(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5178370号
(P5178370)

(45)発行日　平成25年4月10日(2013.4.10)　(24)登録日　平成25年1月18日(2013.1.18)

(51)Int.Cl.　F I
G10L 21/0308 (2013.01)　G10L 21/02 203Z
G10L 21/028 (2013.01)　G10L 21/02 201D
G10L 21/0272 (2013.01)　G10L 21/02 202B

請求項の数 2　(全 12 頁)

(21)出願番号　特願2008-191382 (P2008-191382)
(22)出願日　平成20年7月24日 (2008.7.24)
(65)公開番号　特開2009-42754 (P2009-42754A)
(43)公開日　平成21年2月26日 (2009.2.26)
審査請求日　平成22年11月26日 (2010.11.26)
(31)優先権主張番号　60/954889
(32)優先日　平成19年8月9日 (2007.8.9)
(33)優先権主張国　米国 (US)

(73)特許権者　000005326
本田技研工業株式会社
東京都港区南青山二丁目１番１号
(74)代理人　110000800
特許業務法人創成国際特許事務所
(72)発明者　武田　龍
埼玉県和光市本町８－１　株式会社ホンダ・リサーチ・インスティチュート・ジャパン内
(72)発明者　中臺　一博
埼玉県和光市本町８－１　株式会社ホンダ・リサーチ・インスティチュート・ジャパン内

最終頁に続く

(54)【発明の名称】音源分離システム

(57)【特許請求の範囲】
【請求項１】
スピーカから環境に対して音響として出力される既知信号を保存する既知信号記憶手段と、
マイクロホンと、
前記マイクロホンからの出力信号を周波数変換することにより現在のフレームの観測信号を生成する第１処理部と、
周波数領域における前記既知信号の周波数成分と前記スピーカから前記マイクロホンまでの周波数領域の音響の伝達関数との畳み込みとして現在フレームの元信号が表現されている第１モデルと、前記元信号および未知信号を包含するように前記観測信号が表現されている第２モデルとにしたがって、前記第１処理部により生成された現在フレームの前記観測信号を、適応フィルタを経由した前記元信号と前記未知信号とに独立成分分析法により分離することにより前記未知信号を抽出する第２処理部とを備えていることを特徴とする音源分離システム。
【請求項２】
請求項１記載の音源分離システムにおいて、
前記適応フィルタｈ^の更新式は次の（ａ）式とし、
$h\hat{}(\omega,f+1)=h\hat{}(\omega,f)-\mu_1\varphi(E(\omega,f))X^*(\omega,f)$・・・（ａ）
とし、
ただし、前記（ａ）式において、

ωは周波数を意味し、
ｆ＋１，ｆは更新後および更新前のフレームを意味し、
μ_1は所定の係数であり、
Ｅ(ω,ｆ)は前記未知信号であり、
関数φは確率変数ｅの密度関数p_x（ｘ）により次の（ｂ）式により定義されるものであり、
$\phi(x)=-(d/dx)\log p_x(x)$・・・（ｂ）
$X^*(\omega,f)$は前記元信号Ｘ(ω,ｆ)の複素共役であることを特徴とする音源分離システム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、音源分離システムに関する。

【背景技術】

【０００２】

ユーザとロボットとの自然な対話を実現する上で、ロボットの発話中にユーザの発話（いわゆるバージイン）を許容することは不可欠である。ロボットにマイクロホンが搭載されている場合、ロボット自身の発話がマイクロホンに入り込むので、バージインは相手の発話を認識する上で大きな障害となる。

【０００３】

そこで、図４に示されている構成の適応フィルタが用いられている。自発話の除去は、スピーカＳからマイクロホンＭへの伝達系ｈを近似するフィルタｈ^の推定問題として取り扱われる。マイクロホンＭから入力された観測信号ｙ(ｋ)から推定信号ｙ^(ｋ)が差し引かれることにより相手発話が取り出される。

【０００４】

適応フィルタの１つとしてＮＬＭＳ（Normalized Least Mean Squares）法が提案されている。ＮＬＭＳ法によれば、時間領域において、線形かつ時間不変な伝達系を経て観測される信号ｙ(ｋ)が、元信号ベクトル$x(k)={}^t(x(k),x(k-1),..x(k-N+1))$（「Ｎ」はフィルタ長を表わす。「ｔ」は転置を意味する。）と、伝達系のインパルス応答$h={}^t(h_1,h_2,..h_N)$との畳み込みを用いて関係式（１）により表現される。

【０００５】

$y(k)={}^t x(k)h$ ..(1)

