

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4162604号
(P4162604)

(45) 発行日 平成20年10月8日(2008.10.8)

(24) 登録日 平成20年8月1日(2008.8.1)

(51) Int.Cl. F I
G 1 O L 15/20 (2006.01) G 1 O L 15/20 3 7 O D
G 1 O L 21/02 (2006.01) G 1 O L 15/20 3 7 O E
 G 1 O L 15/20 3 7 O F
 G 1 O L 21/02 1 0 1 B

請求項の数 7 (全 17 頁)

(21) 出願番号	特願2004-3108 (P2004-3108)	(73) 特許権者	000003078 株式会社東芝
(22) 出願日	平成16年1月8日(2004.1.8)		東京都港区芝浦一丁目1番1号
(65) 公開番号	特開2005-195955 (P2005-195955A)	(74) 代理人	100109900 弁理士 堀口 浩
(43) 公開日	平成17年7月21日(2005.7.21)	(72) 発明者	天田 皇 神奈川県川崎市幸区小向東芝町1番地 株 式会社東芝 研究開発センター内
審査請求日	平成17年2月10日(2005.2.10)	(72) 発明者	河村 聡典 神奈川県川崎市幸区小向東芝町1番地 株 式会社東芝 研究開発センター内
		(72) 発明者	小柴 亮典 神奈川県川崎市幸区小向東芝町1番地 株 式会社東芝 研究開発センター内

最終頁に続く

(54) 【発明の名称】 雑音抑圧装置及び雑音抑圧方法

(57) 【特許請求の範囲】

【請求項1】

雑音信号と目的信号が混合した入力信号から雑音信号を抑圧する雑音抑圧装置において、前記入力信号から雑音信号成分を推定する雑音推定手段と、前記入力信号から目的信号区間と雑音信号区間を判定する区間判定手段と、前記入力信号と前記推定雑音信号とから第1の抑圧係数に応じて雑音抑圧をする雑音抑圧手段と、前記入力信号と前記推定雑音信号とから前記第1の抑圧係数よりも大きな第2の抑圧係数に応じて雑音抑圧をする雑音過剰抑圧手段と、目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を前記入力信号に乗じた補正用信号を生成する補正用信号生成手段と、前記補正用信号と前記雑音過剰抑圧手段の出力とを加算する加算手段と、前記区間判定手段の判定結果に応じて前記雑音抑圧手段の出力信号と前記加算手段の出力信号とを切替える切替手段とを具備したことを特徴とする雑音抑圧装置。

【請求項2】

雑音信号と目的信号が混合した入力信号から雑音信号を抑圧する雑音抑圧装置において、前記入力信号から雑音信号成分を推定する雑音推定手段と、前記入力信号から目的信号区間と雑音信号区間を判定する区間判定手段と、前記入力信号と前記推定雑音信号とから第1の抑圧係数を算出する抑圧係数算出手段と、前記入力信号と前記推定雑音信号とから前記第1の抑圧係数よりも大きな第2の抑圧係数を算出する過剰抑圧係数算出手段と、前記入力信号から目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を生成する補正用係数生成手段と、前記補正用係数と前記第2の抑圧係数とを加算する加算

手段と、前記区間判定手段の判定結果に応じて前記第 1 の抑圧係数と前記加算手段で加算された係数とを切替える切替手段と、前記切替手段により切替えられた係数を前記入力信号に乗じる乗算手段とを具備したことを特徴とする雑音抑圧装置。

【請求項 3】

前記区間判定手段は、前記入力信号と前記推定雑音信号とから目的信号区間と雑音信号区間とを判定することを特徴とする請求項 1 または請求項 2 記載の雑音抑圧装置。

【請求項 4】

雑音信号と目的信号が混合した複数の入力信号から雑音信号を抑圧する雑音抑圧装置において、前記複数の入力信号から目的信号が強調される統合信号を生成する統合信号生成手段と、前記統合信号から雑音信号成分を推定する雑音推定手段と、前記複数の入力信号から目的信号区間と雑音信号区間を判定する区間判定手段と、前記統合信号と前記推定雑音信号とから第 1 の抑圧係数を算出する抑圧係数算出手段と、前記統合信号と前記推定雑音信号とから前記第 1 の抑圧係数よりも大きな第 2 の抑圧係数を算出する過剰抑圧係数算出手段と、前記統合信号から目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を生成する補正用係数生成手段と、前記補正用係数と前記第 2 の抑圧係数とを加算する加算手段と、前記区間判定手段の判定結果に応じて前記第 1 の抑圧係数と前記加算手段で加算された係数とを切替える切替手段と、前記切替手段により切替えられた係数を前記統合信号に乗じる乗算手段とを具備したことを特徴とする雑音抑圧装置。

【請求項 5】

雑音信号と目的信号が混合した複数の入力信号から雑音信号を抑圧する雑音抑圧装置において、前記複数の入力信号から目的信号が強調される統合信号を生成する統合信号生成手段と、前記複数の入力信号から目的信号が抑圧された目的音除去信号を生成する目的音除去信号生成手段と、前記統合信号と前記目的音除去信号から雑音信号成分を推定する雑音推定手段と、前記複数の入力信号から目的信号区間と雑音信号区間を判定する区間判定手段と、前記統合信号と前記推定雑音信号とから第 1 の抑圧係数を算出する抑圧係数算出手段と、前記統合信号と前記推定雑音信号とから前記第 1 の抑圧係数よりも大きな第 2 の抑圧係数を算出する過剰抑圧係数算出手段と、前記統合信号から目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を生成する補正用係数生成手段と、前記補正用係数と前記第 2 の抑圧係数とを加算する加算手段と、前記区間判定手段の判定結果に応じて前記第 1 の抑圧係数と前記加算手段で加算された係数とを切替える切替手段と、前記切替手段により切替えられた係数を前記統合信号に乗じる乗算手段とを具備したことを特徴とする雑音抑圧装置。

【請求項 6】

雑音信号と目的信号が混合した複数の入力信号から雑音信号を抑圧する雑音抑圧装置において、前記複数の入力信号から周波数帯域毎に目的信号が強調されるサブバンド統合信号を生成するサブバンド統合信号生成手段と、前記サブバンド統合信号からサブバンド毎の雑音信号成分を推定する雑音推定手段と、前記複数の入力信号からサブバンド毎に目的信号区間と雑音信号区間を判定する区間判定手段と、前記サブバンド統合信号と前記推定雑音信号とからサブバンド毎に第 1 の抑圧係数を算出する抑圧係数算出手段と、前記サブバンド統合信号と前記推定雑音信号とからサブバンド毎に前記第 1 の抑圧係数よりも大きな第 2 の抑圧係数を算出する過剰抑圧係数算出手段と、前記サブバンド統合信号からサブバンド毎に目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を生成する補正用係数生成手段と、サブバンド毎に前記補正用係数と前記第 2 の抑圧係数とを加算する加算手段と、前記区間判定手段の判定結果に応じてサブバンド毎に前記第 1 の抑圧係数と前記加算手段で加算された係数とを切替える切替手段と、前記切替手段により切替えられた係数をサブバンド毎に前記サブバンド統合信号に乗じる乗算手段と、サブバンド毎の前記乗算手段で係数を乗じたサブバンド統合信号を合成する合成手段とを具備したことを特徴とする雑音抑圧装置。

