

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2013年8月22日(22.08.2013)



(10) 国際公開番号
WO 2013/122183 A1

- (51) 国際特許分類:
G10L 21/0208 (2013.01) H04B 3/23 (2006.01)
G10L 21/0216 (2013.01) H04R 3/02 (2006.01)
- (21) 国際出願番号: PCT/JP2013/053645
- (22) 国際出願日: 2013年2月15日(15.02.2013)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2012-033159 2012年2月17日(17.02.2012) JP
- (71) 出願人: 株式会社日立製作所(HITACHI, LTD.)
[JP/JP]; 〒1008280 東京都千代田区丸の内一丁目
6番6号 Tokyo (JP).
- (72) 発明者: 戸上 真人(TOGAMI Masahito); 〒1858601
東京都国分寺市東恋ヶ窪一丁目280番地 株
式会社日立製作所 中央研究所内 Tokyo (JP). 川
口 洋平(KAWAGUCHI Yohei); 〒1858601 東京都
国分寺市東恋ヶ窪一丁目280番地 株式会
社日立製作所 中央研究所内 Tokyo (JP).

- (74) 代理人: 平木 祐輔, 外(HIRAKI Yusuke et al.); 〒
1056232 東京都港区愛宕2丁目5番1号 愛宕
グリーンヒルズMORIタワー32階 Tokyo
(JP).
- (81) 指定国(表示のない限り、全ての種類の国内保
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,
BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,
CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES,
FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN,
IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS,
LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH,
PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保
護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW,
MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシ
ア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ
(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR,
GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT,

[続葉有]

(54) Title: DEREVERBERATION PARAMETER ESTIMATION DEVICE AND METHOD, DEREVERBERATION/ECHO-CANCELLATION PARAMETER ESTIMATION DEVICE, DEREVERBERATION DEVICE, DEREVERBERATION/ECHO-CANCELLATION DEVICE, AND DEREVERBERATION DEVICE ONLINE CONFERENCING SYSTEM

(54) 発明の名称: 残響除去パラメータ推定装置及び方法、残響・エコー除去パラメータ推定装置、残響除去装置、残響・エコー除去装置、並びに、残響除去装置オンライン会議システム

図 1

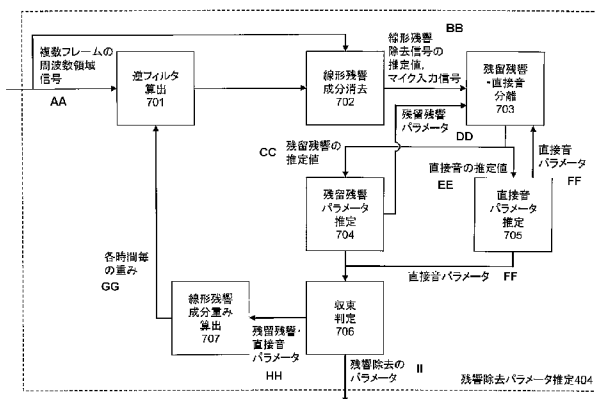


FIG. 11:
404 Dereverberation parameter estimation
701 Inverse filter calculation
702 Linear reverberation components removal
703 Residual reverberation and direct sound separation
704 Residual reverberation parameter estimation
705 Direct sound parameter estimation
706 Convergence evaluation
707 Linear reverberation component weighting calculation
AA Frequency domain signals of multiple frames
BB Estimation value for linear reverberation removal signal, microphone input signal
CC Estimation value for residual reverberation
DD Residual reverberation parameter
EE Estimation value for direct sound
FF Direct sound parameter
GG Weighting per time
HH Residual reverberation / direct-sound parameter
II Residual reverberation parameter

(57) Abstract: In conventional dereverberation schemes, it had been difficult to determine with precision a linear reverberation removal filter for removing non-fluctuating reverberation components when fluctuating reverberation components are present. The present invention provides an integrated algorithm for a dereverberation scheme using a linear filter and a dereverberation scheme using a nonlinear filter, and the algorithm has a function for measuring the amount of fluctuation in a transfer function per time for the nonlinear filter of the subsequent stage and controls the intensity of the nonlinear filter per time on the basis of that function. This allows for a configuration such that intensive nonlinear processing is applied only when the fluctuation of the transfer function is large; therefore, distortion in the speech component can be kept to a minimum. Further, the amount of fluctuation in the transfer function per time according to the abovementioned function is fed back to the estimation of the linear filter of the preceding stage, and the weighting for linear filter estimation values in the time periods where the transfer function fluctuates is lowered, thereby making it possible to mitigate the influence of a decrease in the precision of transfer function estimation.

(57) 要約:

[続葉有]



NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI 添付公開書類:
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, — 国際調査報告 (条約第 21 条(3))
NE, SN, TD, TG).

従来の残響除去方式では、変動性残響成分が存在する際には、非変動性残響成分を除去するための線形の残響除去フィルタを高精度に求めることが困難であった。線形フィルタを用いた残響除去方式と非線形フィルタを用いた残響除去方式の統合アルゴリズムであって、後段の非線形フィルタにおいて伝達関数の揺らぎの量を時間毎に計測する機能を有し、その機能を基に時間毎の非線形フィルタの強さを制御する。これにより、伝達関数の揺らぎが大きい時だけ、強い非線形処理を掛けるような構成とすることになるため、音声成分の歪みを最小限に小さく保つことができる。また、同機能による時間毎の伝達関数の揺らぎの量を前段の線形フィルタの推定にフィードバックさせて、線形フィルタ推定値に伝達関数が揺らぐような時間帯の重みを下げることで、伝達関数の推定精度が落ちる影響を軽減させることが可能である (図 11 参照)。

明 細 書

発明の名称：

残響除去パラメータ推定装置及び方法、残響・エコー除去パラメータ推定装置、残響除去装置、残響・エコー除去装置、並びに、残響除去装置オンライン会議システム

技術分野

[0001] 本発明は、マイクロホンに入った音の中から、所望の音だけをピックアップする雑音除去及び／又は残響成分の除去技術に関する技術分野に属する。

背景技術

[0002] マイクロホンに含まれる音の中から、部屋の壁や天井などで音が反射した結果生じる残響成分を高精度に除去し、話者の口元から直接マイクに届く成分（直接音）成分のみを高精度にピックアップするための残響除去技術が検討されている（例えば非特許文献1参照）。

[0003] 図1は、室内における音の伝搬過程の模式図を示している。図1に示されるように、話者の口元から発せられ、マイクで受音される成分は、話者の口元からマイクロホンまで最短経路を辿って伝搬する成分である直接音成分（図1の(i)参照）と、壁や床、また室内の人に反射し跳ね返った音がマイクに混入する成分である残響成分と、に大きく分けることができる。

[0004] これまでの残響除去技術は、マイクに入った音を、伝搬特性の逆フィルタを線形フィルタとして推定し、推定した線形フィルタをマイク入力信号に重畳している。これにより、マイク入力信号を直接音成分と残響成分の二つに分けて、残響の無い直接音成分のみを抽出する技術と捉える事ができる。従来は、この残響成分の伝搬経路が時間変動しない非変動性残響成分（図1の(ii)参照）として捉えてきた。

[0005] しかし、実際には、非変動性残響成分の他に、室内で動く人（頭等が動く）などに反射した後、マイクに伝わる成分である変動性残響成分（図1の(iii)参照）も存在する。この変動性残響成分は、線形フィルタでは十分消去

しきれない成分と捉えることができる。このような変動性残響成分については、非線形フィルタによって除去することが提案されている（例えば特許文献 1 参照）。

先行技術文献

特許文献

[0006] 特許文献1：特許第4 1 0 7 6 1 3号公報

非特許文献

[0007] 非特許文献1：T. Yoshioka et al., “Blind separation and dereverberation of speech mixtures by joint optimization,” IEEE Trans. ASLP, vol. 19, no. 1, pp. 69-84, Jan. 2011

非特許文献2：K. Kinoshita et al., “Suppression of Late Reverberation Effect on Speech Signal Using Long-Term Multiple-step Linear Prediction,” IEEE Trans. ASLP, vol. 17, no. 4, pp. 534-545, 2009

発明の概要

発明が解決しようとする課題

[0008] 従来の線形フィルタを用いた残響除去方式（例えば非特許文献 1 参照）では、話者の口元からマイクまで音が伝達する時の伝達関数は変化しないという前提で残響成分を除去する。

[0009] しかし、例えば会議シーン等においては、話者の顔の向きや話者以外の人（頭等）が動くなどの影響を受けて、伝達関数が時間毎に変化する場合が多い。このような場合、まず、線形フィルタを推定するのに用いる信号の中に、伝達関数が揺らぐような時間帯が含まれていると、伝達関数の推定精度が落ちてしまい、結果として残響除去性能が落ちるという課題がある（課題 1）。また、たとえ線形フィルタが高精度に求まったとしても、線形フィルタでは伝達関数が変化しないような場合以外は残響を落とすことができず、伝達関数が揺らぐような時間帯における残響抑圧性能が悪いという課題がある（課題 2）。

[0010] 一方、非線形フィルタを用いた非線形の残響除去技術では、残響成分を落とす量を制御するパラメータを大きくすることで残響除去量を増やすことが可能である。

[0011] しかし、パラメータを適切に設定できないと本来取りたい音声成分を歪ませる量を増やしてしまうことになり、結果的に取りたい音の抽出精度が下がることになる。

[0012] 本発明はこのような状況に鑑みてなされたものであり、マイク入力信号から非変動性及び変動性の残響成分を精度良く除去して、直接音の抽出精度を高めるためのパラメータを推定する技術及びそのパラメータを用いて残響成分をマイク入力信号から除去する技術を提供するものである。

課題を解決するための手段

[0013] (i) 上記課題を解決するため、本発明は、線形フィルタを用いた残響除去方式と非線形フィルタを用いた残響除去方式の統合アルゴリズムを提案する。より具体的には、後段の非線形フィルタにおいて伝達関数の揺らぎの量を時間毎に計測する機能を有し、その機能を基に時間毎の非線形フィルタの強さを制御（パラメータの生成を制御）する。これにより、伝達関数の揺らぎが大きい時だけ、強い非線形処理を掛けるような構成とすることになるため、音声成分の歪みを最小限に小さく保つことができる。また、非線形処理によって得られる変動性残響成分の推定値を、線形フィルタのパラメータ生成処理にフィードバックして非変動性残響成分除去の精度をさらに向上させる。

[0014] (ii) 図2は、マイクロホン内に含まれる非変動性残響成分と変動性残響成分を除去する方法を模式的に示す図である。図2に示されるように、伝搬過程が変動しない非変動性残響成分は、伝搬過程の逆フィルタに相当する線形フィルタにより抑圧される。一方、所望の直接音及び変動性残響成分は線形フィルタでは抑圧されずに残る。その後、空間／非線形のフィルタリングを施すことにより、線形フィルタでは抑圧しきれなかった変動性残響成分が抑圧され、最終的に所望の直接音のみを得ることができる。

[0015] つまり、本発明を全体としてみると、伝搬過程が揺らがない残響成分だけ

でなく、伝搬過程が揺らぐ残響成分も落とすことが可能な構成となっている（課題 1 の解決）。つまり、本発明では、線形フィルタによって音声入力信号に含まれる非変動性残響成分を除去して線形残響除去信号の推定値を生成し、非線形フィルタを用いて、線形残響除去信号の推定値に含まれる変動性残響成分及び直接音成分の推定値を生成する。そして、変動性残響成分及び直接音の推定値に基づいて、非線形フィルタのパラメータを構成する変動残響成分及び直接音成分のパラメータを更新する。さらに、更新された変動性残響成分及び直接音成分のパラメータに基づいて、線形フィルタのパラメータを逐次的に更新するようにしている。

[0016] (iii) また、同機能による時間毎の伝達関数の揺らぎの量を前段の線形フィルタの推定にフィードバックさせて、線形フィルタ推定値に伝達関数が揺らぐような時間帯の重みを下げることで、伝達関数の推定精度が落ちる影響を軽減させることが可能である（課題 2 の解決）。

[0017] 図 3 は、時間毎の重みの違いを模式的に示す図である。線形フィルタを学習する際には、できるだけ線形フィルタで抑圧すべき対象である非変動性残響成分が大きく存在する時間帯のみをピックアップして学習すべきといえる。そこで、本発明の重み決定プロセスでは、直接音成分と変動性成分が少ない時間帯を選ぶことで、学習対象の非変動性残響成分が大きく存在している時間帯だけを抽出することが可能となっている。

[0018] 本発明に関連する更なる特徴は、本明細書の記述、添付図面から明らかになるものである。また、本発明の態様は、要素及び多様な要素の組み合わせ及び以降の詳細な記述と添付される特許請求の範囲の様態により達成され実現される。