【０００６】

推定フィルタ$h^=({}^t(h_1^,h_2^,..h_N^)$は、関係式（２）により表わされる観測信号と推定信号との誤差ｅ(ｋ)の二乗平均を最小化することにより得られる。推定フィルタｈ^を求めるためのオンラインアルゴリズムは、正則化のための小さな正数値を用いて関係式（３）により表現される。なお、関係式（３）において$||x(k)||^2+\delta$により学習係数が正規化されない場合がＬＭＳ法である。

【０００７】

$e(k)=y(k)-{}^t x(k)h^$ ..(2)

【０００８】

$h^(k)=h^(k-1)+\mu_{NLMS}x(k)e(k)/(||x(k)||^2+\delta)$ ..(3)

【０００９】

また、ＩＣＡ（独立成分分析）法が提案されている。ＩＣＡ法はノイズを仮定して設計されているため、自発話区間の検出が不要であり、かつ、ノイズが存在しても分離可能であり、その結果としてバージイン問題の解決に適している。たとえば、時間領域ＩＣＡ法が提案されている（非特許文献１参照）。音源の混合過程はノイズｎ(ｋ)およびＮ＋１次の行列Ａを用いて関係式（４）により表わされる。

【００１０】

$({}^t(y(k),{}^t x(k))=A\ {}^t(n(k),{}^t x(k))$,

$A_{ii}=1(i=1..N+1), A_{1j}=h_{j-1}(j=2..N+1),$
$A_{ik}=0(k\neq i)$ ..(4)

【００１１】

ＩＣＡによれば関係式（５）における分離行列Ｗが推定される。

【００１２】

$^{t}(e(k),^{t}x(k))=W\ ^{t}(y(k),^{t}x(k)),$
$W_{11}=a,\ W_{ii}=1(i=2..N+1),$
$W_{1j}=h_{j}(j=2..N;1), W_{ik}=0(k\neq i)$ ..(5)

【００１３】

分離行列Ｗの第１行第１列成分$W_{11}=a=1$の場合が従来の適応フィルタのモデルであり、ＩＣＡ法と最も異なっている点である。自然勾配法にしたがってＫＬ情報量が最小化されることにより、オンラインアルゴリズムを表わす関係式（６）および（７）にしたがって最適な分離フィルタが求められる。

【００１４】

$$\hat{h}(k+1)=\hat{h}(k)+\mu_{1}[\{1-\phi(e(k))e(k)\}\hat{h}(k)-\phi(e(k))x(k)] \quad ..(6)$$

【００１５】

$$a(k+1)=a(k)+\mu_{2}[1-\phi(e(k))e(k)]a(k) \quad ..(7)$$

【００１６】

関数φは確率変数ｅの密度関数$p_{x}(x)$により関係式（８）により定義される。

【００１７】

$$\phi(x)=-(d/dx)\log p_{x}(x) \quad ..(8)$$

【００１８】

さらに、周波数領域ＩＣＡ法が提案されている（非特許文献２参照）。一般的に周波数領域では畳み込み混合が瞬時の混合とみなせるため、時間領域ＩＣＡ法よりも収束性に優れている。この手法によれば、窓長Ｔおよびシフト長Ｕによる短時間フーリエ解析が実行されることにより、時間周波数領域での信号が得られる。元信号ｘ(t)および観測信号ｙ(ｔ)のそれぞれはフレームｆおよび周波数ωを変数とする関数Ｘ(ω,ｆ)およびＹ(ω,ｆ)のそれぞれにより表現される。観測信号ベクトルＹ(ω,ｆ)$=^{t}$(Ｙ(ω,ｆ),Ｘ(ω,ｆ))の分離過程は、推定された元信号ベクトル$\hat{Y}(\omega,f)=^{t}(E(\omega,f),X(\omega,f))$を用いて関係式（９）により表現される。

【００１９】

$$\hat{Y}(\omega,f)=W(\omega)Y(\omega,f),\ W_{21}(\omega)=0, W_{22}(\omega)=1 \quad ..(9)$$

【００２０】

分離行列の学習は周波数ごとに独立に行われる。学習は非ホロノミック拘束適用によるＫＬ情報量最小化に基づく関係式（１０）により表わされる反復学習則にしたがって行われる（非特許文献３参照）。

【００２１】

$$W^{(j+1)}(\omega)=W^{(j)}(\omega)-\alpha\{\text{off-diag}<\phi(\hat{Y})\hat{Y}^{H}>\}W^{(j)}(\omega) \quad ..(10)$$

【００２２】

αは学習係数であり、（ｊ）は更新回数であり、＜・＞は平均値であり、ｏｆｆ－ｄｉａｇＸは行列Ｘの対角要素を０に置換する演算を表わし、非線形関数φ（ｙ）は関係式（１１）により定義されている。

