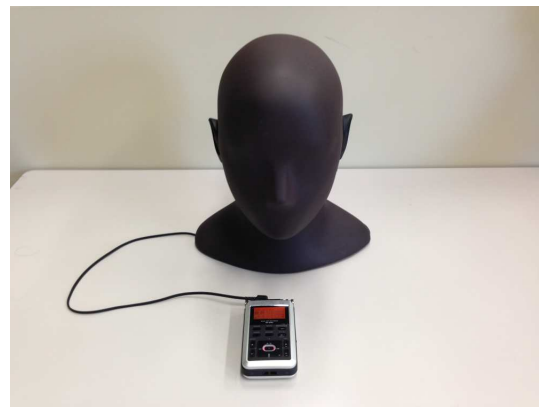


Fig. 1 ダミーヘッド

2.2 トランスオーラル再生

バイノーラル音源をステレオ再生すると左右の音が混ざり合ってしまう臨場感は損なわれる。2チャンネルのステレオスピーカを用いて高臨場感再生を可能にする手法がトランスオーラル再生¹⁾である。トランスオーラル再生を実現するためには、スピーカから耳元へ至るクロストーク成分をキャンセルするために2入力2出力系(H)の逆システム(G)が必要となる⁴⁾。この逆システムがFig.2の補正フィルタ G となる。トランスオーラルシステムの伝達関数行列は次の関係式を満たす(複素変数 z は省略)。

$$\begin{bmatrix} H_{LL} & H_{LR} \\ H_{RL} & H_{RR} \end{bmatrix} \begin{bmatrix} G_{LL} & G_{LR} \\ G_{RL} & G_{RR} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (1)$$

$$H_{LL}G_{LL} + H_{LR}G_{RL} = 1, H_{RL}G_{LR} + H_{RR}G_{RR} = 1 \quad (2)$$

$$H_{LL}G_{LR} + H_{LR}G_{RR} = 0, H_{RL}G_{LL} + H_{RR}G_{RL} = 0 \quad (3)$$

式(2)は音源の通過領域であり、式(3)はクロストークのキャンセルを表す。上式を満たす逆フィルタ G を実現できれば、耳元でバイノーラル音源が再現される。つまり、録音時の信号を耳元で再現できるので3次元の高臨場感が実現される。トランスオーラルシステムによる3次元高臨場感再生において、未知の2入力2出力系の逆システムを直接同定することが重要であり、その精度が聴取時に得られる臨場感に大きく影響を与える。

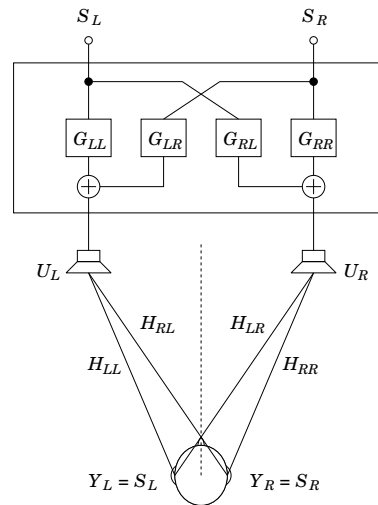


Fig. 2 トランスオーラルシステム

3. 4経路同時逆同定法に基づくトランスオーラル再生

本研究では、Fig.3に示すように未知の2入力2出力系の逆システムの4経路を同時に同定し、トランスオーラルシステムを構築する^{5),6)}。

未知系の4つの経路の伝達関数をそれぞれ H_{LL} , H_{LR} , H_{RL} , H_{RR} とし、スピーカの出力を U_L, U_R とする。このとき聴取者の両耳に届く信号 Y_L, Y_R は

$$\begin{bmatrix} Y_L \\ Y_R \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{LR} \\ H_{RL} & H_{RR} \end{bmatrix} \begin{bmatrix} U_L \\ U_R \end{bmatrix} \quad (4)$$