【請求項 7】

雑音信号と目的信号が混合した入力信号から雑音信号を抑圧する雑音抑圧方法において、

10

20

30

40

50

前記入力信号から雑音信号成分を雑音推定手段により推定し、前記入力信号から目的信号区間と雑音信号区間を区間判定手段により判定し、前記入力信号と前記推定雑音信号とから第1の抑圧係数に応じて雑音抑圧手段により雑音抑圧し、前記入力信号と前記推定雑音信号とから前記第1の抑圧係数よりも大きな第2の抑圧係数に応じて雑音過剰抑圧手段により雑音抑圧し、目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を前記入力信号に乗じた補正用信号を補正用信号生成手段により生成し、前記補正用信号と前記雑音過剰抑圧手段の出力とを加算手段により加算し、前記区間判定手段の判定結果に応じて前記雑音抑圧手段の出力信号と前記加算手段の出力信号とを切替手段により切替えることを特徴とする雑音抑圧方法。

【発明の詳細な説明】

10

【技術分野】

【0001】

本発明は、ハンズフリー通話や音声認識等で用いられる雑音抑圧技術の一つであり、入力音響信号から目的とする音声信号を強調して出力する技術に関する。

【背景技術】

【0002】

実環境での音声認識や携帯電話の実用化に伴い、雑音の重畳した信号から雑音を取り除き音声信号のみを強調する信号処理方法が重要になってきている。スペクトルサブトラクション(Spectral Subtraction: S S)は、効果的で実現がしやすいためしばしば用いられる(例えば、非特許文献1を参照)。

20

【0003】

スペクトルサブトラクションにはミュージカルノイズと呼ばれる聴覚上不自然に聞こえる音が生成される問題がある。これは雑音区間で特に顕著であり、実際にはバラツキが存在する入力信号(雑音信号)からその平均値を引き去ることで、消し残しの成分が不連続に存在することに起因する。この問題を解決する方法として、過剰抑圧を行うという方法がある。過剰抑圧とは推定ノイズよりも大きな値を引去り、雑音の変動成分も含めて抑圧してしまう方法である。なお、減算により負の値になる場合は最小値で置き換えるなどの処理が行われる。しかし、過剰抑圧は、音声区間で抑圧量が過剰になり、音声が歪んでしまうといった問題があった(例えば、非特許文献2を参照)。

【0004】

30

また、ミュージカルノイズの発生した区間に何らかの処理を施して目立たなくする方法、例えば入力信号などに小さなゲインをかけて加えるなどの方法もあるが、この方法ではミュージカルノイズが知覚できなくなるまで十分な信号を重畳すると、重畳した信号により雑音レベルが上がり、雑音抑圧の効果が失われかねない問題があった。

【0005】

【非特許文献1】S.Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans., ASSP-27, No.2, pp.113-120, 1979

【非特許文献2】Z.Goh, K.Tan and B.T.G.Tan, "Postprocessing Method for Suppressing Musical Noise Generated by spectral Subtraction", IEEE Trans., SAP-6, No. 3, May 1998

40

【発明の開示】

【発明が解決しようとする課題】

【0006】

上述したように、抑圧係数を大きくして過剰抑圧することは、ミュージカルノイズを押さえる効果はあるものの、音声区間での歪みを生みやすいという問題があった。また、ミュージカルノイズに入力信号を重畳するなどの後処理を用いた手法では、ミュージカルノイズを知覚できなくなるに十分な音量を重畳すると、雑音抑圧の効果が失われる問題があった。

【0007】

本発明は、このような課題を解決するためになされたものであり、雑音区間ではミュー

50

ジカルノイズが発生せず、音声区間での歪みも発生しない雑音抑圧装置、及び雑音抑圧方法を提供することを目的とする。

【課題を解決するための手段】

【0009】

上記の課題を解決するために本発明に係る雑音抑圧装置は、雑音信号と目的信号が混合した入力信号から雑音信号を抑圧する雑音抑圧装置において、前記入力信号から雑音信号成分を推定する雑音推定手段と、前記入力信号から目的信号区間と雑音信号区間を判定する区間判定手段と、前記入力信号と前記推定雑音信号とから第1の抑圧係数に応じて雑音抑圧をする雑音抑圧手段と、前記入力信号と前記推定雑音信号とから前記第1の抑圧係数よりも大きな第2の抑圧係数に応じて雑音抑圧をする雑音過剰抑圧手段と、目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を前記入力信号に乗じた補正用信号を生成する補正用信号生成手段と、前記補正用信号と前記雑音過剰抑圧手段の出力とを加算する加算手段と、前記区間判定手段の判定結果に応じて前記雑音抑圧手段の出力信号と前記加算手段の出力信号とを切替える切替手段とを具備したことを特徴とする。

10

【0011】

また、雑音信号と目的信号が混合した入力信号から雑音信号を抑圧する雑音抑圧装置において、前記入力信号から雑音信号成分を推定する雑音推定手段と、前記入力信号から目的信号区間と雑音信号区間を判定する区間判定手段と、前記入力信号と前記推定雑音信号とから第1の抑圧係数を算出する抑圧係数算出手段と、前記入力信号と前記推定雑音信号とから前記第1の抑圧係数よりも大きな第2の抑圧係数を算出する過剰抑圧係数算出手段と、前記入力信号から目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を生成する補正用係数生成手段と、前記補正用係数と前記第2の抑圧係数とを加算する加算手段と、前記区間判定手段の判定結果に応じて前記第1の抑圧係数と前記加算手段で加算された係数とを切替える切替手段と、前記切替手段により切替えられた係数を前記入力信号に乗じる乗算手段とを具備したことを特徴とする。