【００２３】

$$\phi(y_{i})=\tanh(|y_{i}|)\exp(i\theta(y_{i})) \quad ..(11)$$

【００２４】

また、既知音源から既知音源への伝達特性は定数で表わされるため、分離行列Ｗの第１行成分のみが更新される。

【非特許文献１】J.Yang et al., A New Adaptive Filter Algorithm for System Identification Using Independent Component Analysis, Proc. ICASSP2007, pp.1341-1344,20

07
【非特許文献２】S.Myabe et al., Double-Talk Free Spoken Dialogue Interface Combining Sound Field Control with Semi-Blind Source Separation, Proc. ICASSP2006, pp.809-812,2006
【非特許文献３】Sawada et al., Polar Coordinate based Nonlinear Function for Frequency-Domain Blind Source Separation, IEICE Trans., Fundamentals, 3, E-86A, pp. 505-510, 2003
【発明の開示】
【発明が解決しようとする課題】
【００２５】
しかし、従来の周波数領域ＩＣＡ法には次のような問題点があった。第１の問題は残響に対応するために窓長Ｔを長く取る必要があり、その分だけ演算処理遅延および音源分離性能の低下を招くという点である。第２の問題点は窓長Ｔを環境に応じて変更する必要があり、他の雑音抑圧手法などの接続が煩雑になる点である。
【００２６】
そこで、本発明は、残響または反響の影響を軽減することにより音源分離精度の向上を図ることができるシステムを提供することを解決課題とする。
【課題を解決するための手段】
【００２７】
第１発明の音源分離システムは、スピーカから環境に対して音響として出力される既知信号を保存する既知信号記憶手段と、マイクロホンと、前記マイクロホンからの出力信号を周波数変換することにより現在のフレームの観測信号を生成する第１処理部と、周波数領域における前記既知信号の周波数成分と前記スピーカから前記マイクロホンまでの周波数領域の音響の伝達関数との畳み込みとして現在フレームの元信号が表現されている第１モデルと、前記元信号および未知信号を包含するように前記観測信号が表現されている第２モデルとにしたがって、前記第１処理部により生成された現在フレームの前記観測信号を、適応フィルタを経由した前記元信号と前記未知信号とに独立成分分析法により分離することにより前記未知信号を抽出する第２処理部とを備えていることを特徴とする。
【００２８】
第１発明の音源分離システムによれば、第１モデルおよび第２モデルにしたがって観測信号から未知信号が抽出される。特に、第１モデルによれば、現在フレームの元信号が現在および過去フレームの既知信号の合成信号として表現されている。このため、窓長を変更させることなく、既知信号の残響または反響が観測信号に及ぼす影響を軽減しながら未知信号が抽出されうる。したがって、残響の影響を軽減するための演算処理負荷を軽減しながら、未知信号に基づく音源分離精度の向上を図ることができる。
【００２９】
第１発明の音源分離システムによれば、現在フレームの元信号が周波数領域における既知信号の周波数成分およびその伝達関数の畳み込みにより表現されている。このため、窓長を変更させることなく、元信号の残響または反響が観測信号に及ぼす影響を軽減しながら未知信号が抽出されうる。したがって、残響の影響を軽減するための演算処理負荷を軽減しながら、未知信号に基づく音源分離精度の向上を図ることができる。
【００３０】
第１発明の音源分離システムによれば、第２モデルにおいて適応的に分離フィルタが設定されるので、窓長を変更させることなく、元信号の残響または反響が観測信号に及ぼす影響を軽減しながら未知信号が抽出されうる。したがって、残響の影響を軽減するための演算処理負荷を軽減しながら、未知信号に基づく音源分離精度の向上を図ることができる。
【００３１】
第２発明の音源分離システムは、第１発明の音源分離システムにおいて、前記適応フィルタｈ^の更新式は次の（ａ）式とし、

$h\hat{}(\omega,f+1)=h\hat{}(\omega,f)-\mu_1\varphi(E(\omega,f))X^*(\omega,f)$・・・（ａ）

とし、ただし、前記（ａ）式において、ωは周波数を意味し、ｆ＋１，ｆは更新後および更新前のフレームを意味し、μ_1は所定の係数であり、Ｅ(ω,ｆ)は前記未知信号であり、関数φは確率変数ｅの密度関数p_x（ｘ）により次の（ｂ）式により定義されるものであり、