のように表すことができる。ここで観測信号 Y_L, Y_R を逆フィルタの入力信号とすると出力信号は U_L, U_R

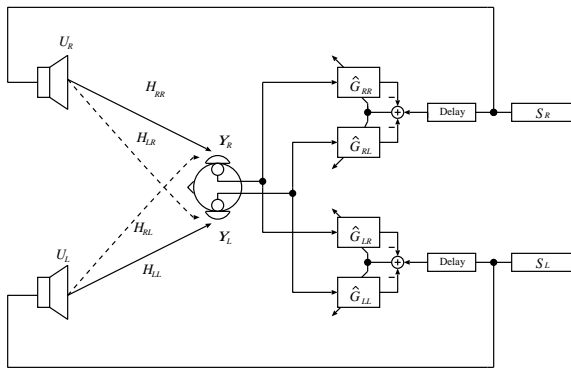


Fig. 3 4経路同時逆システム同定方式

と一致することが望まれる。すなわち、

$$\begin{bmatrix} U_L \\ U_R \end{bmatrix} = \begin{bmatrix} \hat{G}_{LL} & \hat{G}_{LR} \\ \hat{G}_{RL} & \hat{G}_{RR} \end{bmatrix} \begin{bmatrix} Y_L \\ Y_R \end{bmatrix} \quad (5)$$

しかし、逆システムを効果的に同定する際には、

$$\begin{bmatrix} \hat{G}_{LL} & \hat{G}_{LR} \\ \hat{G}_{RL} & \hat{G}_{RR} \end{bmatrix} \begin{bmatrix} Y_L \\ Y_R \end{bmatrix} \rightarrow z^{-D} \begin{bmatrix} U_L \\ U_R \end{bmatrix}$$

に収束するよう、信号 S_L, S_R に遅延 D を与えたものとの評価誤差が0となるよう適応的にフィルタを制御する。このとき、

$$\begin{aligned} & \begin{bmatrix} \hat{G}_{LL} & \hat{G}_{LR} \\ \hat{G}_{RL} & \hat{G}_{RR} \end{bmatrix} \begin{bmatrix} H_{LL} & H_{LR} \\ H_{RL} & H_{RR} \end{bmatrix} \\ &= \begin{bmatrix} \hat{G}_{LL}H_{LL} + \hat{G}_{LR}H_{RL} & \hat{G}_{LL}H_{LR} + \hat{G}_{LR}H_{RR} \\ \hat{G}_{RL}H_{LL} + \hat{G}_{RR}H_{RL} & \hat{G}_{RL}H_{LR} + \hat{G}_{RR}H_{RR} \end{bmatrix} \\ &= z^{-D} \mathbf{I} \end{aligned} \quad (6)$$

となれば、クロストークをキャンセルすることができる。

4. 4経路同時逆同定法の解析

3章の2入力2出力系の逆同定法に基づくトランスオーラルシステムをシミュレーションで評価を行った。Fig.3の方式に基づき、各種信号を用いて、NLMSアルゴリズムと高速 H_∞ アルゴリズムにより未知系を逆同定し、得られたそれぞれの逆フィルタを実際の音声信号に使用してトランスオーラル再生をシミュレーションし、比較・検討した。

逆同定で用いる未知系の入力信号は、白色雑音、2次もしくは16次のAR信号および実際の音声信号

の、左右無相関な4種類のステレオ信号を用いた。対象とする未知系の各経路としては無響室で測定したタップ数342の4つのインパルス応答を用いた。サンプリング周波数は16 kHzとし、未知系の入力信号としては、白色雑音、2次、16次のAR信号と実際の音声信号を使用した。N次のAR信号は次のモデルで生成された。

$$\begin{aligned} u(k) &= \alpha_1 u(k-1) + \alpha_2 u(k-2) \\ &+ \dots + \alpha_N u(k-N) + w_k \end{aligned} \quad (7)$$

ここで、2次のARモデルの場合は $\alpha_1 = 0.7, \alpha_2 = 0.1$ とし、16次のARモデルは母音を近似した。 w_k は平均値0、分散 $\sigma_{w_k}^2 = 0.04$ の定常なガウス白色雑音とした。また、入力信号に白色雑音を用いる場合は上記 w_k をそのまま入力信号とした。

$$u(k) = w_k \quad (8)$$

異なる乱数の種を使用した2種類の無相関な白色雑音を用いて左右で異なる入力信号 u_L, u_R を生成した。音声信号は、Fig.4.のステレオ音声(相関係数0.0190)を用いた。入力信号のデータ長は $L = 16000$

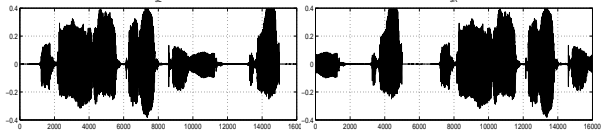
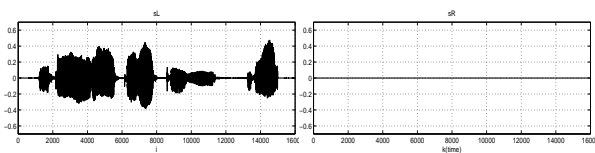