[0019] 本明細書の記述は典型的な例示に過ぎず、本発明の特許請求の範囲又は適用例を如何なる意味に於いても限定するものではないことを理解する必要がある。

発明の効果

[0020] 本発明により広い部屋同士をつなぐビデオ会議システムにおいて、残響の

影響が少ないクリアな音声で快適な音声通話を可能にする残響除去パラメータを推定することが可能となる。

図面の簡単な説明

[0021] [図1]音の伝搬過程を説明するための模式図である。

[図2]マイクロホン内に含まれる非変動性残響成分と変動性残響成分の除去過程（保温発明の基本概念）を説明するための模式図である。

[図3]時間毎の重みの大きさの違いについて説明するための模式図である。

[図4]本発明による拠点毎会議システムの概略構成を示す図である。

[図5]本発明による遠隔会議システムの全体構成を示す図である。

[図6]中央演算装置102内で実行する残響除去プログラムのブロック構成を示す図である。

[図7]残響除去302の効果の一例を示す図である。

[図8]残響除去302の詳細なブロック構成（フローチャート相当）を示す図である。

[図9]各マイクロホンのデータ取得について説明するための図である。

[図10]残響除去パラメータ推定404と、オンライン残響除去405の実行タイミングについてのタイミングチャートである。

[図11]残響除去パラメータ推定404の詳細なブロック構成（フローチャート相当）を示す図である。

[図12]逆フィルタ算出701の詳細なブロック構成（フローチャート相当）を示す図である。

[図13]マイクチャンネル毎に残響除去を行う構成を示す概念図である。

[図14]残留残響・直接音分離703の具体的なブロック構成（フローチャート相当）を示す図である。

[図15]残留残響パラメータ推定704の詳細なブロック構成（フローチャート相当）を示す図である。

[図16]直接音パラメータ推定705の詳細なブロック構成（フローチャート相当）を示す図である。

[図17]オンライン残響除去405の具体的な構成（フローチャート相当）を示す図である。

[図18]本発明の第2の実施形態による残響除去処理の構成を示す図である。

[図19]残響除去とエコーキャンセラの双方の性能を向上させることが可能な構成を示す図である。

[図20]残響除去及びエコーキャンセラ2001の具体的な構成を示す図である。

[図21]オフラインパラメータ推定1800の詳細なブロック構成を示す図である。

[図22]残響・音響エコー消去用パラメータ推定1801の詳細なブロック構成（フローチャート相当）を示す図である。

[図23]残響除去とエコーキャンセラを同時実行する構成における重み係数の決定（ $R_{x, f, \tau(i)}$ の影響）について説明するための図である。

[図24]オンライン音響エコー・残響除去2301の詳細な構成（フローチャート相当）を示す図である。

[図25]分散構成による残響除去のブロック構成を示す図である。

発明を実施するための形態

[0022] 本発明は、広い部屋で使われる遠隔会議システムなどで使用されることを想定し、複数設けられたマイクに対する入力信号から残響雑音（非変動及び変動性の残響成分）を取り除き、口元で直接マイクによって音を取ったようにする（直接音採取）ための技術を提供する。第1の実施形態では、残響除去パラメータの推定及びそれを用いたリアルタイム残響除去処理について示されている。また、第2の実施形態では、残響除去処理において、残響除去パラメータ推定処理で求めた過去の残響除去パラメータを複数組持っておき、時間毎に最も良いフィルタを選択して用いる処理について示されている。さらに、第3の実施形態では、残響及びエコー（反響）を除去するパラメータの推定及びそれを用いたリアルタイム残響及びエコー除去処理について示されている。また、第4の実施形態では、残響除去パラメータ推定処理をサ

ーバ側で実行する分散処理について示されている。

[0023] 以下、添付図面を参照して本発明の実施形態について説明する。添付図面では、機能的に同じ要素は同じ番号で表示される場合もある。なお、添付図面は本発明の原理に則った具体的な実施形態と実装例を示しているが、これらは本発明の理解のためのものであり、決して本発明を限定的に解釈するために用いられるものではない。

[0024] 本実施形態では、当業者が本発明を実施するのに十分詳細にその説明がなされているが、他の実装・形態も可能で、本発明の技術的思想の範囲と精神を逸脱することなく構成・構造の変更や多様な要素の置き換えが可能であることを理解する必要がある。従って、以降の記述をこれに限定して解釈してはならない。

[0025] 更に、本発明の実施形態は、汎用コンピュータ上で稼動するソフトウェアで実装しても良いし専用ハードウェア又はソフトウェアとハードウェアの組み合わせで実装しても良い。

[0026] (1) 第1の実施形態

＜各拠点のシステム構成＞

図4は、本発明の第1の実施形態による遠隔会議の各拠点に設置される会議システムのハードウェア概略構成を示す図である。各拠点の会議システム100では、マイクロホンアレイ105が、各会議室の中の音声波形を集音する。マイクロホンアレイ105は、単一のマイクロホンまたは複数のマイクロホン素子からなることを想定する。

[0027] 集音したアナログの音声波形は、A/D変換機104でアナログ信号からデジタル信号に変換される。変換されたデジタル音声波形は、中央演算装置102（プロセッサとも言う）で残響除去処理が施された後、HUB108を介してパケットに変換されネットワークに放出される。

[0028] 中央演算装置102は、不揮発性メモリ101に記憶されているプログラム、及びプログラムで用いるパラメータを読み込み、該プログラムを実行する。また、プログラム実行時に用いるワークメモリは、揮発性メモリ103

上に確保され、残響除去に必要な各種パラメータの記憶領域が定義される。残響除去パラメータは、中央処理装置 102 によって推定され、それが揮発性メモリ 103 に格納される。そして、格納された残響パラメータが再度中央処理装置 102 によって読み出され、新たな推定処理に用いられる。

[0029] 中央演算装置 102 は、遠隔会議における、他拠点（遠端）の音声波形を、ネットワーク経由で、HUB 108 から受け取る。受け取った遠端音声波形（デジタル音声波形）は、中央演算装置 102 経由で、D/A 変換機 106 に送られて、デジタル信号からアナログ信号に変換された後、変換されたアナログの音声波形は、スピーカアレイ 107 から放出される。

[0030] スピーカアレイ 107 は、単一のスピーカ素子または複数のスピーカ素子で構成される。また、各拠点の映像情報は、一般的なカメラ 109 で撮像され、HUB 108 を経由して他拠点に送信される。他拠点の映像情報は、ネットワーク経由で HUB 108 に送られ、更に中央演算装置 102 を経由して、各拠点に設置されたディスプレイ 110 上で表示される。なお、カメラ 109 を複数台設置したり、ディスプレイ 110 を複数台設置するような構成を、取っても良い。

[0031] <遠隔会議システムの全体構成>

図 5 は、本発明の遠隔会議システムの全体構成例を示す図である。遠隔会議システムは、N 個の各拠点会議システム 100-1、100-2、・・・100-N（拠点数を N とする）と、各拠点の音声や映像の流れを制御する MCU 202 と、各拠点の計算機では処理できないような、処理量の大きい計算を実行する会議情報計算サーバ 201 と、を有し、それぞれネットワークを介して接続されている。なお、当該システムは、当業者であれば既知のシステムであるため、詳細な説明は割愛する。また、会議情報計算サーバ 201 において、各拠点の会議システムで実行される残響除去に必要な一部の処理を実行するような構成を採っても良い。

[0032] <残響除去処理>

図 6 は、本発明の第 1 の実施形態において、中央演算装置 102 内で実行

される処理（プログラム）のブロック構成を示す図である。マイクロホンアレイ 105 から A/D 変換機 104 経由で得られたデジタル音声波形は、エコーキャンセラ 301 で処理され、音響エコー成分が除去される。ここで、音響エコー成分とは、スピーカアレイ 107 から出力された音声波形が各拠点の壁や天井などで反射した後、マイクロホンアレイ 105 に混入する成分を指す。エコーキャンセラ 301 は、音響エコー成分を除去するために HUB 108 経由で得られる NLMS 法など当業者であれば既知の構成により実現される。スピーカアレイ 107 から出力する信号として、遠端音声波形が用いられる。

[0033] エコーキャンセル後の音声波形は、残響除去 302 に送られ、残響成分が除去され、残響成分を除去した後の音声波形が出力される。音声波形は、時系列信号であるため、残響除去プログラムは A/D 変換後の音声波形が一定量蓄積される毎に実行される。

[0034] <残響除去の効果>

図 7 は、残響除去 302 の効果（一例）を説明するための図である。マイク入力信号（図 7 の (i) 参照）は、マイクロホンアレイ 105 を構成するある一つのマイクロホンの入力波形を時間一周波数領域に変換したスペクトログラムを示している。横軸が時間軸で、縦軸が周波数を示している。図 7 において、明るい時間一周波数成分ほど、音量が大きいことを意味する。

[0035] 残響の無い信号（理想）（図 7 の (ii) 参照）は、部屋の残響が存在しない直接音のみの波形を示している。マイク入力信号と残響の無い信号（理想）を比較すると、前者は、時間方向にインクがにじんだように成分が拡散している様子が見て取れる。本成分が、本発明で捉える残響成分である。

[0036] 残響除去後の信号（図 7 の (iii) 参照）は、本発明の残響除去 302 で処理した後の波形である。残響成分が除去され、残響の無い信号（理想）に近付いている様子が分かる。

[0037] <残響除去処理>

図 8 は、第 1 の実施形態による残響除去（処理）302（図 6 参照）の詳細

細なブロック構成を示す図である。

[0038] (i) フレーム化処理

図8に示されるように、エコーキャンセル後のデジタル音声波形は、フレーム化（処理）401に送られる。ここで、デジタル音声波形を $q(m, t)$ と記載し、 m を、マイクロホンアレイ105を構成するマイクロホンのインデックス、 t をA/D変換処理のサンプルインデックスとする。

[0039] フレーム化401では、マイクロホン素子毎のデジタル音声波形が一定量蓄積されるたびに、フレーム単位の音声波形を出力する。一方、一定量蓄積されるまでは、何も出力されない。該一定量をフレームシフトと呼び、 S （ポイント）と記載する。このようなフレームシフトを行うのは、音声は徐々に周波数成分が変化して伝わるので、その遷移を正確に捉えるためである。また、フレーム化401で出力される各マイクロホン素子毎の音声波形は、フレームシフトよりも大きい、 P ポイント分の音声波形とする。

[0040] フレームの単位をフレームインデックスとよび、 τ と記載する。そして、 M 番目のマイクロホン素子におけるフレームインデックス τ の出力信号は、 $t = S\tau$ から $t = S\tau + P - 1$ までのポイントの音声波形となり、これを式（1）で定義する。

[数1]

$$\mathbf{k}(m, \tau) = [x(m, S\tau) \cdots x(m, S\tau + P - 1)] \quad \cdots (1)$$

[0041] 図9は、フレーム形成、及びフレームシフトの処理を説明するための図である。フレームインデックスが1の場合、「フレーム1」は、データ取得開始直後 P ポイント分のデータの音声波形で構成される。次のフレーム「フレーム2」は、フレーム1から S ポイントずらしたポイントから P ポイント分のデータの音声波形で構成される。

[0042] フレーム化401は、各マイクロホン素子のフレーム単位の音声波形 $\mathbf{k}(m, \tau)$ を出力する。

[0043] (ii) 周波数分解処理

周波数分解（処理）402は、フーリエ変換処理など、同業者であれば一般的に使われている周波数変換処理により、各マイクロホン素子の音声波形を時間周波数領域信号に変換する。時間周波数領域信号に変換した m 番目のフレーム単位の信号を $X_m(f, \tau)$ と定義する。ここで、 f は時間周波数領域における周波数インデックスとする。また、各マイクロホンの時間領域信号を時間周波数毎にまとめたベクトルを $X(f, \tau) = [X_1(f, \tau), X_2(f, \tau), \dots, X_m(f, \tau), \dots, X_M(f, \tau)]$ と記載する。 M はマイクロホンの数を意味する。そして、各マイクロホンの周波数領域信号は、バッファリング（処理）403及びオンライン残響除去（処理）405に送られる。

[0044] (iii) バッファリング処理

バッファリング（処理）403は、時間領域信号を蓄積し、蓄積量が一定量に達した場合のみ、蓄積した信号を出力し、それ以外の場合は何も出力しない。各マイクロホンに蓄積する量は、 T フレーム（例えば、300フレーム）分とする。ある程度の統計量を用いないとパラメータ推定が適切に（安定的に）行えないため、 T フレーム分の音声データを蓄積してから残響パラメータ推定処理を実行するようにする。また、例えば、会議中に話者が切り替わると、今までオンライン残響除去405で使用していた残響除去パラメータは適切なものではなくなるため、再度パラメータ推定を実行し、当該パラメータを更新する。ただし、話者の切り替わりを音声波形のみから検出するのは困難であるため、本実施形態では、 T フレームごとに残響除去パラメータを更新するようにしている。別の言い方をすれば、本実施形態では、一度パラメータを推定すると、次のパラメータ推定の処理が終了するまで、現在のパラメータを用いて残響除去処理が実行される。つまり、本実施形態の残響除去処理では、常に最新の推定パラメータを用いている。なお、話者が切り替わったことが検知できれば、話者切り替わりのタイミングで残響除去パラメータを更新するようにしても良い。

[0045] (iv) 残響除去パラメータ推定処理

残響除去パラメータ推定（処理）４０４は、バッファリング４０３が出力するＴフレーム分のデータを基に、残響除去を行うためのパラメータを推定し、推定したパラメータを出力する。残響除去パラメータ推定（処理）４０４の更なる詳細については、図１１を参照して後述する。

[0046] (v) オンライン残響除去処理

オンライン残響除去（処理）４０５は、推定された残響除去パラメータをリアルタイムで活用する。残響除去パラメータ推定４０４は、Ｔフレーム分のデータが溜まる毎に処理を実施するのに対して、オンライン残響除去４０５は、リアルタイムに残響除去を行うことが必要であるため、１フレーム分のデータ毎に処理を実施する。オンライン残響除去４０５は、残響が含まれた１フレーム分のデータ中の残響成分を除去した後の信号を出力する。

[0047] 図１０は、残響除去パラメータ推定４０４と、オンライン残響除去４０５の実行タイミングを説明するためのタイミングチャートである。残響除去パラメータ推定４０４では、上述のように、Ｔフレーム分のデータを得るたびにパラメータ推定処理が実行される。推定したパラメータがオンライン残響除去４０５で使われ始めるまでの各処理遅延フレーム長（Ｕフレーム）は、残響除去パラメータ推定４０４での処理量に依存し、異なる。