$\varphi(x)=-(d/dx)\log p_x(x)$・・・（ｂ）

$X^*(\omega,f)$は前記元信号Ｘ(ω,ｆ)の複素共役であることを特徴とする。

【発明を実施するための最良の形態】

【００３３】

本発明の音源分離システムの実施形態について図面を用いて説明する。

【００３４】

図１に示されている音源分離システムはマイクロホンＭと、スピーカＳと、電子制御ユニット（ＣＰＵ，ＲＯＭ，ＲＡＭ／Ｏ回路、Ａ／Ｄ変換回路等の電子回路などにより構成されている。）１０とにより構成されている。電子制御ユニット１０は第１処理部１１と、第２処理部１２と、第１モデル格納部１０１と、第２モデル格納部１０２と、自発話格納部１０４とを備えている。各処理部はたとえば演算処理回路、または、メモリと、メモリからプログラムを読み出してそのプログラムにしたがって担当する演算処理を実行する演算処理装置（ＣＰＵ）とにより構成されている。

【００３５】

第１処理部１１はマイクロホンＭからの出力信号を周波数変換することにより現在のフレームｆの観測信号（周波数ω成分）Ｙ(ω,ｆ)を生成する。第２処理部１２は第１処理部１１により生成された現在フレームの観測信号Ｙ(ω,ｆ)に基づき、第１モデル格納部１０１に格納されている第１モデルと、第２モデル格納部１０２に格納されている第２モデルとにしたがって未知信号Ｅ(ω,ｆ)を抽出する。電子制御ユニット１０は自発話格納部（既知信号記憶手段）１０４に格納されている既知信号をスピーカＳから音声または音響として出力させる。

【００３６】

マイクロホンＭはたとえば図２に示されているように電子制御ユニット１０が搭載されているロボットＲの頭部Ｐ１に配置されている。なお、音源分離システムはロボットＲのほか、車両（四輪自動車）、複数の音源が存在する環境に接する任意の機械や装置に搭載されうる。また、マイクロホンＭの数および配置は任意に変更されうる。ロボットＲは脚式移動ロボットであり、人間と同様に基体Ｐ０と、基体Ｐ０の上方に配置された頭部Ｐ１と、基体Ｐ０の上部に上部両側から延設された左右の腕体Ｐ２と、左右の腕体Ｐ２のそれぞれの先端に連結されている手部Ｐ３と、基体Ｐ０の下部から下方に延設された左右の脚体Ｐ４と、左右の脚体Ｐ４のそれぞれに連結されている足部Ｐ５とを備えている。基体Ｐ０はヨー軸回りに相対的に回動しうるように上下に連結された上部および下部により構成されている。頭部Ｐ１は基体Ｐ０に対してヨー軸回りに回動する等、動くことができる。腕体Ｐ２は肩関節機構、肘関節機構および手根関節機構のそれぞれにおいて１～３軸回りの回動自由度を有している、手部Ｐ３は、手掌部から延設され、人間の手の親指、人差指、中指、薬指および小指のそれぞれに相当する５つの指機構を備え、物体の把持動作等が可能に構成されている。脚体Ｐ４は股関節機構、膝関節機構および足関節機構のそれぞれにおいて１～３軸回りの回動自由度を有している。ロボットＲは音源分離システムによる音源分離結果に基づき、左右の脚体Ｐ４を動かして移動する等、適当な動作をすることができる。

【００３７】

前記構成の音源分離システムの機能について説明する。まず第１処理部１１によりマイクロホンＭからの出力信号が取得される（図３／Ｓ００２）。また、第１処理部１１によりこの出力信号がＡ／Ｄ変換された上で周波数変換されることにより、フレームｆの観測信号Ｙ(ω,ｆ)が生成される（図３／Ｓ００４）。

【００３８】

続いて第２処理部１２により、第１モデルおよび第２モデルにしたがって、第１処理部１１により生成された観測信号Ｙ(ω,ｆ)から元信号Ｘ(ω,ｆ)が分離されることにより、未知信号Ｅ(ω, ｆ)が抽出される（図３／Ｓ００６）。

【００３９】

第１モデルによれば、現在および過去の所定数Ｍのフレームにわたる元信号を包含するように現在フレームｆの元信号Ｘ(ω,ｆ)が表現されている。第１モデルによれば、次フレームに入り込んだ反響音が時間周波数領域における畳み込みにより表現されている。具体的には、あるフレームｆの周波数成分がＭフレームにわたって観測信号の周波数成分に影響を及ぼすという仮定のもと、元信号Ｘ(ω,ｆ)が、遅延した既知信号（具体的には元信号の遅延ｍの周波数成分）Ｓ(ω,ｆ－ｍ＋１)およびその伝達関数Ａ(ω,ｍ)の畳み込みとして関係式（１２）により表現されている。