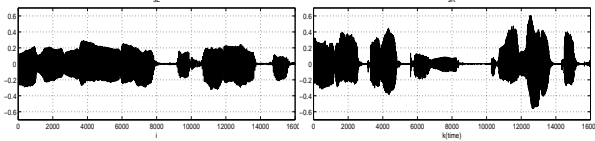
Fig. 4 同定に使用したステレオ音声信号(左右)

とし、逆システムの各経路のインパルス応答はタップ数 $N_{iv} = 512$ 、遅延時間 $D = 246$ と設定した。観測雑音は平均0、分散 $\sigma_v^2 = 1.0 \times 10^{-4}$ の白色雑音を用いた。入力信号 u_L, u_R として、白色雑音、2次、16次のAR信号および実際の音声信号の4つをそれぞれ用いた。音声信号として、左(S_L)がアナウンス音声、右(S_R)が無音のステレオ音声(Fig.5)と、左右(S_L, S_R)がそれぞれ別のアナウンス音声を用いたステレオ音声(Fig.6)を使用した。



(a)元音源の音声信号 アナウンス: s_L (左) (b)元音源の音声信号 無音: s_R (右)

Fig. 5 元音源の音声信号 s_L, s_R :片方が無音のステレオ音声;データ長 $L = 16000$



(a)元音源の音声信号 アナウンス: s_L (左) (a)元音源の音声信号 アナウンス: s_R (右)

Fig. 6 元音源の音声信号:左右がそれぞれ別のアナウンス音声 s_L, s_R ;データ長 $L = 16000$

シミュレーションでは、まず元音源 s_L, s_R と逆フィルタ $\hat{g}_{LL}, \hat{g}_{LR}, \hat{g}_{RL}, \hat{g}_{RR}$ との畳み込みからスピーカ出力 u_L, u_R を得た($\hat{g}_{lm} * s_n = \mathcal{Z}^{-1}\{\hat{G}_{lm}S_n\}$).

$$u_L(k) = \sum_{i=0}^{511} \hat{g}_{LL}(i)s_L(k-i) + \sum_{i=0}^{511} \hat{g}_{LR}(i)s_R(k-i) \quad (9)$$

$$u_R(k) = \sum_{i=0}^{511} \hat{g}_{RR}(i)s_R(k-i) + \sum_{i=0}^{511} \hat{g}_{RL}(i)s_L(k-i) \quad (10)$$

この u_L, u_R とスピーカとマイクとの音響系との畳み込みから、耳元への出力 y_L, y_R を得ることでトランスオーラル再生し、その結果を比較した。

$$y_L(k) = \sum_{i=0}^{341} h_{LL}(i)u_L(k-i) + \sum_{i=0}^{341} h_{LR}(i)u_R(k-i) \quad (11)$$

$$y_R(k) = \sum_{i=0}^{341} h_{RR}(i)u_R(k-i) + \sum_{i=0}^{341} h_{RL}(i)u_L(k-i) \quad (12)$$

元音源の音声信号と、耳元への出力信号が遅延を除いて等しいとき($y_L(k) = s_L(k-D), y_R(k) = s_R(k-D)$), 左右音声の分離が実現できているといえる。

Fig.8,10はNLMSアルゴリズム、Fig.9,11は高速 H_∞ フィルタに基づく4経路同時逆同定法を用いたトランスオーラルシステムにおける耳元への出力 y_L, y_R を示している。未知系の入力信号として白色雑音を使用した使用した場合、NLMSアルゴリ

ズムを用いた適応フィルタ、高速 H_∞ フィルタ共に良好なクロストークキャンセルが得られている。しかし、AR信号、実際の音声信号を用いて逆同定した場合、高速 H_∞ フィルタで求めた逆フィルタによる y_L, y_R は、わずかに乱れが見られるものの、左右の信号分離は高精度で行われている。一方、NLMSアルゴリズムによるものは大きな乱れが見受けられ、クロストークキャンセルができていないことがわかる。これより、高速 H_∞ フィルタによる逆フィルタがより良好であることがわかる。

5. 3次元高臨場感再生装置の試作



Fig. 7 H_∞ トランスオーラル再生を用いた3次元高臨場感再生装置

Fig.7は4経路同時逆同定法に基づくトランスオーラル再生を用いた3次元高臨場感再生装置である。PC,内蔵アンプ、スピーカー、バイノーラルマイク、ディスプレイによって構築されており、PCにインストールされている独自開発された「3D立体音響ソフト」によって、3次元立体音響を誰でも手軽に体験できる。