20

【0013】

また、前記区間判定手段は、前記入力信号と前記推定雑音信号とから目的信号区間と雑音信号区間とを判定することを特徴とする。

【0014】

また、雑音信号と目的信号が混合した複数の入力信号から雑音信号を抑圧する雑音抑圧装置において、前記複数の入力信号から目的信号が強調される統合信号を生成する統合信号生成手段と、前記統合信号から雑音信号成分を推定する雑音推定手段と、前記複数の入力信号から目的信号区間と雑音信号区間を判定する区間判定手段と、前記統合信号と前記推定雑音信号とから第1の抑圧係数を算出する抑圧係数算出手段と、前記統合信号と前記推定雑音信号とから前記第1の抑圧係数よりも大きな第2の抑圧係数を算出する過剰抑圧係数算出手段と、前記統合信号から目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を生成する補正用係数生成手段と、前記補正用係数と前記第2の抑圧係数とを加算する加算手段と、前記区間判定手段の判定結果に応じて前記第1の抑圧係数と前記加算手段で加算された係数とを切替える切替手段と、前記切替手段により切替えられた係数を前記統合信号に乗じる乗算手段とを具備したことを特徴とする。

30

40

【0015】

また、雑音信号と目的信号が混合した複数の入力信号から雑音信号を抑圧する雑音抑圧装置において、前記複数の入力信号から周波数帯域毎に目的信号が強調されるサブバンド統合信号を生成するサブバンド統合信号生成手段と、前記サブバンド統合信号からサブバンド毎の雑音信号成分を推定する雑音推定手段と、前記複数の入力信号からサブバンド毎に目的信号区間と雑音信号区間を判定する区間判定手段と、前記サブバンド統合信号と前記推定雑音信号とからサブバンド毎に第1の抑圧係数を算出する抑圧係数算出手段と、前記サブバンド統合信号と前記推定雑音信号とからサブバンド毎に前記第1の抑圧係数よりも大きな第2の抑圧係数を算出する過剰抑圧係数算出手段と、前記サブバンド統合信号からサブバンド毎に目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係

50

数を生成する補正用係数生成手段と、サブバンド毎に前記補正用係数と前記第2の抑圧係数とを加算する加算手段と、前記区間判定手段の判定結果に応じてサブバンド毎に前記第1の抑圧係数と前記加算手段で加算された係数とを切替える切替手段と、前記切替手段により切替えられた係数をサブバンド毎に前記サブバンド統合信号に乗じる乗算手段と、サブバンド毎の前記乗算手段で係数を乗じたサブバンド統合信号を合成する合成手段とを具備したことを特徴とする。

【0016】

また、雑音信号と目的信号が混合した入力信号から雑音信号を抑圧する雑音抑圧方法において、前記入力信号から雑音信号成分を雑音推定手段により推定し、前記入力信号から目的信号区間と雑音信号区間を区間判定手段により判定し、前記入力信号と前記推定雑音信号とから第1の抑圧係数に応じて雑音抑圧手段により雑音抑圧し、前記入力信号と前記推定雑音信号とから前記第1の抑圧係数よりも大きな第2の抑圧係数に応じて雑音過剰抑圧手段により雑音抑圧し、目的信号区間の出力に残留する雑音信号とのレベルの違いを補正する係数を前記入力信号に乗じた補正用信号を補正用信号生成手段により生成し、前記補正用信号と前記雑音過剰抑圧手段の出力とを加算手段により加算し、前記区間判定手段の判定結果に応じて前記雑音抑圧手段の出力信号と前記加算手段の出力信号とを切替手段により切替えることを特徴とする。

【発明の効果】

【0017】

本発明によれば、音声区間に歪みを生むことなく雑音区間に不自然な消し残し音を発生させることなく雑音を抑圧することができる。

【発明を実施するための最良の形態】

【0018】

以下、図面を参照して本発明の実施の形態について詳細に説明する。

【実施例1】

【0019】

図1は本発明の第1の実施形態に係る雑音抑圧装置の構成を表すブロック図である。図1に示されるように、第1の実施形態の雑音抑圧装置は、音響信号を入力するための入力端子101と、音響信号を周波数領域に変換する周波数変換部102と、この出力から推定雑音を求める雑音推定部103と、周波数変換部102から雑音推定部103の出力を用いて雑音が抑圧された信号を生成する雑音抑圧部104と、同じくより強く雑音が抑圧された信号を生成する雑音過剰抑圧部105と、周波数変換部102の出力から雑音レベルを補正する信号を生成する雑音レベル補正用信号生成部106と、雑音過剰抑圧部105と雑音レベル補正用信号生成部106との出力を加算する加算部107と、入力信号から音声区間か雑音区間かを判定する音声・雑音判定部108と、音声・雑音判定部108の出力により雑音抑圧部104の出力と加算部107の出力を選択する切替部109と、この出力を時間領域に変換する周波数逆変換部110から構成される。

【0020】

入力端子101には、

【0021】

【数1】

$$x(t) = s(t) + n(t)$$

【0022】

で表される信号が入力される。ここで $x(t)$ はマイクなどで受音した時間波形を表す信号であり、 $s(t)$ はその中の目的信号成分(例えば音声)であり、 $n(t)$ は非目的信号成分(例えば周囲の雑音)である。入力された信号 $x(t)$ は周波数変換部102においてDFTなどを用いて所定の窓幅で周波数領域に変換され $X(f)$ を得る。(fは周波数を表す。)

10

20

30

40

50

【 0 0 2 3 】

雑音推定部103では、 $X(f)$ から雑音信号の推定値 $Ne(f)$ を推定する。この推定には、例えば $s(t)$ が音声信号の場合、非発話区間が存在するので、その区間は $x(t) = n(t)$ となり、その区間の平均値を $Ne(f)$ とする。これを用いて、

【 0 0 2 4 】

【数2】

$$|Se(f)| = |X(f) - \alpha Ne(f)|$$

【 0 0 2 5 】

として、音声の推定値 $|Se(f)|$ を得る。これを時間領域にもどすことで、音声のみを推定することができる。 $|Se(f)|$ は振幅値のみで位相項がないので、これには入力信号 $X(f)$ の位相項を用いるのが一般的である。(数2)は振幅スペクトルで行う方法であるが、パワースペクトルを用いる方法もあり、一般的な表記を用いると

10

【 0 0 2 6 】

【数3】

$$|Se(f)|^b = |X(f)|^b - \alpha |Ne(f)|^b$$

【 0 0 2 7 】

と表すことができる。スペクトルサブトラクションをフィルタ演算とみなして、

20

【 0 0 2 8 】

【数4】

$$Se(f) = \left\{ \frac{(|X(f)|^b - \alpha |Ne(f)|^b)}{|X(f)|^b} \right\}^{\left(\frac{1}{a}\right)} X(f)$$

【 0 0 2 9 】

と表記することもできる。 $(a, b) = (1, 1)$ の場合は振幅スペクトルを用いたスペクトルサブトラクション(数2)と等価になる。また $(a, b) = (2, 2)$ の場合はパワースペクトルを用いたスペクトルサブトラクションとなる。さらに、 $(a, b) = (1, 2)$ かつ $\alpha = 1$ の場合はWienerフィルタの形式となる。これらは実現上においては(数4)で統一的に記述できる同種の手法と見なすことができる。