【００４０】

$$X(\omega,f)=\Sigma_{m=1-M}A(\omega,m)S(\omega,f-m+1) \quad ..(12)$$

【００４１】

図５には当該畳み込みの模式図が示されている。畳み込まれた未知信号Ｅ(ω,ｆ)と、通常の伝達過程を経た既知音（自発話信号）Ｓ(ω,ｆ)との混合が観測音Ｙ(ω,ｆ)であるとみなされる。これは、一様ＤＴＦ（Discrete Fourier Transform）フィルタバンクによる一種のマルチレート処理に相当する。

【００４２】

第２モデルによれば、適応フィルタ（分離フィルタ）ｈ^を経由した元信号Ｘ(ω,ｆ)と観測信号Ｙ(ω,ｆ)とを包含するように未知信号Ｅ(ω,ｆ)が表現されている。具体的には、第２モデルによる分離過程は、元信号ベクトルＸ、未知信号Ｅ、観測音スペクトルＹおよび分離フィルタｈ^およびｃに基づき、関係式（１３）～（１５）にしたがってベクトル表現されている。

【００４３】

$$^t(E(\omega,f),\ ^tX(\omega,f))=C\ ^t(Y(\omega,f),^tX(\omega,f)),$$
$$C_{11}=c(\omega),C_{ii}=1(i=2..M+1),$$
$$C_{1j}=h_{j-1}\hat{}(j=2..M+1),C_{ki}=0(k\neq i) \quad ..(13)$$

【００４４】

$$X(\omega,f)=^t(X(\omega,f),X(\omega,f-1),..X(\omega,f-M+1)) \quad ..(14)$$

【００４５】

$$h\hat{}(\omega)=(h_1\hat{}(\omega),h_2\hat{}(\omega)..h_M\hat{}(\omega)) \quad ..(15)$$

【００４６】

この表現は複素数が用いられるほかは時間領域ＩＣＡ法と同一であるが、収束性の観点から周波数領域ＩＣＡ法においてよく利用されている関係式（１１）が用いられた。これによりフィルタｈ^の更新は関係式（１６）により表現される。

【００４７】

$$h\hat{}(f+1)=h\hat{}(f)-\mu_1\phi(E(f))X^*(f) \quad ..(16)$$

【００４８】

Ｘ*(ｆ)はＸ(ｆ)の複素共役を表わしている。なお、周波数インデックスωは省略されている。

【００４９】

分離フィルタｃに関する更新がないため、分離フィルタｃは分離行列の初期値c_0のままである。初期値c_0は誤差Ｅの対数密度関数の導関数φ(ｘ)に対して適切に定められるスケーリング係数である。関係式（１６）から明らかなようにフィルタ更新時に誤差（未知信号）Ｅが適切にスケーリングされていれば、その学習は阻害されない。このため、スケーリング係数ａがなんらかの方法にしたがって求められ、これを用いて関数φ(ａＥ)が適用されれば、分離行列の初期値c_0が１であっても差し支えない。スケーリング係数の学習則は時間領域ＩＣＡ法と同様の関係式（７）が用いられればよい。これは、関係式（７）によれば実質的にｅを正規化するスケーリング係数が求められているからである。時

間領域ＩＣＡ法におけるｅはａＥに相当する。

【００５０】

以上から第２モデルによる学習則は関係式（１７）～（１９）により表現される。

【００５１】

$$E(f)=Y(f)-{}^{t}X(f)\hat{h}(f) \quad ..(17)$$

【００５２】

$$\hat{h}(f+1)=\hat{h}(f)+\mu_1\phi(a(f)E(f))X^{*}(f) \quad ..(18)$$

【００５３】

$$a(f+1)=a(f)+\mu_2[1-\phi(a(k)E(k))a^{*}(f)E^{*}(f)]a(f) \quad ..(19)$$

【００５４】

非線形関数φ(ｘ)がｔａｎｈ(|ｘ|)ｅｘｐ(ｉθ(ｘ))等、ｒ(|ｘ|,θ((ｘ))ｅｘｐ(ｉθ(ｘ))という形式を満たしていればａは実数となる。

【００５５】

前記機能を発揮する音源分離システムによれば、第１モデルおよび第２モデルにしたがって、観測信号Ｙ(ω,ｆ)から未知信号Ｅ(ω,ｆ)が抽出される（図３／Ｓ００２～Ｓ００６参照）。第１モデルによれば、現在フレームｆの観測信号Ｙ(ω,ｆ)が現在および過去の所定数Ｍのフレームにわたる元信号Ｘ(ω,ｆ－ｍ＋１)（ｍ＝１～Ｍ）の合成信号として表現されている（関係式（１２）参照）。また、第２モデルにおいて適応的に分離フィルタ$\hat{h}$が設定される（関係式（１６）～（１９）参照）。このため、窓長を変更させることなく、元信号(ω,ｆ)の残響または反響が観測信号Ｙ(ω,ｆ)に及ぼす影響を軽減しながら未知信号Ｅ(ω,ｆ)が抽出されうる。したがって、既知信号Ｓ(ω,ｆ)の残響の影響を軽減するための演算処理負荷を軽減しながら、未知信号Ｅ(ω,ｆ)に基づく音源分離精度の向上を図ることができる。