[0048] オンライン残響除去４０５は、各フレームの時間領域信号を、処理時点で得られている最新の残響除去パラメータを用いて残響除去実施する。

[0049] このような構成を採ることで、残響除去パラメータの推定が遅延した場合であっても、残響除去をリアルタイムに実行することが可能となる。

[0050] (vi) 時間領域変換処理

図８に戻り、時間領域変換（処理）４０６は、マイクロホン毎に１フレーム分の周波数領域信号に対して逆フーリエ変換などの周波数領域から時間領域への変換処理を実行することにより、時間領域信号に戻し、戻した時間領域信号を残響除去後の音声波形として出力する。この時間領域変換処理を実行することにより、各フレームにおける、（Ｐ－Ｓ）ポイント分の重複領域の音声パワーを調整することができる。

[0051] <残響除去推定パラメータ処理の詳細>

図 1 1 は、残響除去パラメータ推定（処理）4 0 4 の詳細ブロック構成を示す図である。残響除去パラメータ推定 4 0 4 は、繰り返し計算により、残響除去性能を逐次的に高めるような構成になっている。つまり、本実施形態において、2 種類のパラメータ X 及び Y（ここでは、X を非変動性残響除去パラメータ、Y を変動性残響除去パラメータとする）を求める場合、X をまず固定して、Y を変化させて Y の適正值を求める。次に、X を変化させ、X の適正值を求める。このように、図 1 1 は、X と Y を交互に求めて収束させていく処理を表している。また、残響除去パラメータ推定 4 0 4 では、周波数インデックス毎に、独立して処理が行われる。したがって、本残響除去パラメータ推定処理を周波数毎に異なる中央演算装置（プロセッサ）で実行する構成も可能である。

[0052] (i) 逆フィルタ算出処理

周波数毎に得られた、マイクロホン毎の T フレーム分のデータは、まず逆フィルタ算出（処理）7 0 1 に送られ、残響除去のための線形フィルタが算出される。

[0053] 図 1 2 は、逆フィルタ算出（処理）7 0 1 の詳細な構成を示す図である。

[0054] フィルタ計算（処理）1 4 0 3 は、線形フィルタを式（2）によって算出する。残響成分は過去の信号に由来する成分であるため、式（2）は現在の信号と過去の信号、及び過去の信号同士がどの位の相関を有しているか算出するための演算式である。現在の信号と過去の信号との相関を求めるだけでは、過去の信号を現在の信号から除去しすぎてしまう可能性があるため、過剰な信号除去を回避すべく、過去の信号同士の相関をも考慮した演算となっている。なお、相関を算出するための演算は、T フレーム分の音声信号に対して実行される。

[数2]

$$A_f = \text{ivec}(P_f^{-1} Q_f) \quad \dots (2)$$

[0055] $\tau(i)$ を、Tフレーム分のデータのうち、i番目のデータのフレームインデックスとすると、式(2)における P_f は、重み付き共分散行列であり、重み付き共分散行列計算1402において式(3)によって定義される。式(3)は、過去の信号同士の相関を算出するための演算式である。ここで、Hは、行列またはベクトルの共役転置を表す演算子とする。

[数3]

$$P_f = \left(\sum_{i=1}^T \left(U_{f,\tau(i)} U_{f,\tau(i)}^H \right) \otimes R_{x,f,\tau(i)}^{-1} \right) \cdots \quad (3)$$

[0056] さらに、式(2)における Q_f は、重み付き相関行列計算1401において、式(4)で算出される。式(4)は、現在の信号と過去の信号との相関を算出するための演算式である。

[数4]

$$Q_f = \text{vec} \left(\sum_{i=1}^T R_{x,f,\tau(i)}^{-1} x_{f,\tau(i)} U_{f,\tau(i)}^H \right) \cdots \quad (4)$$

[0057] また、式(3)における「 $\otimes$ 」は、行列またはベクトルの転置を表す演算子を表している。なお、式(3)において、



は、クロネッカ・デルタ積を表す演算子である。

[0058] さらに、式(3)において、 $U_{f,\tau(i)}$ は、式(5)で定義される。

[数5]

$$U_{f,\tau(i)} = [x_{f,\tau(i)-D}^H \cdots x_{f,\tau(i)-L_1}^H]^H \cdots \quad (5)$$

[0059] D、L1を予め定めるパラメータとする。Dは残響成分のうち、初期残響に相当するフレーム長に設定することが望ましい。L1は、後部残響の継続フレーム長に対応するパラメータであり、後部残響が大きい環境では大きい値に設定することが望ましい。 $R_{x,f,\tau(i)}$ は、線形残響成分重み算出707

が出力する行列とする。線形残響成分重み算出（処理）707を行わない場合、もしくは、Tフレーム分のデータ毎に、初めて逆フィルタ算出701を実行する場合は、 $R_{x, f, \tau(i)}$ を単位行列に設定する。式（2）において、「 vec 」は、行列をベクトルに変換するためのオペレータとする。行列 A を vec オペレータにより、ベクトルに変換する例を式（6）に示す。

[数6]

$$vec(A) = [a_{11} \ a_{21} \cdots \ a_{12} \ a_{22} \ \cdots \ a_{1n} \ a_{2n} \ \cdots] ^T \quad \cdots (6)$$

[0060] 式（6）において、 a_{mn} は行列 A の m 行 n 列目の成分を意味する。 $ivec$ は、 vec オペレータの逆演算子とし、ベクトルを行列に変換する。行列に変換する際には、行数の任意性があるが、式（2）においては、出力される行列の行数がマイク数に一致するような行列を出力するものとする。式（2）で求めた A_f は、式（7）により、ブロック毎に区分される。

[数7]

$$A_f = [W_{f,D} \cdots W_{f,L_1}] \quad \cdots (7)$$

[0061] 逆フィルタ算出701は、 A_f を出力して処理を終了する。

[0062] (ii) 線形残響成分消去処理

線形残響成分消去（処理）702は、逆フィルタ算出701が出力する A_f を利用して非変動残響成分を除去した信号 $g_{f, \tau(i)}$ を式（8）によって取得する。

[数8]

$$g_{f, \tau(i)} = x_{f, \tau(i)} - \sum_{l=D}^{L_1} W_{f,l} x_{f, \tau(i)-l} \quad \cdots (8)$$

[0063] 線形残響成分消去（処理）702は、定性的には、マイクチャンネル毎に、別々のFIRフィルタを動作させて、チャンネル毎の残響除去信号を得るような方式と捉えることができる。なお、図13に、マイクチャンネル毎に

残響除去を行うような構成として捉えた概念が示されている。このように、本実施形態による線形残響除去処理は、Tフレーム分の時間周波数信号全てに対して実行される。そして、線形残響除去後の信号は、残留残響・直接音分離703に送られる。

[0064] (iii) 残留残響・直接音分離処理

残留残響・直接音分離（処理）703は、線形残響除去後の信号を直接音と残響音（推定値）の2つに分離する。

[0065] 図14は、残留残響・直接音分離703の詳細なブロック構成を示す図である。残留残響・直接音分離処理についても、Tフレーム分の時間周波数信号全てに対して実行する。

[0066] フレーム毎の残響除去信号の推定値は、直接音フィルタ係数推定1102で推定したフィルタ係数 $W_{n,f,\tau(i)}$ を各フレームの時間周波数信号に重畳することで、式(9)のように算出される。

[数9]

$$y_{n,f,\tau(i)} = W_{n,f,\tau(i)} g_{f,\tau(i)} \quad \dots (9)$$

[0067] ここで、nは音源のインデックスを意味する変数であり、1からNまでの整数とする。Nが音源数であり、複数の音源が存在する場合においても、Nを2以上とすることで、同時に複数音源の残響除去及び直接音の分離が可能となる。

[0068] 図14において、直接音フィルタ推定（処理）1102は、式(10)を用いて、直接音を抽出するためのフィルタ（ミラーフィルタ） $W_{n,f,\tau(i)}$ を音源毎に算出する。式(10)は、直接音のパワーが全体のパワーに占める割合（直接音のパワー／（直接音のパワー＋残留残響のパワー））を求めていることと実質的に等価である。

[数10]

$$W_{n,f,\tau(i)} = R_{s(n),f,\tau(i)} R_{x,f,\tau(i)}^{-1} \quad \dots (10)$$

[0069] ここで、 $R_{s(n), f, \tau(i)}$ は、音源毎、及びフレーム毎の共分散行列であり、時間毎の目的音分散推定（処理）1104において、式（11）を用いて算出される。

[数11]

$$\mathbf{R}_{s(n), f, \tau(i)} = v_{s(n), f, \tau(i)} \mathbf{C}_{s(n), f} \quad \dots (11)$$

[0070] ここで、 $v_{s(n), f, \tau(i)}$ 、 $\mathbf{C}_{s(n), f}$ は、 n 番目の直接音成分に関するパラメータであり、繰り返し計算の中で、逐次的に更新されるパラメータである。 $v_{s(n), f, \tau(i)}$ の初期値は1であり、 $\mathbf{C}_{s(n), f}$ は、ランダムな正定エルミート行列とする。

[0071] よって、残留残響成分は、 $W_{rev, l, m, f, \tau(i)}$ を各フレームの時間周波数信号に重畳することで、式（12）のように算出される。

[数12]

$$\mathbf{y}_{rev, l, m, f, \tau(i)} = \mathbf{W}_{rev, l, m, f, \tau(i)} \mathbf{g}_{f, \tau(i)} \quad \dots (12)$$

[0072] ここで、 l は、逆フィルタのタップインデックスに対応したインデックス、 m はマイクインデックスとなっている。つまり、逆フィルタのタップインデックス及びマイクインデックス毎に残留残響成分が算出される。残留残響フィルタ係数推定1103では、 $W_{rev, l, m, f, \tau(i)}$ を式（13）によって算出する。式（13）は、残留残響のパワーが全体のパワーに占める割合（残留残響のパワー／（直接音のパワー＋残留残響のパワー））を求めていることと実質的に等価である。

[数13]

$$\mathbf{W}_{rev, l, m, f, \tau(i)} = \mathbf{R}_{rev, l, m, f, \tau(i)} \mathbf{R}_{x, f, \tau(i)}^{-1} \quad \dots (13)$$

[0073] ここで、 $\mathbf{R}_{rev, l, m, f, \tau(i)}$ は、逆フィルタのタップインデックス毎、フレーム毎の共分散行列であって、時間毎の残留残響分散推定（処理）1105において、式（14）に従って算出される。

[数14]

$$\mathbf{R}_{rev,l,m,f,\tau(i)} = \left| \mathbf{x}_{f,\tau(i)-l}(m) \right|^2 \mathbf{C}_{rev,l,m,f} \quad \dots (14)$$

[0074] $\mathbf{x}_{f,\tau(i)}(m)$ は、周波数インデックス f 、フレームインデックス $\tau(i)$ の m 番目のマイクロホンの時間周波数領域信号である。 $\mathbf{C}_{rev,l,m,f}$ は各タップインデックス、マイクインデックス毎の残留残響成分の共分散行列であり、繰り返し計算の中で、逐次的に更新されるパラメータである。初期値は、ランダムな正定エルミート行列とする。

[0075] 以上のように、分離された残留残響の推定値、及び直接音の推定値は、残留残響パラメータ推定（処理）704 及び直接音パラメータ推定（処理）705 に、それぞれ送られる。

[0076] (iv) 残響パラメータ推定処理

残留残響パラメータ推定（処理）704 は、変動性残響成分の統計量などのパラメータを推定する。図15は、残留残響パラメータ推定（処理）704の詳細を示す図である。

[0077] 図15において、残留残響二次統計量算出（処理）1301は、残留残響・直接音分離703から送られてきた残留残響の推定値に対して式（15）を適用し、フレーム毎の残留残響成分の二次統計量（残留残響のパワー）を算出する。

[数15]

$$\mathbf{V}_{rev,l,m,f,\tau(i)} = \mathbf{y}_{rev,l,m,f,\tau(i)} \mathbf{y}_{rev,l,m,f,\tau(i)}^H + (\mathbf{I} - \mathbf{W}_{rev,l,m,f,\tau(i)}) \mathbf{R}_{rev,l,m,f,\tau(i)} \quad \dots (15)$$

[0078] $\mathbf{V}_{rev,l,m,f,\tau(i)}$ は、主軸算出（処理）1302に送られて、 $\mathbf{C}_{rev,l,m,f}$ が式（16）によって更新される。

[数16]

$$\mathbf{C}_{rev,l,m,f} = \sum_{i=1}^T \frac{1}{\left| \mathbf{x}_{f,\tau(i)}(m) \right|^2} \mathbf{V}_{rev,l,m,f,\tau(i)} \quad \dots (16)$$

[0079] (v) 直接音パラメータ推定処理

直接音パラメータ推定（処理）705は、直接音の統計量などのパラメータを推定する。図16は、直接音パラメータ推定705の詳細を示す図である。

[0080] 図16において、直接音二次統計量算出（処理）1201は、残留残響・直接音分離703から送られてきた残留残響の推定値に対して式（17）を適用し、各音源のフレーム毎の直接音の共分散行列（各マイクに入力される直接音のパワー及び各マイクに信号が到達するまでの時間差成分）を算出する。

[数17]

$$\mathbf{V}_{s(n),f,\tau(i)} = \mathbf{y}_{n,f,\tau(i)} \mathbf{y}_{n,f,\tau(i)}^H + (\mathbf{I} - \mathbf{W}_{n,f,\tau(i)}) \mathbf{R}_{s(n),f,\tau(i)} \quad \dots (17)$$