30

【 0 0 3 0 】

ここで、一般に $X(f)$ は複素数であり、

【 0 0 3 1 】

【数5】

$$X(f) = |X(f)| \exp(j \arg(X(f)))$$

40

【 0 0 3 2 】

と表される。 $|X(f)|$ は $X(f)$ の大きさを、 $\arg(X(f))$ は位相を、 j は虚数単位である。周波数変換部102からは $X(f)$ の大きさが出力されるが、ここでは指数 b を付加した一般的な表現を用いることにする。その理由は、スペクトルサブトラクションは(数3)で述べたように、いくつかのバリエーションが存在するためである。 b の値は1または2である場合が多い。雑音推定部103は $|X(f)|^b$ から推定雑音 $|Ne(f)|^b$ を求める。これには $|X(f)|^b$ から雑音区間と見なされる区間の平均値を用いる。

【 0 0 3 3 】

50

例えば、雑音区間において

【0034】

【数6】

$$|Ne(f, n)|^b = \delta |Ne(f, n-1)|^b + (1-\delta) |X(f)|^b$$

【0035】

とする方法などがある。ただし、 $|Ne(f, n)|^b$ は現在フレームの $|Ne(f)|^b$ で $|Ne(f, n-1)|^b$ は1つ前のフレームの値であり、 δ は $0 < \delta < 1$ なる値で、平滑化の度合いを制御する。音声区間か否かは $|X(f)|^b$ の大きさが大きい区間を音声区間とする方法や、 $|X(f)|^b$ と $|Ne(f, n)|^b$ の比率を求め、 $|X(f)|^b$ がある比率よりも大きくなる区間を音声とするなどの方法がある。

10

【0036】

雑音抑圧部104、雑音過剰抑圧部105では周波数変換部102の出力 $|X(f)|^b$ から雑音推定部103の出力 $|Ne(f)|^b$ を引き去り出力信号 $|S(f)|^b$ を出力する。これには(数3)の方法を用いるが、推定雑音 $|Ne(f)|$ が入力信号 $|X(f)|$ よりも大きい場合などいくつかの処理方法がある。ここでは、

【0037】

【数7】

$$|Se(f)|^b = \text{Max}(|X(f)|^b - \alpha |Ne(f)|^b, \beta |X(f)|^b)$$

20

【0038】

を用いることにする。ここで、 $\text{Max}(x, y)$ は x, y の大きい方を表し、 α は抑圧係数、 β はフロアリング係数である。 α の値は大きいほど多くの雑音を取り除くことができるので、雑音抑圧効果は大きくなるが、音声が存在する区間では音声成分も引き去られ出力信号に歪みを生じる。 β は正の小さな値で演算結果が負になることを抑止する。例えば $(\alpha, \beta) = (1.0, 0.01)$ 等である。

【0039】

本発明においては雑音抑圧部104の抑圧係数(s)よりも過剰雑音抑圧部105の抑圧係数(n)を大きくしている。また、過剰雑音抑圧部105では大きな抑圧係数が用いられているため雑音の平均的なパワー(雑音レベル)が雑音抑圧部104に比べて下がる。これを補償する手段として雑音レベル補正用信号生成部106を用いている。

30

【0040】

ここでは、入力信号 $|X(f)|^b$ にゲインをかけた信号

【0041】

【数8】

$$|M(f)|^b = (1-\alpha s) |X(f)|^b$$

40

【0042】

を生成し、これを雑音過剰抑圧部105の出力に加算部107において加える。

切替部109では雑音抑圧部104と加算部107の出力を選択して出力信号を生成する。切替は音声・雑音判定部108の判定結果に基づき、音声区間では雑音抑圧部104の出力を、雑音区間では加算部107の出力を選択する。音声・雑音判定部108での判定方法は様々な方法が存在するが、例えば信号のパワーと閾値を用いて判定する方法などがある。

【0043】

最後に周波数逆変換部110で切替部109の出力を周波数領域から時間領域に変換され音声の強調された時間信号が得られる。さらに、フレーム単位で処理する場合はオーバーラップアドにより時間的に連続した信号を生成する場合でも適用できる。また周波数逆変換部

50

110を用いずに時間領域に変換せず周波数領域のまま出力してもよい。

【0044】

雑音過剰抑圧部105と雑音レベル補正用信号生成部106に関して詳細に説明する。上述したようにスペクトルサブトラクションにはミュージカルノイズと呼ばれる雑音区間での引き残しが不自然な音になる現象が存在する。図2を用いてこの現象を模式的に説明する。図2(a)は周波数変換した入力信号のある周波数 f の振幅値 ($|X(f)|$) をフレーム(時刻)ごとに表したものである。ここでは $b=1$ として指数部を省略して表す。空白の箱が $|X(f)|$ の雑音成分で斜線の箱が音声成分である。3本の点線のうち中央が雑音推定部により出力された推定雑音の大きさ $|N_e(f)|$ を示し、上側が過剰抑圧を行う場合の値 $n|N_e(f)|$ で、下側が通常の抑圧を行う場合の値 $s|N_e(f)|$ である。まず、 $n=1$ で抑圧を行うと $|N_e(f)|$ だけ振幅が減少し図2(b)のようになる。これは通常のスเปクトルサブトラクションであり、雑音区間のノイズが減少し、音声が強調されている。しかし、雑音区間に引き残し成分が間欠的に存在しミュージカルノイズとなって聞こえる。また、音声区間では引きすぎにより音声成分の一部が欠けてしまう。これは音声の歪みとなって知覚される。

10

【0045】

図2(c)は $n|N_e(f)|$ で過剰抑圧を行った場合である。雑音区間は完全に抑圧されミュージカルノイズは発生していないが、音声成分がかなり削られ大きな歪みが発生する。図2(d)は $s|N_e(f)|$ で抑圧を行った場合である。音声成分に歪みは出ていないが、雑音区間に信号が間欠的に残る現象がまだ存在する。本発明は、図2(e)に示すように音声区間と雑音区間を予め区別しておき、音声区間では歪みの生じない図2(d)の方法で抑圧し、雑音区間は過剰抑圧により図2(c)のように強い抑圧を行いミュージカルノイズを完全に除去される。

20

【0046】

ところで、図2(e)では雑音区間では雑音は完全に除去されているものの、音声区間では歪みを発生させないかわりに雑音も残っているため、この雑音を知覚され、雑音レベルが不連続に聞こえる場合がある。この問題を解決するため、図2(f)に示したように雑音区間にのみ、入力信号のレベルを低減させた信号を加算することにより雑音レベルを揃える。以上が本発明の模式的な説明である。厳密には雑音と音声を加算した信号の振幅はそれぞれの振幅の和になるとは限らないなど正確な表現になっていない点は考慮しておく必要がある。