【００５６】

ここで、関係式（３）および（１８）を比較する。適用領域を除けば、本願発明の拡張周波数領域ＩＣＡ法はスケーリング係数ａおよび関数φによりＬＭＳ（ＮＬＭＳ）法における推定フィルタと相違している。簡単のため、定義域が時間領域（実数）であり、ノイズ（未知信号）が標準正規分布にしたがうと仮定すると、関数φは関係式（２０）により表わされる。

【００５７】

$$\phi(x)=-(d/dx)\log(\exp(-x^2/2))/(2\pi)^{1/2}=x \quad ..(20)$$

【００５８】

これは、関係式（１８）右辺第２項に含まれるφ(ａＥ(ｔ))Ｘ(ｔ)がａＥ(ｔ)Ｘ(ｔ)と表現されることを意味するので、関係式（１８）は関係式（３）と等価になる。これは、関係式（３）において学習係数が適切に定められればＬＭＳ法でもDouble-Talk状態においてフィルタ更新が可能であることを意味する。換言すると、ノイズがガウス分布にしたがっており、かつ、学習係数がノイズのパワーに応じて適切に設定されている場合、ＬＭＳ法はＩＣＡ法と等価な動作をする。

【００５９】

図６にＬＭＳ法およびＩＣＡ法のそれぞれによる分離例が示されている。観測音は前半では自発話のみである一方、後半では自発話と相手発話とが混じっている。ＬＭＳ法によればノイズがない区間では拘束に収束しているが、ノイズがあるDouble-Talk状態では不安定な動作を示している。これに対して、ＩＣＡ法によれば収束は遅いもののノイズがある区間でも安定である。

【００６０】

続いて、Ａ．時間領域ＮＬＭＳ法、Ｂ．時間領域ＩＣＡ法、Ｃ．周波数領域ＩＣＡ法およびＤ．本願発明の手法のそれぞれの連続音源分離性能の実験結果について説明する。

【００６１】

実験に際して図７に示されているように４．２ｍ×７ｍの広さの部屋（残響時間（ＲＴ

６０）が約０．３秒）において、サンプリングレート１６ｋＨｚでインパルス応答が録音された。自発話に対応するスピーカＳはマイク付近に設置され、マイクＭに対するスピーカＳが向く方向を正面方向とした。相手発話に対応するスピーカはマイクに向けて設置された。マイクＭとスピーカとの距離は１．５ｍとされた。録音されたインパルス応答を畳み込んだＡＳＪ－ＪＮＡＳの評価用データセット２００文（男女各１００文）が評価用データとして用いられた。この２００文を相手発話とし、自発話にはその中の一文（約７秒）を用いた。混合されたデータは、相手発話および自発話の始まりは揃っているが終わりは揃っていない。

【００６２】

音源分離エンジンとしてＪｕｌｉｕｓが使用された（1216856767017_0参照）。クリーン音声２００話者（男性１００人、女性１００人）分のＡＳＪ－ＪＮＡＳ新聞記事読み上げ、および、音素バランス文計１５０文で学習したトライフォン（３状態８混合のＨＭＭ）が音響モデルとして使用された。ＭＦＣＣ（１２＋Δ１２＋ΔＰｏｗ）２５次元が音源分離特徴量として用いられた。認識に用いられた音声は学習データに含まれていない。

【００６３】

実験条件を一致させるため、時間領域におけるフィルタ長が約０．１２８秒に設定された。これにより、手法Ａおよび手法Ｂのフィルタ長は２０４８（約０．１２８秒）となる。手法Ｄでは窓長Ｔが１０２４（０．０６４秒）に設定され、シフト長Ｕが１２８（約０．００８秒）に設定され、かつ、遅延フレーム数Ｍが８に設定されることにより手法Ａおよび手法Ｂと条件を一致させた。手法Ｃでは窓長Ｔが２０４８（０．１２８秒）に設定され、シフト長Ｕが手法Ｄと同様に１２８（０．００８秒）に設定された。フィルタの初期値はすべて０に設定され、オンライン処理で分離が実行された。

【００６４】

学習係数の値としては試行錯誤により認識率が最高になる値が選択された。学習係数は収束性および分離性能を左右する因子であるが、最適値から大きく外れていない限り性能を著しく変化させることはない。