[0081] 時変パラメータ算出1202は、 $\mathbf{v}_{s(n),f,\tau(i)}$ を式（18）に従って更新する。ここで、時変パラメータとは、N個のマイクに直接音が到達するまでの時間差を含む情報を意味する。

[数18]

$$\mathbf{v}_{s(n),f,\tau(i)} = \frac{1}{M} \text{trace}(\mathbf{C}_{s(n),f}^{-1} \mathbf{V}_{s(n),f,\tau(i)}) \quad \dots (18)$$

[0082] 更に、主軸算出1203は、 $\mathbf{C}_{s(n),f}$ を式（19）に従って更新する。ここで主軸算出は、N個のマイクに直接音が到達するまでの時間差を考慮して、N次元（N個のマイク）におけるN次元多様体（楕円）の主軸（各マイクに入力される直接音のパワーのばらつき）を求めることを意味する。

[数19]

$$\mathbf{C}_{s(n),f} = \sum_{i=1}^T \frac{1}{\mathbf{v}_{s(n),f,\tau(i)}} \mathbf{V}_{s(n),f,\tau(i)} \quad \dots (19)$$

[0083] 例えば、複数のマイクが会議室内に設置されている場合、ある話者が特定方向から話していると、各フレーム時間の各マイクに入力される音声信号の

パワーは、話者に近いマイクほど大きくなる。主軸算出 1203 は、この各マイクに伝わる音声信号のパワーのばらつきを算出する処理であり、時変パラメータ算出 1202 は、各マイクに到達するまでの音声信号（直接音）の時間差を算出する処理である。より具体的に、2つのマイク m_1 及び m_2 が設置されている場合には、 m_1 及び m_2 に伝わる音声信号のパワーを $m_1 - m_2$ 平面でプロットしたときの、原点とそのプロットとを結ぶ直線の傾きが主軸を表し、その主軸長（原点から当該プロットまでの距離）が時変パラメータ（時間差）を表すことになる。

[0084] (vi) 収束判定処理

推定した直接音のパラメータと残留残響のパラメータは、収束判定（処理）706 に送られる。

[0085] 収束判定 706 は、繰り返し計算を所定回数実行したかどうか、推定したパラメータの値と推定前の値との差が所定値以下か否かなど、一般的な繰り返し計算の場合と同様の尺度で、計算が収束したかどうか判定する。収束していれば、残響除去のパラメータを出力して、残響除去パラメータ推定 404 のブロックを終了する。

[0086] 収束していなければ、処理は、線形残響成分重み算出 702 に移行する。

[0087] (vii) 線形残響成分重み算出処理

直接音や残留残響（変動性残響成分）のパワーは、時間毎に変化するため、上述したように、図 11 の処理によって変数として学習している。従って、T フレームの中で、どのフレームの直接音や残留残響のパワーが大きいかわかることができる（図 3 参照）。パワー（直接音や残留残響のパワーの和）が大きいほど変動成分が大きいことになるので、その時間帯は逆フィルタのパラメータを求めるための情報として使わない方が得策である。そこで、ここでは、例えば、パワー（直接音や残留残響のパワーの和）の大きさの逆数に比例するような重み係数を用いることとしている。

[0088] 線形残響成分重み算出（処理）707 は、 $R_{x, f, \tau(i)}$ を式 (20) に従って更新する。ここで、N は音源数とする。

[数20]

$$\mathbf{R}_{x,f,\tau(i)} = \sum_{n=1}^N v_{s(n),f,\tau(i)} \mathbf{C}_{s(n),f} + \sum_{m=1}^M \sum_{l=D}^{L_1} |\mathbf{x}_{f,\tau(i)-l}(m)|^2 \mathbf{C}_{rev,l,m,f} \quad \dots (20)$$

[0089] そして、時間毎の重みは、逆フィルタ算出 701 にフィードバックされる。なお、演算初期段階では、各時間帯の各成分のパワーは推定できていないため、重み係数を 1 として演算を開始する。そして、図 11 の演算を繰り返すことによって重み係数を適切な値に収束させることができるようになっていく。

[0090] <オンライン残響除去処理の詳細>

図 17 は、オンライン残響除去（処理）405 の具体的な構成を示す図である。オンライン残響除去（処理）405 は、逐次計算によりパラメータ推定精度を高める構成となっている。

[0091] (i) バッファリング処理

バッファリング（処理）801 は、フレーム毎の時間周波数領域信号を揮発性メモリ 103 上に格納する。本実施形態では、格納される時間周波数領域信号は、T フレーム内の信号のうち、最新の時間領域信号から数えて L1 フレーム（例えば、5 フレーム）分とする。

[0092] あるフレームの音声信号が得られたとき、そのフレームの残響成分を除去するには過去のフレームの信号が必要となる。そのため、バッファリング処理 801 によって所定フレーム分の音声信号を溜めて処理対象とする。

[0093] (ii) 線形残響成分消去処理

線形残響成分消去（処理）702 は、格納した L1 フレーム分の時間領域信号を受け取り、残響成分を逆フィルタにより除去する。この際、適用される逆フィルタは、残響除去パラメータ推定（処理）404 が出力した残響除去のパラメータに含まれるフィルタである。

[0094] (iii) 残留残響直接音分離処理

残留残響直接音分離（処理）703 は、線形残響成分消去（処理）702 から残響成分除去信号を受け取り、直接音と残留残響成分に分離し、直接音

を出力する。この際、 $v_{s(n),f,\tau}$ の初期値は1とする。また、 $C_{s(n),f}$ は残響除去パラメータ推定（処理）404が出力した残響除去のパラメータに含まれる共分散行列である。

[0095] (iv) 直接音分離推定処理

音声信号は時間毎にそのパワーが変化するため、時間毎の音声パワーの値を推定する必要がある。例えば、同じ話者が発生しても、時間毎に出力される音量は変化し、パワーが変化するため、リアルタイムに推定値を更新する必要があるからである。そこで、直接音分離推定（処理）802では、非線形パラメータの中の一部のパラメータ（直接音のパラメータ）のみをリアルタイムに推定する。そして、時間毎にリアルタイムに変化する部分には、推定処理を繰り返して推定値の精度を上げるようにする。なお、残留残響（変動性残響成分）のパラメータについては、時間変動が少ないと考えられるため、過去のフレームで学習したパラメータをそのまま用いれば良い。

[0096] 図17において、直接音分散推定（処理）802は、残留除去パラメータ推定404内の直接音パラメータ推定705と同様の流れで、式(21)（式(18)と同じ式）に従って、処理対象のフレームにおける $v_{s(n),f,\tau(i)}$ を算出する。

[数21]

$$v_{s(n),f,\tau} = \frac{1}{M} \text{trace} \left(C_{s,f}^{-1} \mathbf{V}_{s(n),f,\tau(i)} \right) \quad \dots (21)$$

[0097] (v) 収束判定処理

収束判定（処理）706は、推定した直接音分散などのパラメータを用いて、求めたパラメータの収束判定を行う。収束していると判断された場合には、収束判定（処理）706は、推定した直接音を出力して処理を終了する。それ以外の場合、収束判定（処理）706は、推定した直接音分散を基に再度、残留残響直接音分離（処理）703を実行する。

[0098] なお、収束したか否かの判定は、図11で説明した通りである。

[0099] (2) 第2の実施形態

第2の実施形態は、残響除去（処理）302において、残響除去パラメータ推定（処理）404で求めた過去の残響除去パラメータを複数組持っており、時間毎に最も良いフィルタを選択して用いる構成について開示している。

[0100] 図18は、本発明の第2の実施形態による残響除去（処理）302の構成を示す図である。なお、本構成による残響除去302においては、フレーム化（処理）401から残響除去パラメータ推定（処理）404までの構成及び処理と、時間領域変換（処理）406は、既に説明した図8の構成と同一である。

[0101] 図18において、パラメータ書き込み制御（処理）903は、残響除去パラメータ推定404が出力する残響除去パラメータを、新しい残響除去パラメータとして登録するかどうかを切り替える処理である。残響除去パラメータDB901は、予め定める個数分の残響除去パラメータをDB内に記憶している。

[0102] パラメータ書き込み制御903が実行する処理としては、例えば残響除去パラメータDB901内に記憶している残響除去パラメータのうち、記憶された時間が最も古い残響除去パラメータを廃棄し、新しい残響除去パラメータを代わりに記憶するような構成や、残響除去時の尤度（likelihood：ここでは誤差と同義である）の値が最も小さい残響除去パラメータを廃棄し、新しい残響除去パラメータを代わりに記憶するような構成を採っても良い。また、廃棄する残響除去パラメータは、周波数毎に同じタイミングで記憶された残響除去パラメータを廃棄するような構成を取っても良い。

[0103] ここで、残響除去パラメータDB901に記憶されている残響除去パラメータの数をAとすると、オンライン残響除去（処理）405-1～405-Aはそれぞれ、各残響除去パラメータにオンライン処理で残響成分を除去する処理を実行する。

[0104] 最適残響除去音選択（処理）902は、各残響除去パラメータで除去され

た残響除去音の内、一つの残響除去音を選択する。例えば、各残響除去音の中で最も音量が小さい成分を選択するような構成や、尤度値が最も大きくなるような残響除去音を選択するような構成を取っても良い。また、音量成分や尤度値の計算は周波数方向に平均した値を用いても良い。

[0105] 選択された残響除去音は、時間領域変換（処理）406に送られ、時間領域信号に変換されて出力される。例えば、低周波数領域の音声信号についてはパラメータ1を、高周波数領域の音声信号についてはパラメータ2を使って残響除去することが可能となる。これにより、周波数ごとに最適なフィルタを求めることができ、複数人が同時に発話するような場面であっても正確な残響除去処理を実行することが可能となる。また、第2の実施形態では、過去に求めた残響除去パラメータを蓄積しており、話者が切り替わっても過去に求めた最適なパラメータを使うことができるので、残響除去処理を迅速に実行することができるようになる。

[0106] （3）第3の実施形態

第3の実施形態は、残響除去とエコーキャンセラを同じフレームワーク内で実行し、残響除去とエコーキャンセラの双方の性能を向上させることが可能な構成に関するものである。残響除去とエコーキャンセラ消去は別々に動作させることもでき、この構成が単純である（図6参照）。しかし、エコーキャンセラのフィルタは常時形状が変化するため、エコーキャンセラで消去できなかった音声が残響除去に悪影響を及ぼす場合がある。そこで、残響除去及びエコーキャンセラをそれぞれ別々に最適化するのではなく、2つを同時に（互いに及ぼす影響を考慮して）最適化した方がシステム全体のパフォーマンスをより向上させることができる。これを実現するための構成が第3の実施形態に係るものである。

[0107] 図19は、第3の実施形態による処理を実行する中央演算装置の内部構成を概略的に示す図である。

[0108] 中央演算装置102内で実行されるプログラム内で、残響除去及びエコーキャンセラ2001は、デジタル音声波形（マイクロホン入力信号）及び遠

端デジタル音声波形（参照信号）を受信し、それに対して残響除去処理及びエコーキャンセル処理を同時に実行し、残響除去・エコーキャンセル後の音声波形を出力する。

[0109] <残響除去及びエコーキャンセラの構成>

図20は、残響除去及びエコーキャンセラ2001の具体的な構成を示す図である。オフラインパラメータ推定（処理）1800は、デジタル音声波形（マイクロホン入力信号）と遠端デジタル音声波形（参照信号）を受信し、残響除去と音響エコーキャンセラ用のパラメータを推定する。オフラインパラメータ推定1800は、複数フレーム（Tフレーム分）時間領域信号が得られるごとに推定処理を実行する。オフラインパラメータ推定（処理）1800の詳細な構成及び処理については、図21を参照して後述する。

[0110] オンライン音響エコー・残響除去（処理）2301は、オフラインパラメータ推定1800によって推定されたパラメータを用いて、フレーム毎の時間周波数領域信号に対して残響除去及び音響エコー成分の除去を行う。オンライン音響エコー・残響除去（処理）2301の詳細な構成及び処理については、図24を参照して後述する。

[0111] 時間領域変換406は、残響成分及び音響エコー成分が除去された時間周波数領域信号を時間領域信号に変換してそれを出力する。

[0112] <オフラインパラメータ推定処理の詳細>

図21は、オフラインパラメータ推定（処理）1800の具体的なブロック構成を示す図である。

[0113] マイクロホンアレイ105で得られた音声波形をデジタル信号に変換したデジタル音声波形及び遠端デジタル音声波形（参照信号と呼称）は共に、フレーム化（処理）401及び周波数分解（処理）402が施され、時間周波数領域信号に変換される。

[0114] スピーカアレイ107を構成するスピーカ素子のうちb番目の素子の参照信号の時間領域信号を $X_{ref, b}(f, \tau)$ と記載すると、 $X_m(f, \tau)$ 及び $X_{ref, b}(f, \tau)$ は共に、バッファリング403にて複数フレーム分（Tフレーム）バッ

ファに蓄積される。ここで、 $X_{ref, b}(f, \tau) = [X_{ref, 1}(f, \tau), \dots, X_{ref, B}(f, \tau)]$ とし、 B をスピーカ素子数とする。

[0115] そして、バッファリング（処理）403によってTフレーム分のデータが溜まる度に、残響・音響エコー消去用パラメータ推定（処理）1801が実行され、残響・音響エコー消去用パラメータが出力される。残響・音響エコー消去用パラメータ推定（処理）1801の詳細な構成及び処理については、図22を参照して後述する。