30

【0047】

本発明ではミュージカルノイズを消しているのは過剰抑圧であり、入力信号の加算は音声区間との雑音レベルの違いを埋めるために行っている。これは、ミュージカルノイズを入力音声の加算で知覚しにくくする従来の方法とは異なる。従って本発明では、音声区間での抑圧係数を大きくとることにより、雑音区間で付加する信号のレベルを小さくすることが可能であり、この操作によりミュージカルノイズの削減効果が左右されることはない。

【0048】

一方、従来は加算する信号のレベルとミュージカルノイズの知覚されやすさとは密接な関係があり、加算量を少なくするとミュージカルノイズは知覚されやすくなる。(数8)で用いられている入力信号に対するゲイン ($1-s$) は次のように求められる。

40

【0049】

まず、音声区間で歪みを生じないように抑圧係数 s が弱めに設定されるので、 s は1より小さい値となる。したがって、仮に音声区間が雑音のみであった場合、 $(1-s)$ の雑音は引かれずに残ることになる。一方雑音区間では過剰抑圧により雑音はゼロになっている。したがって、その差 $(1-s)$ 分の信号を雑音区間に加えれば音声区間の雑音とレベルがそろうことになる。

【0050】

ところで、音声区間の抑圧量 s が1に近い場合、付加する雑音のゲイン ($1-s$)

50

の値は小さな値となる。このような場合は音声区間と雑音区間の雑音レベルの差が知覚されにくいため、加算そのものを行わないという方法であってもよい。また、分散の大きな雑音の場合は、この方法でもレベル差が完全に補償できないことがあり、その場合は分散を考慮した補償方法を用いることも可能である。

【 0 0 5 1 】

図 2 (g) は全区間雑音と誤った判定がされた場合の過剰抑圧後の状態を模式的に表している。上述している通り、過剰抑圧を行うと雑音区間ではミュージカルノイズは生じないが、音声区間に大きな歪みを生む。ここではこの後に入力信号の加算を行うため、誤って雑音と判断された音声区間には雑音成分とともに音声成分も加算されることになり、一度生じた歪みを回復させる効果がある (図 2 (h))。つまり、音声区間を雑音区間と誤った場合でも、音声区間が誤って抑圧されないことがない音声・雑音判定結果の誤りに対して頑健であるという効果がある。

10

【実施例 2】

【 0 0 5 2 】

図 3 は本発明の第 2 の実施形態に係る雑音抑圧装置の構成を示すブロック図である。第 2 の実施形態の雑音抑圧装置は、上述した第 1 の実施形態におけるスペクトルサブトラクションは伝達関数を乗算する形式にした場合の構成であり、第 1 の実施形態では (数 3) に相当する減算形の抑圧方法であるのに対し、第 2 の実施形態は (数 4) の乗算形に相当する。これらは本質的には同じであるため、以降の実施形態においても (数 3) に相当する減算形の方法で実現することも可能である。第 2 の実施形態と第 1 の実施形態との違いは、雑音抑圧部 104、雑音過剰抑圧部 105、雑音レベル補正用信号生成部 106 が抑圧係数算出部 204、過剰抑圧係数算出部 205、雑音レベル補正用係数生成部 206 にそれぞれ置き換わり、切替部 107 の出力である重み係数を入力信号に乗算する乗算部 211 が加わっている点である。

20

【 0 0 5 3 】

抑圧係数算出部 204 では抑圧係数は、

【 0 0 5 4 】

【数 9】

$$ws(f) = \text{Max} \left(\frac{(|X(f)|^b - \alpha s |Ne(f)|^b)}{|X(f)|^b}, \beta \right)^{\left(\frac{1}{a}\right)}$$

30

【 0 0 5 5 】

と求められ、過剰抑圧係数算出部 205 では

【 0 0 5 6 】

【数 10】

$$wn(f) = \text{Max} \left(\frac{(|X(f)|^b - \alpha n |Ne(f)|^b)}{|X(f)|^b}, \beta \right)^{\left(\frac{1}{a}\right)}$$

40

【 0 0 5 7 】

と求められる。

既に述べたように (a , b) = (1 , 1) の場合は振幅スペクトルを用いたスペクトルサブトラクションと等価であり、(a , b) = (2 , 2) の場合はパワースペクトルを用いたスペクトルサブトラクション、(a , b) = (1 , 2) の場合は Wiener フィルタの形式となる。また、抑圧係数は抑圧係数算出部 204 では s であり、音声区間に歪みを与

50

えない抑圧量が設定されるのに対して、過剰抑圧係数算出部205では n となり、雑音区間でミュージカルノイズを十分に除去するため大きな係数が設定される点も第1の実施形態と同様である。

【0058】

雑音レベル補正用係数生成部206では(数8)に相当する重み係数

【0059】

【数11】

$$w_o(f) = (1 - \alpha s)$$

10

【0060】

と求められる。加算部207では、

【0061】

【数12】

$$w_n(f) = w_n(f) + w_o(f)$$

【0062】

が行われ、音声・雑音判定部208の結果に基づき $w_s(f)$ か $w_n(f)$ を切替部209で選択して最終的な重み係数 $w_w(f)$ を出力する。乗算部211では入力信号のスペクトル $X(f)$ にこの重み係数 $w_w(f)$ をかけ、出力信号 $S(f)$ を

20

【0063】

【数13】

$$S(f) = w_w(f)X(f)$$

【0064】

と求める。

本実施形態は第1の実施形態の表現を伝達関数が乗算される形式に替えただけのものであるが、 $|X(f)|$ の平滑化を行うことで、(数9)(数10)で求める重み係数の局所的な変動を押さえ重み係数の変化を滑らかにすることができ、音質向上につながる。

30

【0065】

一方(数13)の $X(f)$ は平滑化を行うと音がぼやけてしまうので、平滑化を行わない方が望ましい。(数9)(数10)の $X(f)$ の平滑化方法として、例えば(数6)の方法を用いることができる。平滑化に関して本実施例と同等のことを実施例1でも行うことが可能であるが、本実施例の方がより簡単に行える利点がある。

【0066】

また、第1の実施形態と同様に音声区間の抑圧量 s が1に近い場合、付加する雑音のゲイン $(1 - s)$ の値は小さな値となる。このような場合は音声区間と雑音区間の雑音レベルの差が知覚されにくいため、付加そのものを行わなくてもよい。また、分散の大きな雑音の場合は、この方法でもレベル差が完全に補償できないことがあり、その場合は分散を考慮した補償方法を用いることも可能である。