【００６５】

図８に認識結果である単語認識率が示されている。「観測音」は適応フィルタがない状態、すなわち、なんら処理が施されない状態での認識結果を表わしている。「単独発話」は自発話の混合がない状態、すなわち、ノイズがない状態での認識結果を表わしている。クリーン音声の一般的な認識率は約９０％であることから、図８から明らかなように部屋環境の影響によって約２０％も認識率が低下している。手法Ａでは観測音と比較して認識率が０．８７％だけ低下している。これは、自発話と相手発話とが混在するDouble-Talk状態では手法Ａの動作が不安定になることを反映しているためであると推察される。手法Ｂでは観測音と比較して認識率が４．２１％だけ上昇し、手法Ｃでは観測音と比較して認識率が７．５５％だけ上昇している。これは、時間領域で処理が実行される手法Ｂよりも、周波数領域で処理が実行される結果として周波数ごとの特性が反映される手法Ｃのほうがよい結果が得られることを表わしている。手法Ｄでは観測音と比較して認識率が９．６１％だけ上昇しており、従来手法Ａ～Ｃよりも有効な音源分離手法であることが確認された。

【図面の簡単な説明】

【００６６】

【図１】本発明の音源分離システムの構成説明図
【図２】本発明の音源分離システムのロボットへの搭載例示図
【図３】本発明の音源分離システムの機能を示すフローチャート
【図４】適応フィルタの構成に関する説明図
【図５】時間周波数領域における畳み込みに関する説明図
【図６】ＬＭＳ法およびＩＣＡ法による相手発話の分離結果に関する説明図
【図７】実験状況に関する説明図
【図８】音源分離結果としての各手法による単語正解率の比較説明図