使用にあたっては、まず、バイノーラルマイクを両耳に装着し、白色雑音を数秒聴けば、その直後からバイノーラル音源の再生が可能となり、高臨場感が楽しめる。

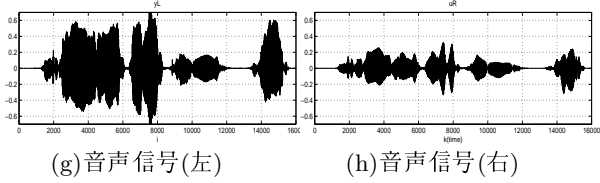
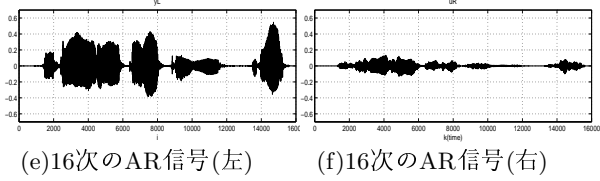
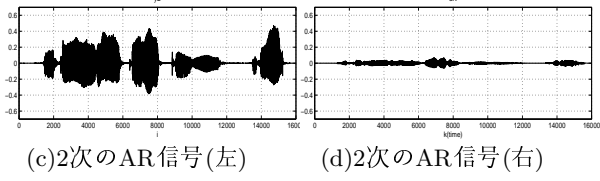
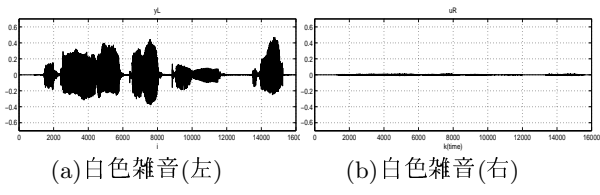


Fig. 8 NLMSアルゴリズム($\mu = 0.5$)による逆フィルタを用いたトランスオーラルシステムにおける耳元への出力信号 y_L, y_R (左:男声、右:無音)

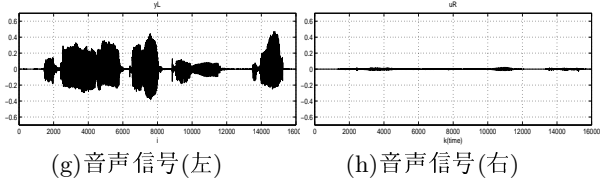
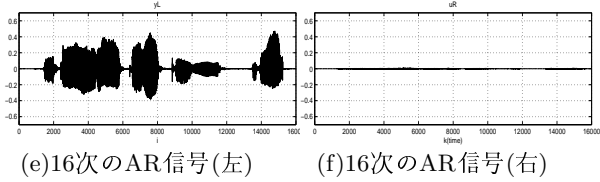
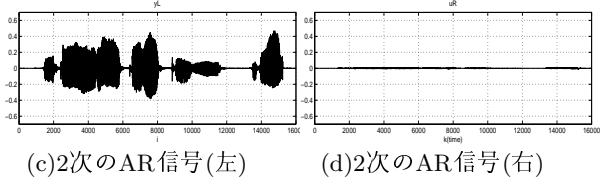
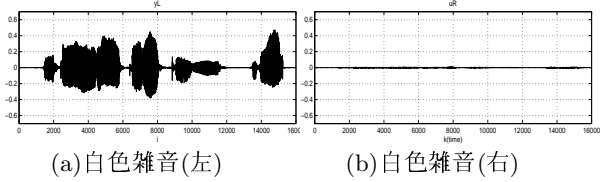


Fig. 9 高速 H_∞ フィルタ($\gamma_f = 32, \epsilon = 1.0$)による逆フィルタを用いたトランスオーラルシステムにおける耳元への出力信号 y_L, y_R (左:男声、右:無音)