40

【実施例3】

【0067】

図4は本発明の第3の実施形態に係る雑音抑圧装置の構成を表すブロック図である。第2の実施形態の音声・雑音判定部208が入力信号 $x(t)$ に基づき判定を行っているのに対し、本実施例の音声・雑音判定部308は推定雑音 $|N(f)|$ と入力信号 $|X(f)|$ に基づき判定を行っている。推定雑音と入力信号との比 S/NR は、

【0068】

【数 1 4】

$$SNR = \frac{\sum_{f=0}^{M-1} |X(f)|^2}{\sum_{f=0}^{M-1} |N(f)|^2}$$

【0069】

となる。本実施形態ではこの値を重み係数の切替えに用いている。SNRは全帯域でなく、音声のパワーが集中している帯域のみで算出するようにしてもよい。

10

【実施例 4】

【0070】

図5に本発明の第4の実施形態に係る雑音抑圧装置の構成を表すブロック図をしめす。第1の実施形態の雑音レベル補正用信号生成部106が入力信号から補正用の信号を生成しているのに対し、本実施形態の雑音レベル補正用信号生成部406は予め保持している重置用信号450から生成している。雑音区間を入力信号とは無関係に白色雑音や聴覚的に聞えの良い雑音にしたい場合など効果的である。

【実施例 5】

【0071】

図6は本発明の第5の実施形態に係る雑音抑圧装置の構成を表すブロック図である。本実施形態は第2の実施形態に対して、N個の入力端子501-1~501-Nと、これを周波数領域に変換する周波数変換部502とその出力を統合して1つの信号を出力する統合信号生成部512と、N個の入力信号から音声・雑音判定508を行う音声・雑音判定部508を備えている点異なる。

20

【0072】

マイクロホンアレーなど複数のマイクを用いて特定の方向の音だけを強調する方法がある。この場合は入力信号が音声か雑音かという問題は特定の方向から到来している信号か否かという問題に置き換えることができる。音声・雑音判定部では複数の入力信号から信号の到来方向をもとに音声か雑音かの判定を下す。例えば、図7のようにマイク2本で正面から到来する信号を音声信号と見なす場合は、受信信号を $X_0(f)$ 、 $X_1(f)$ とし

30

【0073】

て、

【0074】

【数 1 5】

$$Ph = (1/M) \sum_{f=0}^{M-1} \left| \arg(X_0(f)X_1^*(f)) \right|$$

【0075】

を指標として音声区間を検出することができる。

40

ここで、 $X_1^*(f)$ は $X_1(f)$ の共役複素数で \arg は位相を取出す演算子、Mは周波数の成分数である。正面からの信号は2つのマイクに同じ位相で到来するため、片方を共役複素数にして互いに掛け合わせると、位相項はゼロになる。従って、(数15)は理想的に正面から到達した信号に関しては、最小値 $Ph = 0$ となる。それ以外の方向に関して正面からずれるに従い値が増加するので、適当な閾値をもとに音声/雑音の区別を行うことができる。なお、マイクの本数が2本以上の場合は、例えば全てのマイクの組み合わせに対して(数15)を計算するなどの方法がある。

【0076】

信号統合部512では、複数の入力信号から一つの信号を生成する。例えば遅延和アレー

50

と呼ばれる方法では、入力信号の加算を行う。具体的には、統合された信号 $X(f)$ は入力信号 $X_1(f) \sim X_N(f)$ を用いて、

【0077】

【数16】

$$X(f) = 1/N \sum_{i=0}^{N-1} X_i(f)$$

【0078】

と表される。ここで N はマイクの本数である。

10

このようにすることで正面から入力された目的信号は同位相であるため強調され、その他の方向から入力された信号は位相がずれているため弱め合い、その結果目的信号が強調され雑音が抑圧されるので、後段のスペクトルサブトラクションの雑音抑圧効果との相乗効果で、1つのマイクの場合に比べてより高い雑音抑圧性能を実現することができる。

【0079】

また複数のマイクを使って音声区間の検出を行うので、1つのマイクの場合よりも高い検出能力が実現可能である。例えば、横方向からの妨害音が存在する場合、1つのマイクではこれを音声と区別することは困難であるが、複数マイクであれば(数15)のように、位相成分を利用して音声信号(正面からの信号)と区別することができる。

【0080】

20

なお、周波数変換部502の後に統合信号生成部512が構成されているが、周波数変換部502と統合信号生成部512は逆順であってもよい。

【実施例6】

【0081】

図8は本発明の第6の実施形態に係る雑音抑圧装置の構成を表すブロック図である。第6の実施形態は第5の実施形態における統合信号生成部612が目的信号強調部630と目的信号除去部631から構成されている。目的信号強調部630は第5の実施形態と同様に予め設定された目的音方向からの信号(例えば正面)の信号を強調するのに対し、目的信号除去部631は目的信号強調部630の目的音方向とは異なる方向(例えば横方向)を目的音方向とする。その結果、目的信号除去部631では正面から到来する音声信号は弱められ、周囲の音が強調されることになる。このように特定の方向に指向性を形成するユニットはビームフォーマと呼ばれることがある。第5の実施形態で説明した遅延和アレーもビームフォーマの1つである。

30

【0082】

本実施形態においては目的信号強調部630と目的信号除去部631を適応形アレーの代表であるGriffith-Jim形のビームフォーマを用いて実現する構成について説明する。

【0083】

図9にGriffith-Jim形のビームフォーマの一構成例を示す。ビームフォーマの出力 $X(f)$ は、入力信号 $X_0(f)$ 、 $X_1(f)$ と適応フィルタを用いて求められる。入力端子901,902に $X_0(f)$ 、 $X_1(f)$ がそれぞれ入力される。整相化部903では目的音方向の信号の位相が同位相になるように位相が調整される。その出力は加算部904で加算され、減算部905で減算される。この減算により目的音が消去されるため、残りの信号を適応フィルタ906の入力として加算器904の出力から引き去ることで雑音が除去された信号 $X(f)$ が得られる。

40

【0084】

Griffith-Jim形のビームフォーマは妨害音の方向に感度が急峻に落ちる谷状のノッチを作ることが可能であり、この特性は特に目的音信号除去部631が正面からの音声を妨害音とみなして除去するのに適した性質である。

【0085】

さらに、目的音信号除去部631の出力信号は雑音推定部603の入力信号としても用いる。

50

雑音推定部603は自力で $X(f)$ を観測し音声のない区間を見つけこれを平滑化して推定雑音を生成したが、目的信号除去部631の出力は常に雑音のみであるため、雑音の推定に利用される。このことから、これら2つの信号を利用することでより高精度な雑音推定が可能となる。