[0116] <残響・音響エコー消去用パラメータ推定処理の詳細>

図22は、残響・音響エコー消去用パラメータ推定（処理）1801の具体的なブロック構成を示す図である。線形残響成分消去702、残留残響パラメータ推定704、直接音パラメータ推定705、及び収束判定706については、上述した通りの構成及び処理であるので、詳細な説明は割愛する。

[0117] (i) 逆フィルタ算出処理

逆フィルタ算出（処理）1908は、前述の方法（第1の実施形態）で残響（非変動成分）除去用の逆フィルタを求める。逆フィルタを算出する際に、マイク入力信号の代わりに、線形音響エコー消去フィルタ算出（処理）1901の結果を用いて線形音響エコーを消去した信号を使っても良い。この場合、逆フィルタ算出1908は、線形音響エコー消去（処理）1902の機能を有している。つまり、逆フィルタ算出1908は、線形音響エコー消去フィルタ算出1901が算出したエコー消去フィルタを使って、入力信号（複数フレームの周波数領域信号）に含まれる音響エコーを消去した後、残響除去用の逆フィルタを算出する。

[0118] (ii) 線形音響エコー消去フィルタ算出処理

線形音響エコー消去フィルタ算出（処理）1901は、式（22）を用いて音響エコー消去用のフィルタを求める。

[数22]

$$\mathbf{J}_f = \text{ivec}(\mathbf{J}_{p,f}^{-1} \mathbf{J}_{q,f}) \quad \dots (22)$$

[0119] この際、線形音響エコー消去フィルタ算出（処理）1901は、マイク入力信号の代わりに、前述の逆フィルタ算出1908により求めた逆フィルタを使って残響成分を除去した信号を用いて音響エコー消去用のフィルタを求めても良い。この場合、線形音響エコー消去フィルタ算出（処理）1901は、線形残響成分消去（処理）702の機能を有している。つまり、線形音響エコー消去フィルタ算出（処理）1901は、逆フィルタ算出1908で算出した残響除去用の逆フィルタを使って残響を除去した後に、線形音響エコー消去フィルタを算出する。

[0120] 式（22）において、 $\mathbf{J}_{p,f}$ 、 $\mathbf{J}_{u,f,\tau(i)}$ 及び $\mathbf{J}_{q,f}$ はそれぞれ、式（23）、（24）及び（25）のように定義される。

[数23]

$$\mathbf{J}_{p,f} = \left(\sum_{i=1}^T \left(\mathbf{J}_{u,f,\tau(i)} \mathbf{J}_{u,f,\tau(i)}^H \right) \otimes \mathbf{R}_{x,f,\tau(i)}^{-1} \right) \quad \dots (23)$$

[数24]

$$\mathbf{J}_{u,f,\tau(i)} = [\mathbf{x}_{ref,f,\tau(i)}^H \dots \mathbf{x}_{ref,f,\tau(i)-L_2}^H]^H \quad \dots (24)$$

[数25]

$$\mathbf{J}_{q,f} = \text{vec} \left(\sum_{i=1}^T \mathbf{R}_{x,f,\tau(i)}^{-1} \mathbf{x}_{f,\tau(i)} \mathbf{J}_{u,f,\tau(i)}^H \right) \quad \dots (25)$$

[0121] また、音響エコー消去用フィルタは、式（26）のように、タップ毎のフィルタに分割される。

[数26]

$$\mathbf{J}_f = [\mathbf{W}_{ref,f} \cdots \mathbf{W}_{ref,f,L_2}] \quad \cdots \quad (26)$$

[0122] (iii) 線形音響エコー消去処理

線形音響エコー消去（処理）1902は、線形音響エコー消去フィルタ算出1901で算出した音響エコー消去用のフィルタを使って音響エコー成分を消去した信号 $\mathbf{g}_{2,f,\tau(i)}$ を式（27）に従って取得する。

[数27]

$$\mathbf{g}_{2,f,\tau(i)} = \mathbf{g}_{f,\tau(i)} - \sum_{l=0}^{L_2} \mathbf{W}_{ref,f,l} \mathbf{x}_{ref,f,\tau(i)-l} \quad \cdots \quad (27)$$

[0123] (iv) 残留残響・残留音響エコー・直接音分離処理

残留残響・残留音響エコー・直接音分離（処理）1904は、残留残響及び直接音については残留残響・直接音分離703（第1の実施形態）と同一の方法で求める。なお、残留音響エコーについては、式（28）によって求めたで求めた残留音響エコー抽出用フィルタ $\mathbf{W}_{ref,l,b,f,\tau(i)}$ を $\mathbf{g}_{2,f,\tau(i)}$ に重畳することにより、残留音響エコー推定値 $\mathbf{y}_{ref,l,b,f,\tau(i)}$ を式（29）に従って算出する。

[数28]

$$\mathbf{W}_{ref,l,b,f,\tau(i)} = \mathbf{R}_{ref,l,b,f,\tau(i)} \mathbf{R}_{x,f,\tau(i)}^{-1} \quad \cdots \quad (28)$$

[数29]

$$\mathbf{y}_{ref,l,b,f,\tau(i)} = \mathbf{W}_{ref,l,b,f,\tau(i)} \mathbf{g}_{2,f,\tau(i)} \quad \cdots \quad (29)$$

[0124] ここで、 $\mathbf{R}_{ref,l,b,f,\tau(i)}$ は、式（30）によって求めることができる。

[数30]

$$\mathbf{R}_{ref,l,b,f,\tau(i)} = |\mathbf{x}_{ref,f,\tau(i)-l}(b)|^2 \mathbf{C}_{ref,l,b,f} \quad \cdots \quad (30)$$

[0125] 式(30)において、 $C_{ref, l, b, f}$ は、繰り返し計算により更新されるパラメータであり、初期値はランダムな正定エルミート行列とする。

[0126] (v) 残留音響エコーパラメータ推定処理

残留音響エコーパラメータ推定(処理)1906は、 $C_{ref, l, b, f}$ を残留残響パラメータ推定(処理)704(図11:第1の実施形態)と同一の処理により更新する。詳細な説明については上述の通りであるためここでは省略する。

[0127] (vi) 線形残響・エコー成分重み算出処理

線形残響・エコー成分重み算出(処理)1907は、式(31)により、 $R_{x, f, \tau(i)}$ を算出する。

[数31]

$$R_{x, f, \tau(i)} = \sum_{n=1}^N v_{s(n), f, \tau(i)} C_{s(n), f} + \sum_{n=1}^M \sum_{l=D}^{L_1} |x_{f, \tau(i)-l}(m)|^2 C_{rev, l, n, f} + \sum_{b=1}^B \sum_{l=0}^{L_2} |x_{ref, f, \tau(i)-l}(b)|^2 C_{ref, l, b, f} \quad \dots (31)$$

[0128] そして、図23に示されるように、 $R_{x, f, \tau(i)}$ に対して影響を与える変動性残響成分だけでなく変動性音響エコー成分が大きい場合に、それらの重みが小さくなるように高精度に線形フィルタを求めることができる。

[0129] <オンライン音響エコー・残響除去処理>

図24は、オンライン(リアルタイム)音響エコー・残響除去(処理)2301の具体的な構成を示す図である。

[0130] デジタル音声波形(マイク入力音声信号)及び遠端デジタル音声波形(参照信号)は、フレーム化(処理)401及び周波数分解(処理)402が施され、更にバッファリング(処理)403によりバッファリングされる。

[0131] バッファリングされた音声波形は、線形音響エコー・残響除去(処理)1902に送られる。

[0132] 線形音響エコー・残響除去(処理)1902は、受信した音声波形の各フレームのデータに対して非変動性の残響と音響エコー成分を除去する。

[0133] その後、残留残響・残留音響エコー・直接音分離(処理)1904は、直接音のみを抽出する。

[0134] 直接音分散推定（処理）802は、残留残響・残留音響エコー・直接音分離（処理）1904から抽出された直接音を受信し、 $v_{s(n), f, \tau}$ を算出する。当該処理の詳細は、第1の実施形態で説明した通りであるので説明を省略する。

[0135] その後、収束判定（処理）706は、分散推定が収束したかどうか判定し、収束していれば推定した直接音成分を出力する。それ以外の場合、推定した直接音分散値を残留残響・残留音響エコー・直接音分離1904に戻し、再度直接音推定処理が実行される。当該処理の詳細も、第1の実施形態で説明した通りである。

[0136] （4）第4の実施形態

第4の実施形態は、残響除去の中で、特に計算量が大きい残響除去パラメータを会議情報計算サーバ201で実行し、それ以外のリアルタイムの残響除去処理を拠点毎会議システム100で実行するような分散構成に関する。

[0137] 図25は、第4の実施形態による全体のシステム概略構成を示す図である。図25において、フレーム化（処理）401、周波数分解（処理）402、及びバッファリング（処理）403は拠点毎会議システム100で実行される。そして、Tフレーム分時間周波数領域信号が蓄積された後、拠点毎会議システム100は、その時間周波数領域信号を会議情報計算サーバ201に送信する。

[0138] 会議情報計算サーバ201は、拠点毎会議システム100から送られてきたTフレーム分時間周波数領域信号を受信し、その信号に対して残響除去パラメータ推定404を実行する。そして、会議情報計算サーバ201は、推定した残響除去パラメータをサーバから拠点毎会議システム100に送信する。

[0139] 拠点毎会議システム100は、フレーム毎の時間周波領域信号を得る毎に、オンライン残響除去405、及び時間領域変換406を実行し、残響除去音（残響除去された直接音）を取得する。

[0140] なお、会議情報計算サーバ201は、残響除去パラメータ推定（処理）4

０４の構成の代わりに、第３の実施形態で示した残響・音響エコー消去用パラメータ推定（処理）１８０１の構成を含むようにしても良い。この場合、拠点毎会議システム１００は、オンライン残響除去（処理）４０５の構成の代わりに、第３の実施形態で示したオンライン音響エコー・残響除去（処理）２３０１の構成を含むことになる。

[0141] （５）まとめ

（ｉ）本発明の第１の実施形態の残響除去パラメータ推定装置においては、揮発性メモリ等のメモリに残響除去パラメータを格納し、処理に従って逐次更新していく。メモリには、少なくとも、音声入力信号に含まれる非変動性残響成分を除去するための線形フィルタのパラメータと、音声入力信号に含まれる変動性残響成分を除去するための非線形フィルタのパラメータと、が格納される。そして、中央処理装置等のプロセッサが、音声入力信号に含まれる残響成分を除去して直接音を取得するための残響除去パラメータを推定及び更新し、当該残響除去パラメータを、線形フィルタのパラメータ及び非線形フィルタのパラメータとしてメモリに格納する。より具体的には、プロセッサ（逆フィルタ算出７０１及び線形残響成分消去７０２）は、メモリから線形フィルタのパラメータを読み出し、当該線形フィルタによって音声入力信号に含まれる非変動性残響成分を除去して線形残響除去信号の推定値を生成する。次に、プロセッサ（残留残響・直接音分離７０３）は、メモリから非線形フィルタのパラメータを読み出し、当該非線形フィルタを用いて、線形残響除去信号の推定値に含まれる変動性残響成分及び直接音成分の推定値を生成する。続いて、プロセッサ（残留残響パラメータ推定７０４及び直接音パラメータ推定７０５）は、変動性残響成分及び直接音の推定値に基づいて主軸演算（図１５及び１６参照）を実行し、非線形フィルタのパラメータを構成する変動残響成分及び直接音成分のパラメータを更新する。そして、プロセッサ（線形残響成分重み算出７０７及び逆フィルタ算出７０１）は、更新された変動性残響成分及び直接音成分のパラメータに基づいて、線形フィルタのパラメータを逐次的に更新する。つまり、非変動性残響成分を除去

するための線形フィルタのパラメータをある値に固定して、変動性残響成分を除去するための非線形フィルタのパラメータを推定し、得られた推定値に照らして再度線形フィルタのパラメータを更新するようにして、推定パラメータが収束するまでパラメータ推定処理を繰り返す。このようにすることにより、精度良く残響除去するための線形及び非線形フィルタのパラメータ（残響除去パラメータ）を推定することができる。そして、このパラメータを用いて残響除去処理を実行することにより、マイクロホンからの音声入力信号の直接音成分を歪ませることなく、残響成分を精度良く除去することができるようになる。

[0142] また、プロセッサ（線形残響成分重み算出 707）は、更新された変動性残響成分及び直接音成分のパラメータによって、音声入力信号において変動性残響成分と直接音成分の和が大きい時間帯を反映して線形フィルタの重み係数を決定（図 3 参照）し、線形フィルタのパラメータを更新する。このように、線形フィルタの伝達関数が揺らぐ可能性のある時間帯におけるフィルタ係数の重みを小さくすることにより、非変動性残響成分が大きく存在する時間帯の比重を大きくして信号を抽出することができるため、伝達関数の推定精度が落ちる影響を軽減することが可能となる。

[0143] 本発明の実施形態では、音声入力信号として、複数のマイクロホンからの音声信号を想定している。この場合、プロセッサ（残留残響・直接音分離 703）は、複数のマイクロホンのそれぞれからの音声信号に含まれる直接音成分の推定値及び変動性残響成分の推定値を生成する。そして、プロセッサ（残留残響パラメータ推定 704 及び直接音パラメータ推定 705）は、それぞれのマイクロホンからの直接音成分及び変動性残響成分の推定値の二次統計量を抽出し、二次統計量から、複数のマイクの音声信号の直接音成分及び変動性残響成分のそれぞれのパワーのばらつきを表す情報を、変動性残響成分及び直接音成分のパラメータとする。このようにすることにより、音源（話者や、壁及び人物等の反射体における音声反射面（音源から発せられた音が反射する箇所））から各マイクロホンまでの距離が異なる場合に、各マ