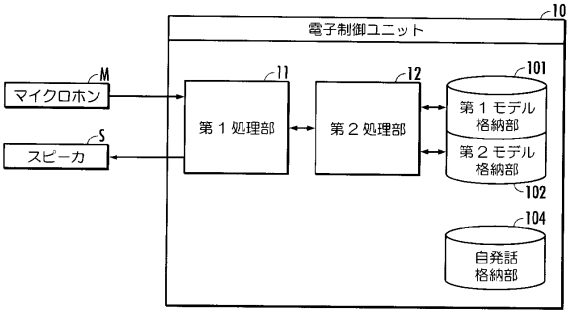
【符号の説明】

【００６７】

１０‥電子制御ユニット、１１‥第１処理部、１２‥第２処理部、Ｓ‥スピーカ、Ｍ‥マイクロホン

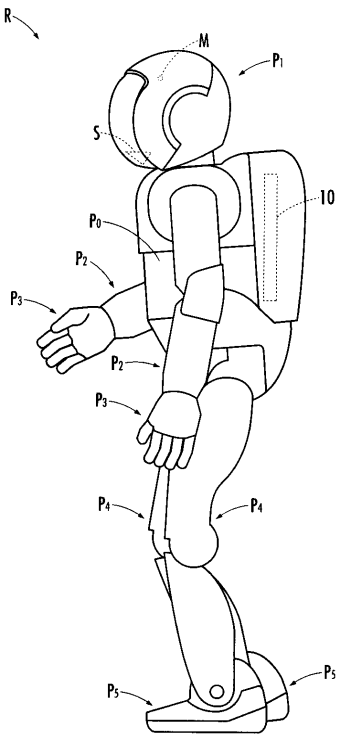
【図１】

FIG.1



【図２】

FIG.2

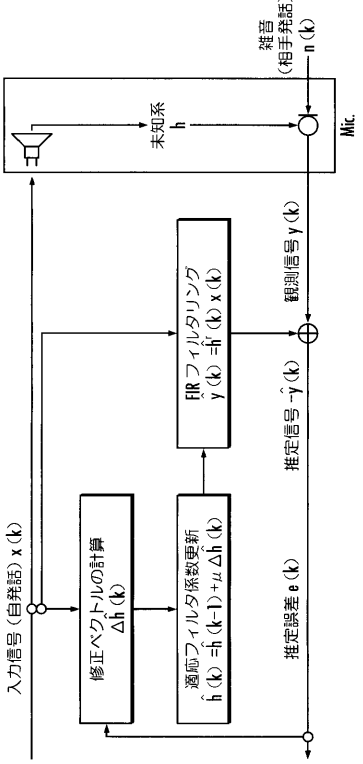


【図３】

FIG.3

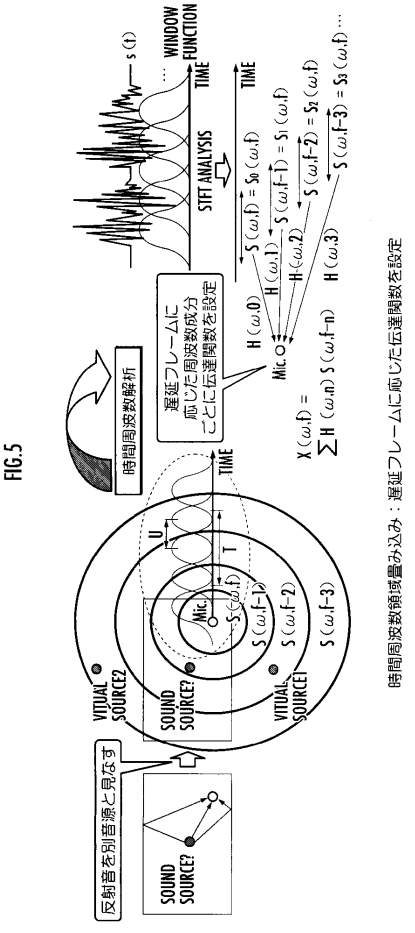
【図４】

FIG.4



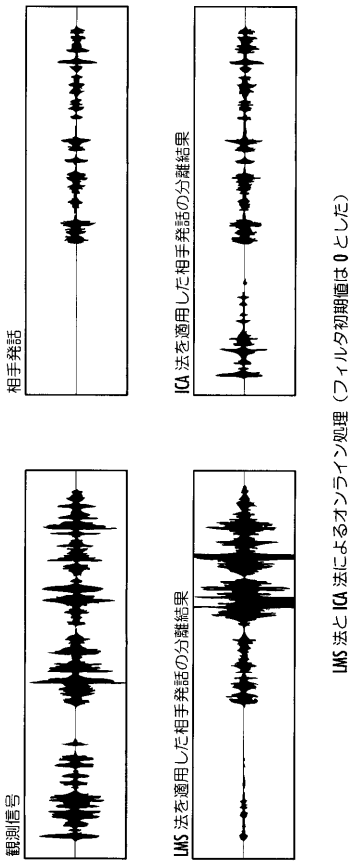
【図５】

FIG.5



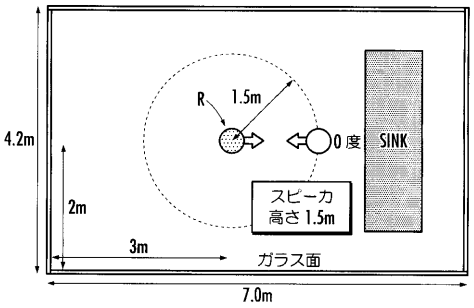
【図６】

FIG.6



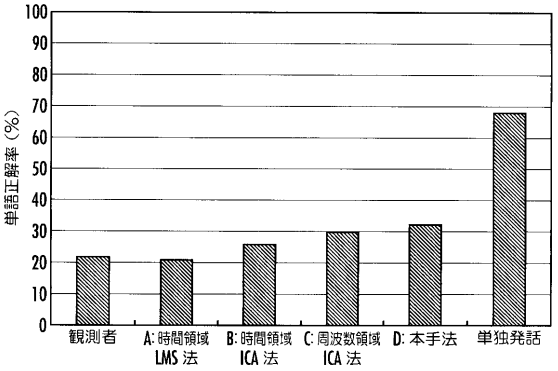
【図７】

FIG.7



【図８】

FIG.8



フロントページの続き

(72)発明者　辻野　広司
埼玉県和光市本町８－１　株式会社ホンダ・リサーチ・インスティチュート・ジャパン内
(72)発明者　奥乃　博
京都府京都市中京区東洞院通三条下る三文字町２０５番地の３－１１０２

審査官　菊池　智紀

(56)参考文献　特開２００６－１６３２３１（ＪＰ，Ａ）
大田健紘 他，"既知雑音除去法の有効性に関する検討"，電子情報通信学会技術研究報告，２００６年　5月１９日，Vol.106，No.78，p.7-12
武田龍 他，"ICAとMFTに基づく音声認識におけるSoft Maskを用いた性能評価"，情報処理学会第69回全国大会予稿集(2)，２００７年　3月　6日，p.2-585－2-586
Shigeki Miyabe et al.，"Double-talk Free Spoken Dialogue Interface Combining Sound Field Control with Semi-blind Source Separation"，Proc. of IEEE ICASSP2006，２００６年　5月１４日，Vol.1，p.I-809－I-812

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１５／２０，２１／０２
ＪＳＴＰｌｕｓ（ＪＤｒｅａｍＩＩ）