【実施例7】

【0086】

図10は本発明の第7の実施形態に係る雑音抑圧装置の構成を表すブロック図である。本実施形態は第5の実施形態における統合信号生成部512の出力 $X(f)$ を帯域分割部740でサブバンドに周波数分割し、各サブバンド毎に雑音抑圧を行う。雑音抑圧はこれまでの実施形態と同様であるが、音声・雑音判定部708は各サブバンド毎に判定を行う。

10

【0087】

音声のスペクトルを周波数方向に眺めると、振幅の出ている区間とそうでない区間が混在しており、山と谷がある。谷の部分の周波数に関しては、雑音区間と考えることができ、雑音レベルの推定や、過剰抑圧といった雑音区間で行う処理が用いられる。サブバンドに分割し各サブバンド毎の雑音/音声の判定に基づいて雑音抑圧を切替えることで、音声区間の品質をより高めることができる。

【0088】

本実施例では複数の入力信号から統合信号を生成後にサブバンドに分割しているが、入力信号を先にサブバンドに分割した後にサブバンド単位で統合信号を求める構成であってもよい。

20

【図面の簡単な説明】

【0089】

【図1】第1の実施形態に係る雑音抑圧装置の構成を表すブロック図。

【図2】入力信号のフレーム毎の振幅を模式した図。

【図3】第2の実施形態に係る雑音抑圧装置の構成を表すブロック図。

【図4】第3の実施形態に係る雑音抑圧装置の構成を表すブロック図。

【図5】第4の実施形態に係る雑音抑圧装置の構成を表すブロック図。

【図6】第5の実施形態に係る雑音抑圧装置の構成を表すブロック図。

【図7】マイクロホンアレイの機能を示す図。

【図8】第6の実施形態に係る雑音抑圧装置の構成を表すブロック図。

30

【図9】Griffith-Jim形のビームフォーマの一構成例を表すブロック図。

【図10】第7の実施形態に係る雑音抑圧装置の構成を表すブロック図。

【符号の説明】

【0090】

101,201,301,401,501-1...501-N, 601-1...601-N, 701-1...701-N,901,902 入力端子

102,202,302,402,502,602,702 周波数変換部

103,203,303,403,503,603 雑音推定部

104,404 雑音抑圧部

204,304,504,604 抑圧係数算出部

105,405 雑音過剰抑圧部

40

205,305,505,605 過剰抑圧係数算出部

106,406 雑音レベル補正用信号生成部

206,306,506,606 補正レベル補正用係数生成部

107,207,307,407,507,607,707,904,905,907 加算部

108,208,308,408,508,608,708 音声・雑音判定部

109,209,309,409,509,609 切替部

110,210,310,410,510,610,710 周波数逆変換部

211,311,511,611 乗算器

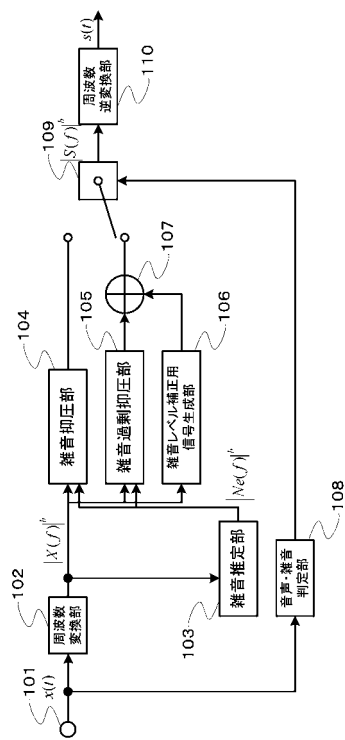
512,612,712 総合信号生成部

450 重畳用信号保存部

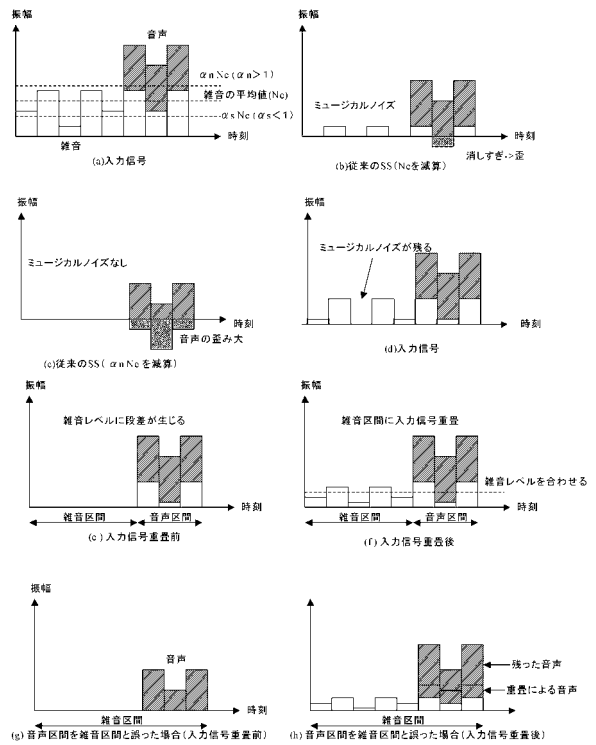
50

- 630 目的信号強調部
- 631 目的信号除去部
- 740 帯域分割部
- 750 サブバンド雑音抑圧部
- 760 帯域統合部
- 903 整相化部
- 906 適応フィルタ