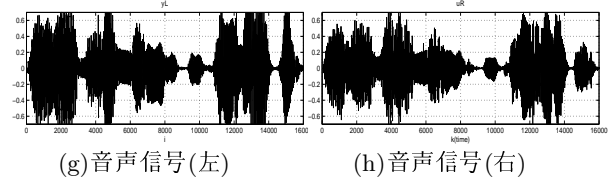
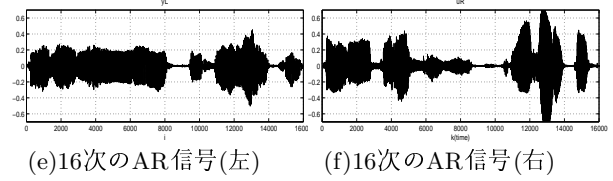
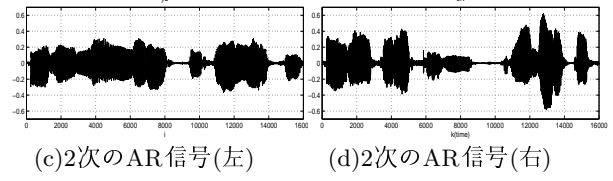
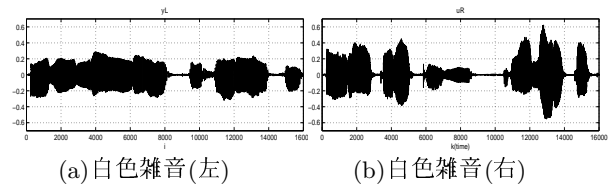


Fig. 10 NLMSアルゴリズム($\mu = 0.5$)による逆フィルタを用いたトランスオーラルシステムにおける耳元への出力信号 y_L, y_R (左:女声、右:男声)

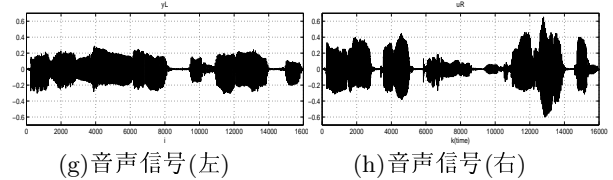
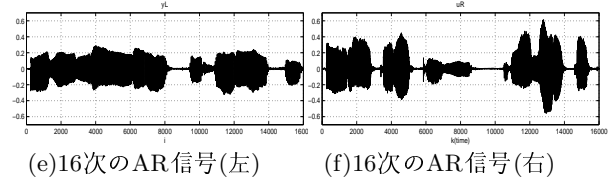
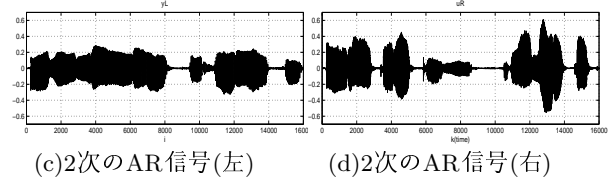
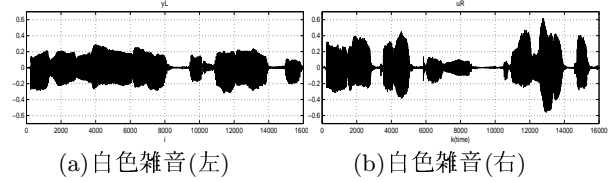


Fig. 11 高速 H_∞ フィルタ($\gamma_f = 32, \epsilon = 1.0$)による逆フィルタを用いたトランスオーラルシステムにおける耳元への出力信号 y_L, y_R (左:女声、右:男声)

6. まとめ

本研究では、4経路同時逆同定法を用いて構成されたトランスオーラルシステムによる音源の分離性能を比較・検討した。逆同定の際、高速 H_∞ フィルタを用いることで優れた効果が得られ、未知系の入力信号が白色雑音以外のAR信号や音声信号でも高い臨場感が得られることを示した。また、独自に開発した3次元高臨場感再生装置では数秒間で未知の2入力2出力系の逆同定が可能となり、手軽に高臨場感が体感できた。今後は、未知系への左右の入力信号の相関性が逆同定へどのように影響していくか、さらに考察を進めていきたい。

参考文献

- 1) 浜田 晴夫, “バイノーラル音場再生系について”, 日本音響学会誌, 48巻4号, pp. 250-257, 1992.
- 2) 西山 清, 最適フィルタリング, 培風館, 2001.
- 3) K. Nishiyama, “An Optimization and Its Fast Algorithm for Time-Variant System Identification”, IEEE Trans. SP, 52, 5, pp. 1335-1342, 2004.
- 4) 小松 洋貴, 西山 清, “トランスオーラルシステムのための逆システムの同定”, 電子情報通信学会ソサイエティ大会講演論文集, p. 70, 2012.
- 5) 新井 清嗣, 勝俣友紀, 佐藤克昌, 西山 清, “ H_∞ 高速同定を用いた3次元高臨場感音場再生システムの開発 -4経路同時同定-”, 日本音響学会秋季研究発表会講演論文集, 3-1-1, 2013.
- 6) 加茂毅一郎, 西山 清, 佐藤克昌, “3次元高臨場感音場再生システムの試作”, 計測自動制御学会50周年記念学術講演論文集, A305, 2014.