イクロホンに入力される音声のパワーのばらつきを考慮し、残響を精度良く除去できるパラメータを推定することが可能となる。

[0144] 第1の実施形態は、さらに、上述した残響除去パラメータ推定装置を有する残響除去装置を提案している。当該装置（図8及び17参照）では、プロセッサ（中央処理装置）は、残響除去パラメータ推定装置から残響除去パラメータを取得して、音声入力信号から非変動性残響成分及び変動性残響成分を除去し、前記直接音成分を出力する。より具体的には、プロセッサ（線形残響成分消去702）は、線形フィルタによって音声入力信号から非変動性残響成分を除去して線形残響除去信号を推定する。次に、プロセッサ（残留残響・直接音分離703）は、非線形フィルタによって線形残響除去信号から変動性残響成分と直接音成分を分離する。そして、プロセッサ（直接音分散推定802及び収束判定706）は、直接音成分における時間毎のパワーの変化を学習し、当該パワーの変化が収束した直接音成分を出力する。このようにすることにより、リアルタイムに残響を精度良く除去することができるようになる。また、直接音分散推定（時間毎に変化する音声パワーに応じて直接音を更新する処理）を実行しているので、時間経過に伴って起こる音声パワーの変化に追従して歪みのない直接音を抽出することができるようになる。

[0145] (ii) 第2の実施形態は、別の残響除去装置を提案している。当該装置（図18）は、残響除去パラメータ推定装置によって得られた残響除去パラメータを格納する残響除去パラメータデータベースを有している。このデータベースには、過去に推定して得られた残響除去パラメータの全て或いは一部が格納されている。当該装置では、プロセッサ（設置されたマイクロホンに対応するオンライン残響除去450-1乃至405-A）が、残響除去パラメータデータベースから対応する複数の残響除去パラメータを取得し、当該複数の残響除去パラメータのそれぞれを用いて音声入力信号から非変動性残響成分及び変動性残響成分を除去し、複数の直接音成分を生成し、当該複数の直接音成分のうち最適な直接音成分を出力する。このように、推定済のパラ

メータを用いて残響除去の処理を実行することができるので、残響除去処理を迅速に行うことができ、よってリアルタイム処理に資することになる。

[0146] (iii) 第3の実施形態では、メモリに、音声入力信号に含まれる非変動性残響成分を除去するための線形残響フィルタのパラメータと、音声入力信号に含まれる非変動性エコー成分を除去するための線形エコー消去フィルタのパラメータと、音声入力信号に含まれる変動性残響成分及び変動性エコー成分を除去するための非線形フィルタのパラメータと、を格納している。また、プロセッサ（中央処理装置）は、音声入力信号に含まれる残響成分及びエコー成分を除去して直接音を取得するための残響・エコー除去パラメータを推定し、当該残響・エコー除去パラメータを、線形残響フィルタのパラメータ、線形エコー消去フィルタのパラメータ、及び非線形フィルタのパラメータとしてメモリに格納する。より具体的に、プロセッサ（逆フィルタ算出1908及び線形音響エコー消去フィルタ算出1901）は、メモリから線形残響フィルタのパラメータのパラメータと、線形エコー消去フィルタのパラメータを読み出し、線形残響フィルタ及び線形エコー消去フィルタによって音声入力信号に含まれる非変動性残響成分及び非変動性エコー成分を除去して線形残響・エコー除去信号の推定値を生成する。この場合、逆フィルタ算出1908と線形音響エコー消去フィルタ算出1901は、互いの処理結果を利用してフィルタを算出する。つまり、線形音響エコー消去フィルタ算出1901は、音声入力信号から非変動残響成分を除去した信号を用いて線形音響エコー消去フィルタを算出する。一方、逆フィルタ算出1908は、音声入力信号から非変動音響エコー成分を除去した信号を用いて逆フィルタを算出する。次に、プロセッサ（残留残響・残留音響エコー・直接音分離1904）は、非線形フィルタを用いて、線形残響・エコー除去信号の推定値に含まれる変動性残響成分、変動性エコー成分、及び直接音成分の推定値を生成する。さらに、プロセッサ（残留残響パラメータ推定704、直接音パラメータ推定705、及び残留音響エコーパラメータ推定1906）は、変動性残響成分、変動性エコー成分、及び直接音の推定値に基づいて、非線形フィ

ルタのパラメータを構成する変動残響成分、変動性エコー成分、及び直接音成分のパラメータを更新する。そして、プロセッサ（線形残響・エコー成分重み算出1907、線形音響エコー消去フィルタ算出1901、及び逆フィルタ算出1908）は、更新された変動性残響成分、変動性エコー成分、及び直接音成分のパラメータに基づいて、線形残響フィルタ及び線形エコー消去フィルタのパラメータを逐次的に更新する。このようにすることにより残響除去とエコーキャンセルを同じフレームワーク内で実行することにより、双方の処理精度を向上させることが可能なパラメータを推定することができるようになる。

[0147] (iv) 第4の実施形態は、処理に負荷が掛かる残響パラメータ推定処理を、会議システムの計算機サーバにおいて実行する構成（図25）について提案している。このようにすることにより、それぞれの拠点における会議システムの処理負荷を軽減することができるため、会議システムのリアルタイム性を担保することができると共に、当該会議システムを構築するためのコストを劇的に削減することができるようになる。

[0148] (v) 本発明は、実施形態の機能を実現するソフトウェアのプログラムコードによっても実現できる。この場合、プログラムコードを記録した記憶媒体をシステム或は装置に提供し、そのシステム或は装置のコンピュータ（又はCPUやMPU）が記憶媒体に格納されたプログラムコードを読み出す。この場合、記憶媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコード自体、及びそれを記憶した記憶媒体は本発明を構成することになる。このようなプログラムコードを供給するための記憶媒体としては、例えば、フレキシブルディスク、CD-ROM、DVD-ROM、ハードディスク、光ディスク、光磁気ディスク、CD-R、磁気テープ、不揮発性のメモ리카ード、ROMなどが用いられる。

[0149] また、プログラムコードの指示に基づき、コンピュータ上で稼動しているOS（オペレーティングシステム）などが実際の処理の一部又は全部を行い

、その処理によって前述した実施の形態の機能が実現されるようにしてもよい。さらに、記憶媒体から読み出されたプログラムコードが、コンピュータ上のメモリに書きこまれた後、そのプログラムコードの指示に基づき、コンピュータのCPUなどが実際の処理の一部又は全部を行い、その処理によって前述した実施の形態の機能が実現されるようにしてもよい。

[0150] さらに、実施の形態の機能を実現するソフトウェアのプログラムコードを、ネットワークを介して配信することにより、それをシステム又は装置のハードディスクやメモリ等の記憶手段又はCD-RW、CD-R等の記憶媒体に格納し、使用時にそのシステム又は装置のコンピュータ（又はCPUやMPU）が当該記憶手段や当該記憶媒体に格納されたプログラムコードを読み出して実行するようにしても良い。

[0151] 最後に、ここで述べたプロセス及び技術は本質的に如何なる特定の装置に関連することはなく、コンポーネントの如何なる相応しい組み合わせによってでも実装できることを理解する必要がある。更に、汎用目的の多様なタイプのデバイスがここで記述した教授に従って使用可能である。ここで述べた方法のステップを実行するのに、専用の装置を構築するのが有益であることが判るかもしれない。また、実施形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。さらに、異なる実施形態にわたる構成要素を適宜組み合わせてもよい。本発明は、具体例に関連して記述したが、これらは、すべての観点に於いて限定の為ではなく説明の為である。本分野にスキルのある者には、本発明を実施するのに相応しいハードウェア、ソフトウェア、及びファームウェアの多数の組み合わせがあることが解るであろう。例えば、記述したソフトウェアは、アセンブラ、C/C++、perl、Shell、PHP、Java（登録商標）等の広範囲のプログラム又はスクリプト言語で実装できる。

[0152] さらに、上述の実施形態において、制御線や情報線は説明上必要と考えられるものを示しており、製品上必ずしも全ての制御線や情報線を示している

とは限らない。全ての構成が相互に接続されていても良い。

[0153] 加えて、本技術分野の通常の知識を有する者には、本発明のその他の実装がここに開示された本発明の明細書及び実施形態の考察から明らかになる。記述された実施形態の多様な態様及び／又はコンポーネントは、データを管理する機能を有するコンピュータ化ストレージシステムに於いて、単独又は如何なる組み合わせでも使用することが出来る。明細書と具体例は典型的なものに過ぎず、本発明の範囲と精神は後続する請求範囲で示される。

符号の説明

[0154] 100…拠点毎会議システム、101…不揮発性メモリ、102…中央演算装置、103…揮発性メモリ、104…A/D変換機、105…マイクロホンアレイ、106…D/A変換機、107…スピーカアレイ、108…HUB、109…カメラ、110…ディスプレイ、201…会議情報計算サーバ、202…MCU、301…エコーキャンセラ、302…残響除去、401…フレーム化、402…周波数分解、403…バッファリング、404…残響除去パラメータ推定、405…オンライン残響除去、406…時間領域変換、701…逆フィルタ算出、702…線形残響成分消去、703…残留残響・直接音分離、704…残留残響パラメータ推定、705…直接音パラメータ推定、706…収束判定、707…線形残響成分重み算出、801…バッファリング、802…直接音分散推定、901…残響除去パラメータDB、902…最適残響除去音選択、903…パラメータ書き込み制御、1102…直接音フィルタ係数推定、1103…残留残響フィルタ係数推定、1104…時間毎の目的音分散推定、1105…時間毎の残留残響分散推定、1201…直接音二次統計量算出、1202…時変パラメータ算出、1203…主軸算出、1301…残留残響二次統計量算出、1302…主軸算出、1401…重み付き相関行列計算、1402…重み付き共分散行列計算、1403…フィルタ計算、1800…オフラインパラメータ推定、1801…残響・音響エコー消去用パラメータ推定、1901…線形音響エコー消去フィルタ算出、1902…線形音響エコー消去、1904…残留残響・残留音響

エコー・直接音分離、１９０６…残留音響エコーパラメータ推定、１９０７
…線形残響・エコー成分重み算出、１９０８…逆フィルタ算出、２００１…
残響除去及びエコーキャンセラ、２３０１…オンライン音響エコー・残響除
去

請求の範囲

[請求項1]

音声入力信号に含まれる非変動性残響成分を除去するための線形フィルタのパラメータと、前記音声入力信号に含まれる変動性残響成分を除去するための非線形フィルタのパラメータと、を格納するメモリと、

前記音声入力信号に含まれる残響成分を除去して直接音を取得するための残響除去パラメータを推定し、当該残響除去パラメータを、前記線形フィルタのパラメータ及び前記非線形フィルタのパラメータとして前記メモリに格納するプロセッサと、を有し、

前記プロセッサは、

前記メモリから前記線形フィルタのパラメータを読み出し、当該線形フィルタによって音声入力信号に含まれる非変動性残響成分を除去して線形残響除去信号の推定値を生成し、

前記メモリから前記非線形フィルタのパラメータを読み出し、当該非線形フィルタを用いて、前記線形残響除去信号の推定値に含まれる変動性残響成分及び直接音成分の推定値を生成し、

前記変動性残響成分及び前記直接音の推定値に基づいて、前記非線形フィルタのパラメータを構成する前記変動残響成分及び前記直接音成分のパラメータを更新し、

前記更新された変動性残響成分及び直接音成分のパラメータに基づいて、前記線形フィルタのパラメータを逐次的に更新する、ことを特徴とする残響除去パラメータ推定装置。

[請求項2]

請求項1において、

前記プロセッサは、前記更新された変動性残響成分及び直接音成分のパラメータによって、前記音声入力信号において前記変動残響成分と前記直接音成分の和が大きい時間帯を反映して前記線形フィルタの重み係数を決定し、前記線形フィルタのパラメータを更新することを特徴とする残響除去パラメータ推定装置。

- [請求項3] 請求項1において、
前記音声入力信号は、複数のマイクからの音声信号を含み、
前記プロセッサは、
前記複数のマイクのそれぞれからの音声信号に含まれる前記直接音成分の推定値及び前記変動性残響成分の推定値を生成し、
前記それぞれのマイクからの前記直接音成分及び前記変動性残響成分の推定値の二次統計量を抽出し、
前記二次統計量から、前記複数のマイクの音声信号の前記直接音成分及び前記変動性残響成分のそれぞれのパワーのばらつきを表す情報を、前記変動性残響成分及び前記直接音成分のパラメータとする、
ことを特徴とする残響除去パラメータ推定装置。
- [請求項4] 請求項1において、
前記プロセッサは、前記更新された変動性残響成分及び直接音成分のパラメータの収束判定を実行し、収束していると判断した場合に、前記残響除去パラメータとして、前記線形フィルタ及び前記非線形フィルタのパラメータを出力し、未収束であると判断した場合には、前記線形フィルタのパラメータの逐次的更新処理を継続する、ことを特徴とする残響除去パラメータ推定装置。
- [請求項5] 音声入力信号に含まれる非変動性残響成分を除去するための線形残響フィルタのパラメータと、前記音声入力信号に含まれる非変動性エコー成分を除去するための線形エコー消去フィルタのパラメータと、
前記音声入力信号に含まれる変動性残響成分及び変動性エコー成分を除去するための非線形フィルタのパラメータと、を格納するメモリと、
前記音声入力信号に含まれる残響成分及びエコー成分を除去して直接音を取得するための残響・エコー除去パラメータを推定し、当該残響・エコー除去パラメータを、前記線形残響フィルタのパラメータ、前記線形エコー消去フィルタのパラメータ、及び前記非線形フィルタ