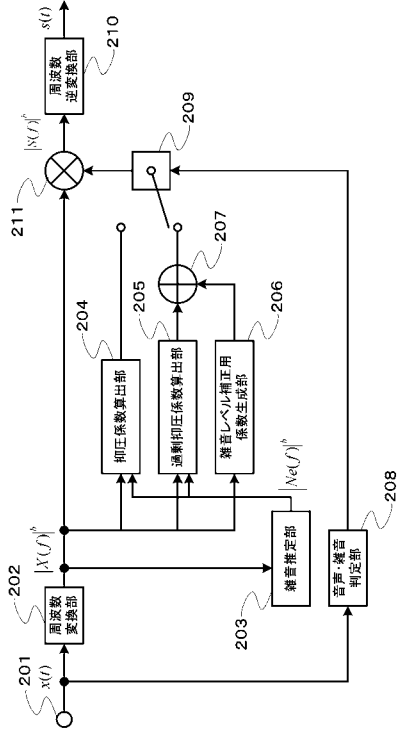
【図1】



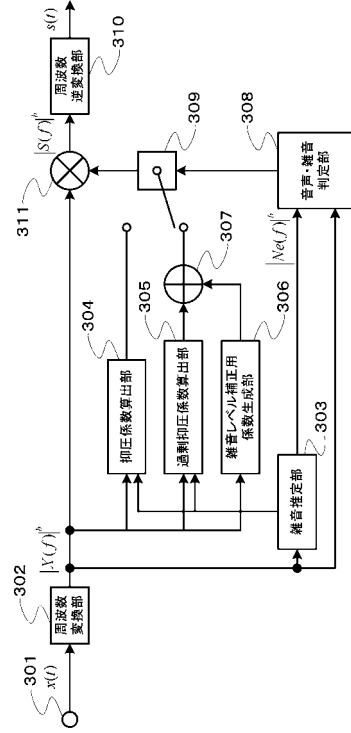
【図2】



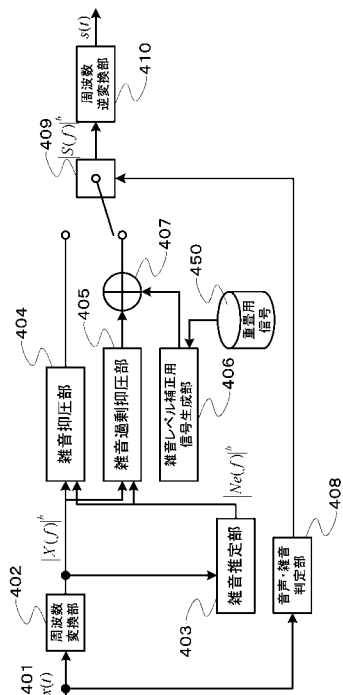
【図 3】



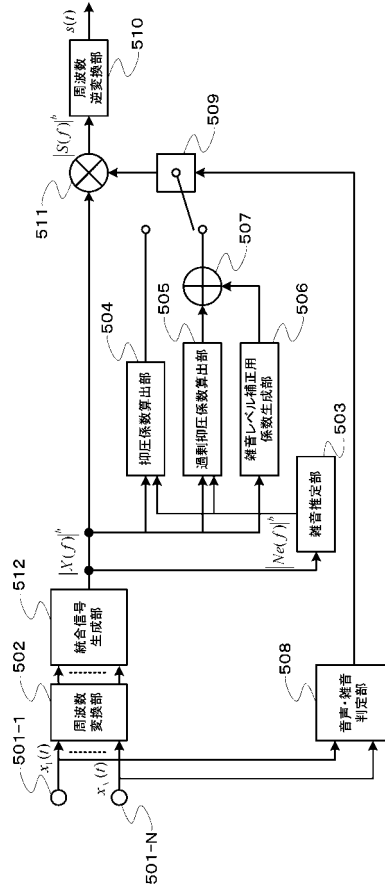
【図 4】



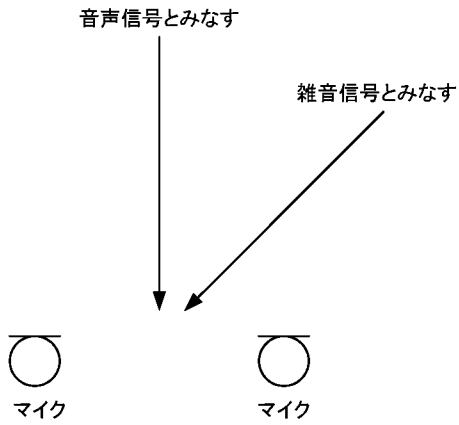
【図 5】



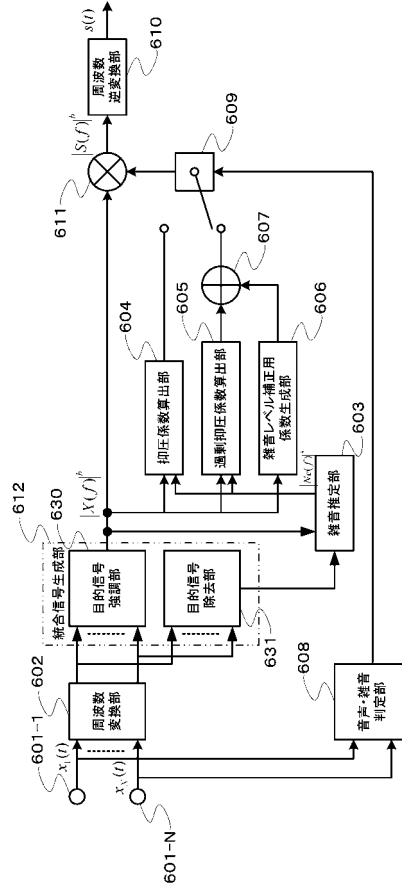
【図 6】



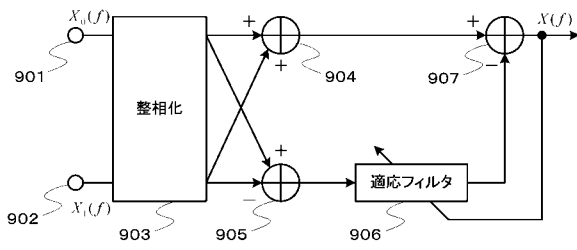
【図 7】



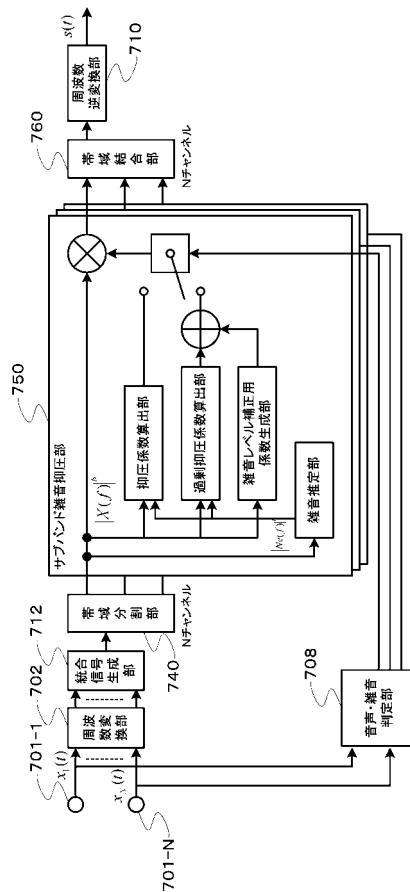
【図 8】



【図 9】



【図 10】



フロントページの続き

審査官 山下 剛史

- (56)参考文献 特開平08 - 221092 (JP, A)
特開平08 - 022297 (JP, A)
国際公開第99 / 050825 (WO, A1)
特開平08 - 167879 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 11 / 02 , 15 / 04 , 15 / 20 , 21 / 02
H04M 1 / 60
H04R 3 / 00