のパラメータとして前記メモリに格納するプロセッサと、を有し、
前記プロセッサは、

前記メモリから前記線形残響フィルタのパラメータと、線形エコー消去フィルタのパラメータを読み出し、前記線形残響フィルタ及び前記線形エコー消去フィルタによって音声入力信号に含まれる非変動性残響成分及び非変動性エコー成分を除去して線形残響・エコー除去信号の推定値を生成し、

前記メモリから前記非線形フィルタのパラメータを読み出し、当該非線形フィルタを用いて、前記線形残響・エコー除去信号の推定値に含まれる変動性残響成分、変動性エコー成分、及び直接音成分の推定値を生成し、

前記変動性残響成分、前記変動性エコー成分、及び前記直接音の推定値に基づいて、前記非線形フィルタのパラメータを構成する前記変動残響成分、前記変動性エコー成分、及び前記直接音成分のパラメータを更新し、

前記更新された変動性残響成分、変動性エコー成分、及び直接音成分のパラメータに基づいて、前記線形残響フィルタ及び前記線形エコー消去フィルタのパラメータを逐次的に更新する、ことを特徴とする残響・エコー除去パラメータ推定装置。

[請求項6] 請求項5において、

前記プロセッサは、前記非変動性残響成分を除去した後の音声信号に対して適用する前記線形エコー消去フィルタのパラメータを更新し、前記非変動性エコー成分を除去した後の音声信号に対して適用する前記線形残響除去フィルタのパラメータを更新することを特徴とする残響・エコー除去パラメータ推定装置。

[請求項7] プロセッサによって、音声入力信号に含まれる残響成分を除去して直接音を取得するための残響除去パラメータを推定し、当該残響除去パラメータを、音声入力信号に含まれる非変動性残響成分を除去する

ための線形フィルタのパラメータ、及び前記音声入力信号に含まれる変動性残響成分を除去するための非線形フィルタのパラメータとして、出力する残響除去パラメータ推定方法であって、

前記プロセッサが、メモリから前記線形フィルタのパラメータを読み出し、当該線形フィルタによって音声入力信号に含まれる非変動性残響成分を除去して線形残響除去信号の推定値を生成するステップと、

前記プロセッサが、前記メモリから前記非線形フィルタのパラメータを読み出し、当該非線形フィルタを用いて、前記線形残響除去信号の推定値に含まれる変動性残響成分及び直接音成分の推定値を生成するステップと、

前記プロセッサが、前記変動性残響成分及び前記直接音の推定値に基づいて、前記非線形フィルタのパラメータを構成する前記変動残響成分及び前記直接音成分のパラメータを更新するステップと、

前記プロセッサが、前記更新された変動性残響成分及び直接音成分のパラメータに基づいて、前記線形フィルタのパラメータを逐次的に更新するステップと、

を有することを特徴とする残響除去パラメータ推定方法。

[請求項8] 音声入力信号に含まれる残響成分を除去する残響除去装置であって、

請求項1に記載の残響除去パラメータ推定装置と、

前記残響除去パラメータ推定装置から前記残響除去パラメータを取得して、前記音声入力信号から前記非変動性残響成分及び前記変動性残響成分を除去し、前記直接音成分を出力するプロセッサと、を有し、

前記プロセッサは、

前記線形フィルタによって前記音声入力信号から前記非変動性残響成分を除去して線形残響除去信号を推定し、

前記非線形フィルタによって前記線形残響除去信号から前記変動性残響成分と前記直接音成分を分離し、

前記直接音成分における時間毎のパワーの変化を学習し、当該パワーの変化が収束した前記直接音成分を出力する、ことを特徴とする残響除去装置。

[請求項9] 音声入力信号に含まれる残響成分を除去する残響除去装置であって、

請求項1に記載の残響除去パラメータ推定装置と、

前記残響除去パラメータ推定装置によって得られた前記残響除去パラメータであって、過去の残響除去パラメータを格納する残響除去パラメータデータベースと、

前記残響除去パラメータデータベースから複数の残響除去パラメータを取得し、当該複数の残響除去パラメータのそれぞれを用いて前記音声入力信号から前記非変動性残響成分及び前記変動性残響成分を除去し、複数の直接音成分を生成し、当該複数の直接音成分のうち最適な直接音成分を出力するプロセッサと、
を有することを特徴とする残響除去装置。

[請求項10] 請求項9において、

前記プロセッサは、前記複数の残響除去パラメータのそれぞれを用いて、

前記線形フィルタによって前記音声入力信号から前記非変動性残響成分を除去して線形残響除去信号を推定し、

前記非線形フィルタによって前記線形残響除去信号から前記変動性残響成分と前記直接音成分を分離し、

前記直接音成分における時間毎のパワーの変化を学習し、当該パワーの変化が収束した前記直接音成分を、前記複数の残響除去パラメータに対応して出力する、ことを特徴とする残響除去装置。

[請求項11] 音声入力信号に含まれる残響成分を除去する残響・エコー除去装置

であって、

請求項 5 に記載の残響・エコー除去パラメータ推定装置と、

前記残響・エコー除去パラメータ推定装置から前記残響・エコー除去パラメータを取得して、前記音声入力信号から前記非変動性及び前記変動性残響成分、並びに非変動性及び変動性エコー成分を除去し、前記直接音成分を出力するプロセッサと、を有し、

前記プロセッサは、

前記線形残響フィルタ及び前記線形エコー消去フィルタによって前記音声入力信号から前記非変動性残響成分及び前記非変動性エコー成分を除去して線形残響・エコー除去信号を推定し、

前記非線形フィルタによって前記線形残響・エコー除去信号から前記変動性残響及びエコー成分と前記直接音成分を分離し、

前記直接音成分における時間毎のパワーの変化を学習し、当該パワーの変化が収束した前記直接音成分を出力する、ことを特徴とする残響・エコー除去装置。

[請求項12]

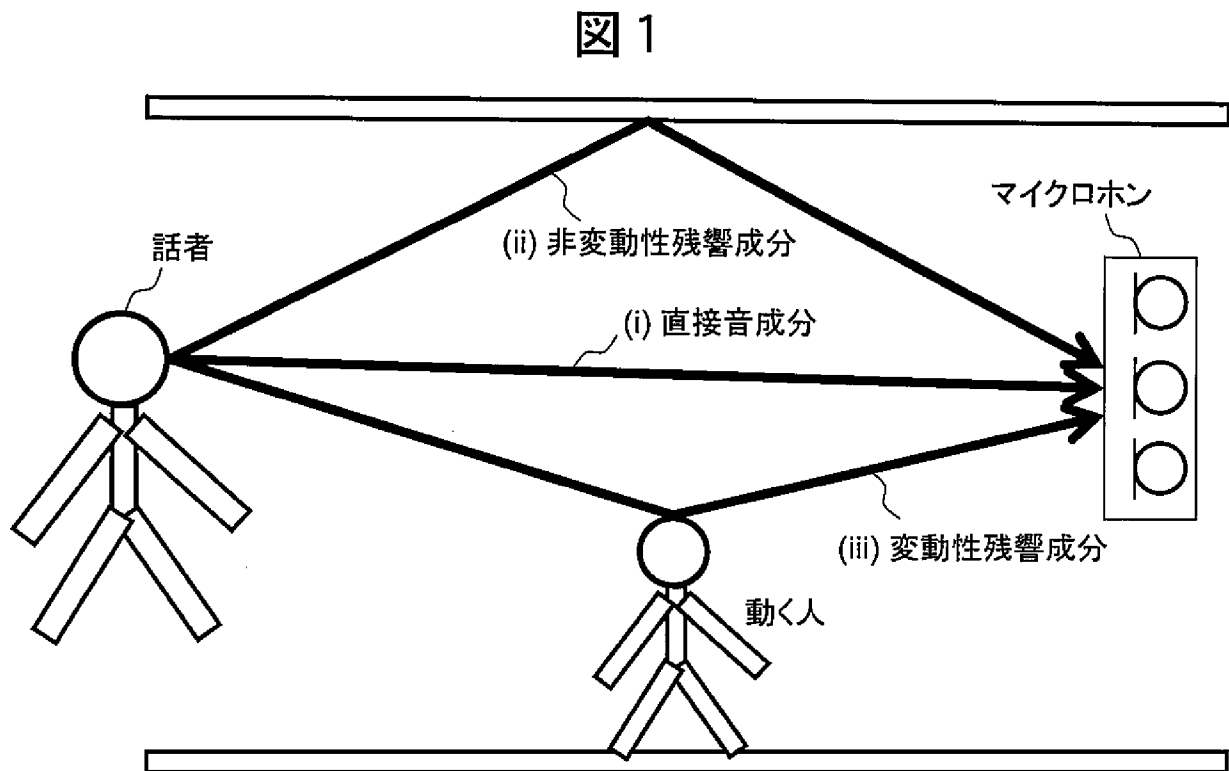
各拠点に配置された複数の拠点毎会議システムと、当該複数の拠点毎会議システムを管理する計算サーバと、を有するオンライン会議システムであって、

前記計算サーバは、請求項 1 に記載の残響除去パラメータ推定装置を有し、それぞれの拠点毎会議システムから受信する音声入力信号を前記残響除去パラメータ推定装置によって処理し、それぞれの拠点毎会議システムに対応した残響除去パラメータを算出し、当該算残響パラメータを該当する拠点毎会議システムに送信し、

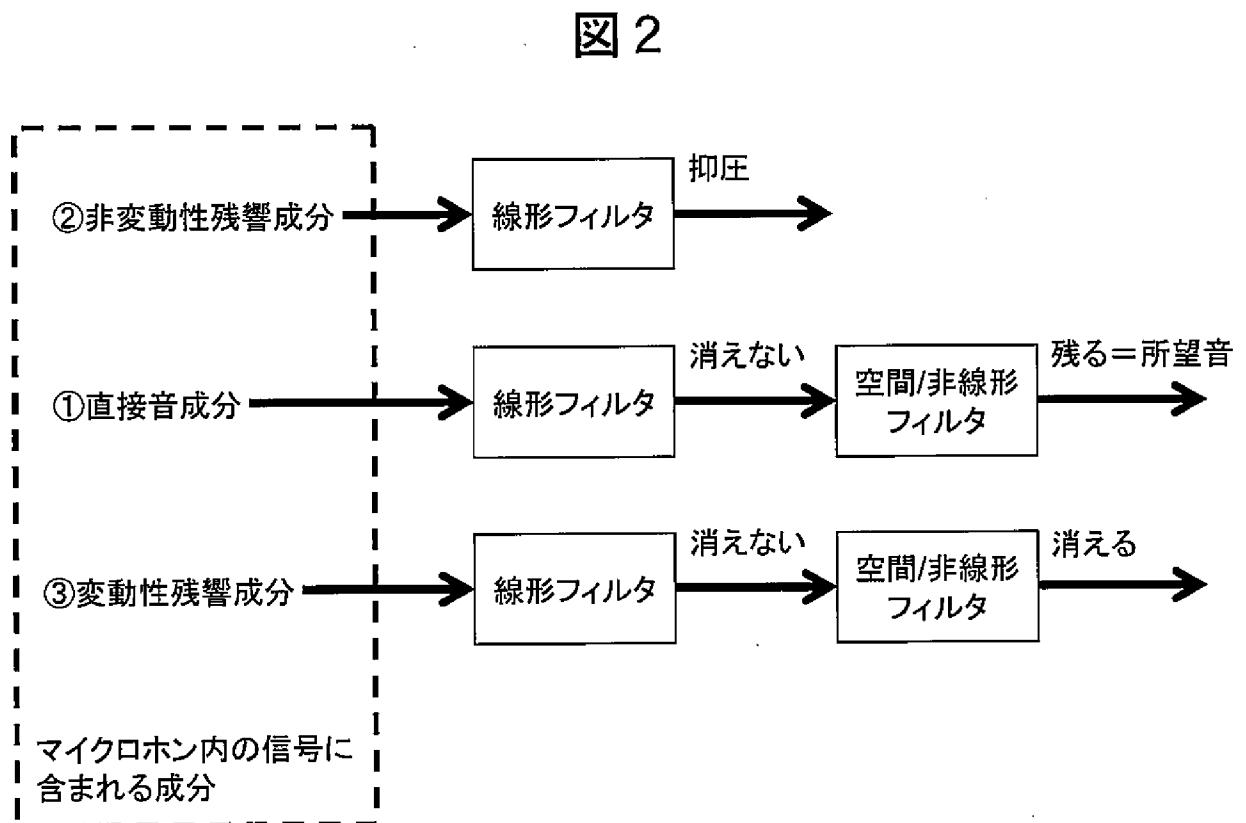
前記複数の拠点毎会議システムのそれぞれの、音声入力信号を前記計算サーバに送信し、前記計算サーバから前記送信した音声入力信号に対応する残響除去パラメータを受信し、当該受信した残響パラメータを用いて、前記音声入力信号から前記非変動性残響成分及び前記変動性残響成分を分離して前記直接音成分をリアルタイムに出力する、

オンライン会議システム。

[図1]

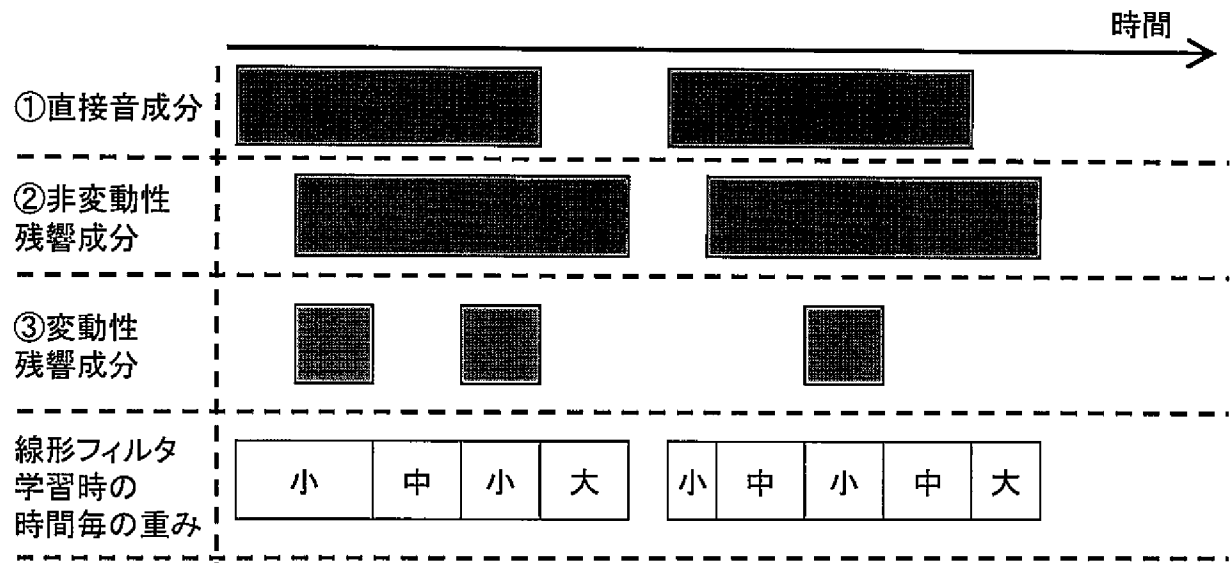


[図2]



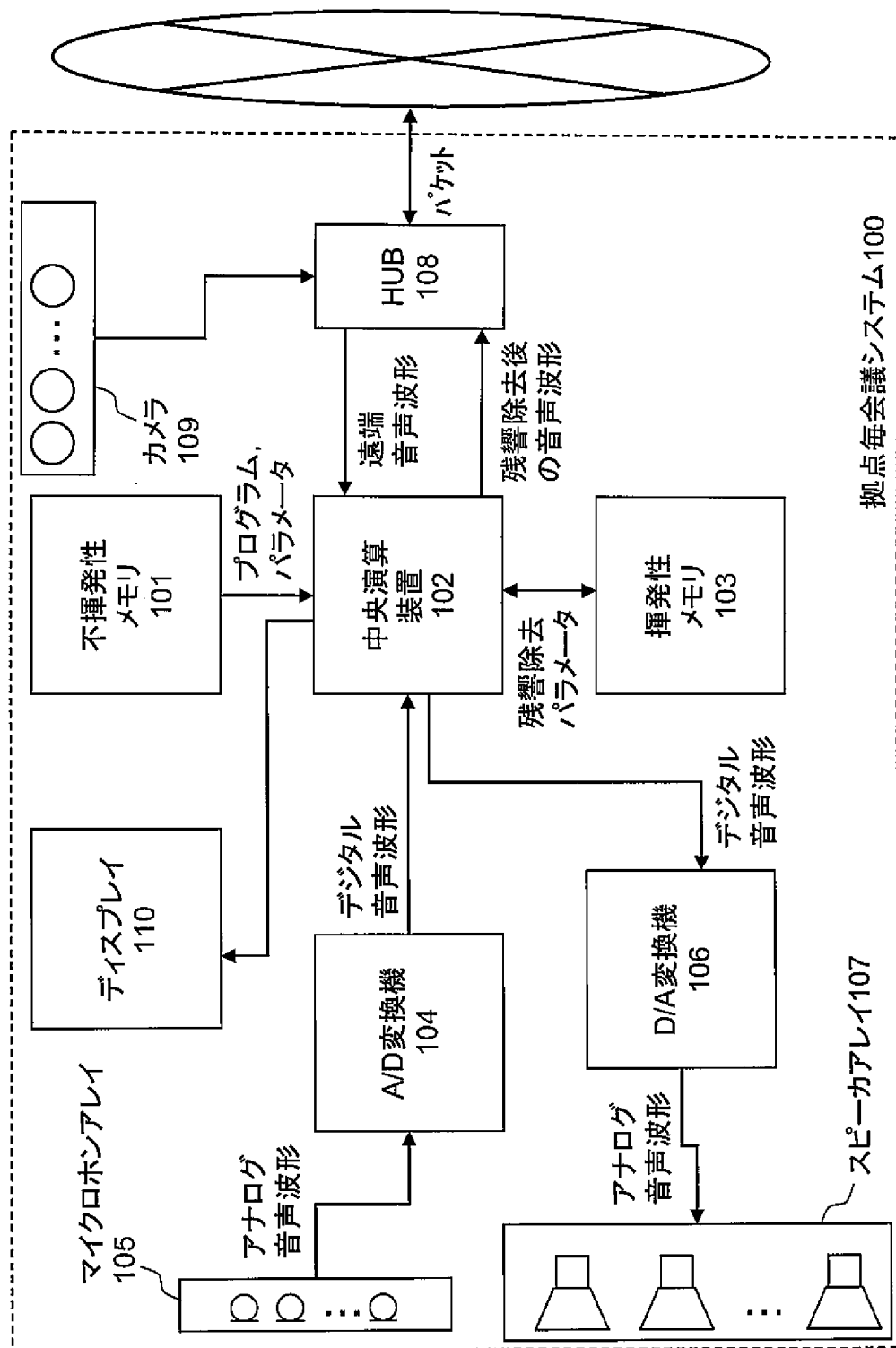
[図3]

図 3



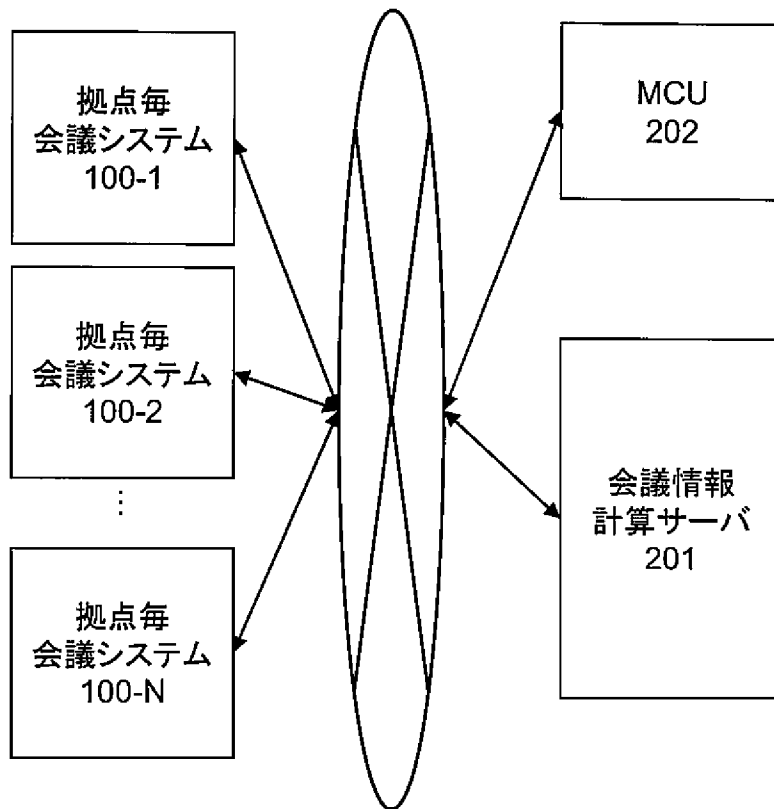
[図4]

図 4



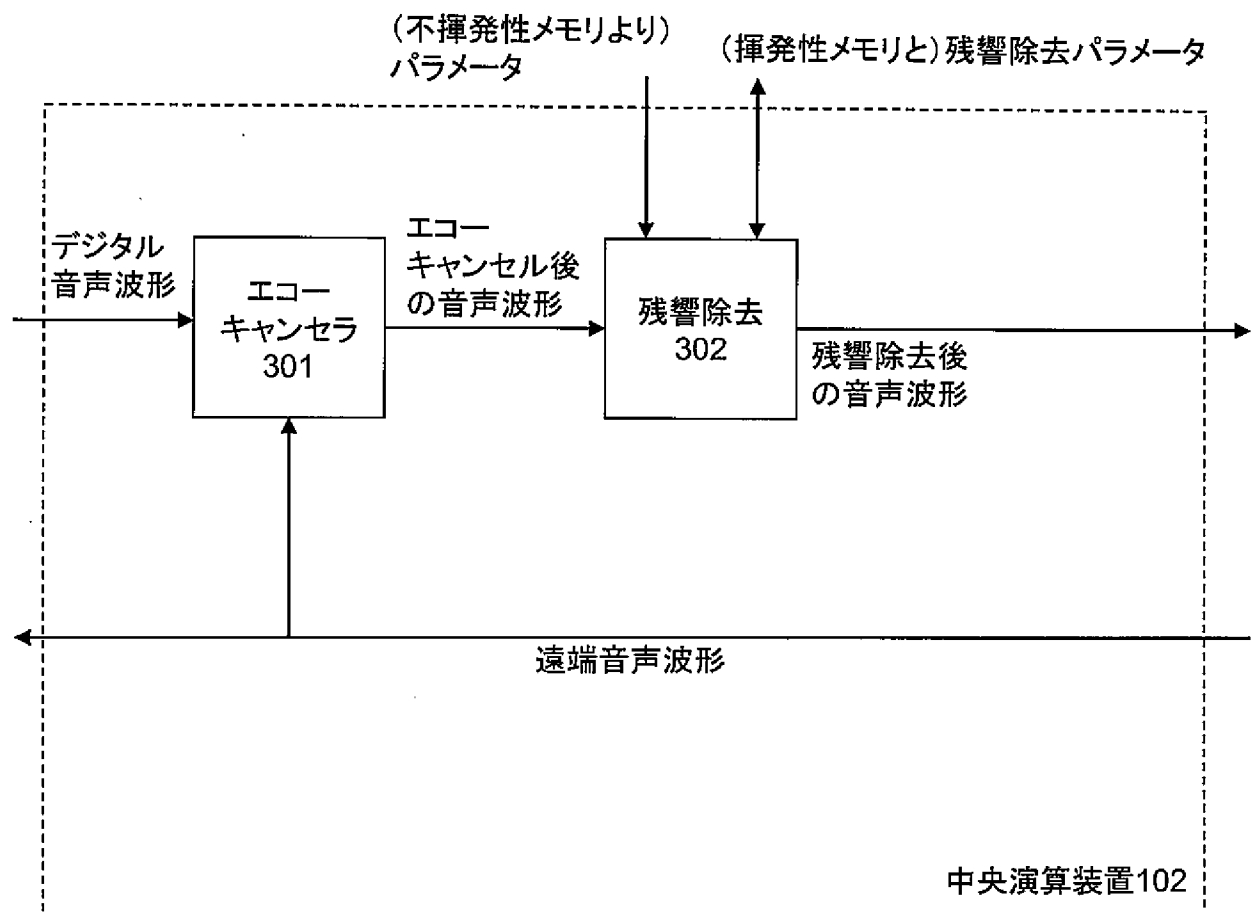
[図5]

図 5



[図6]

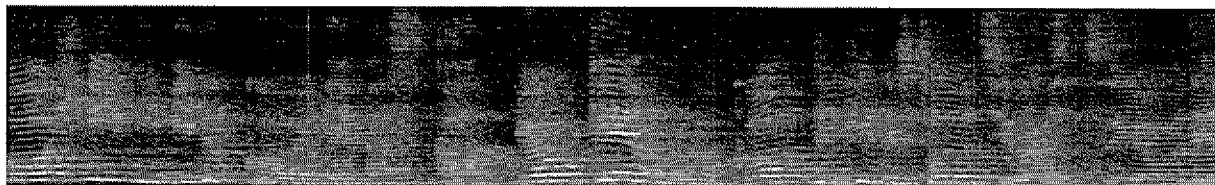
図 6



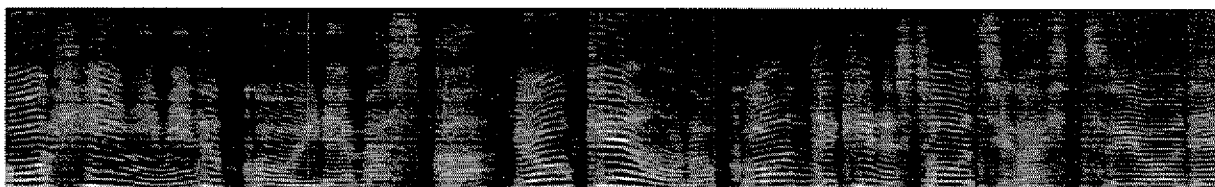
[図7]

図 7

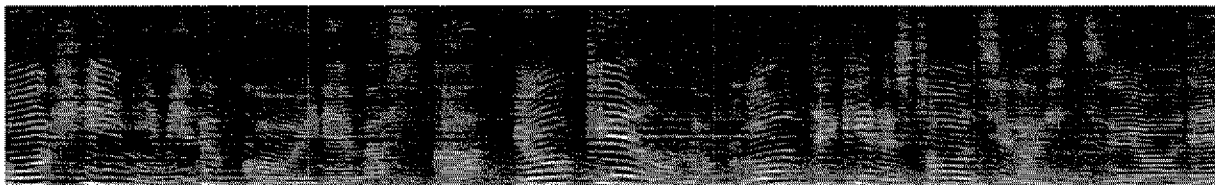
(i) マイク入力信号



(ii) 残響の無い信号(理想)



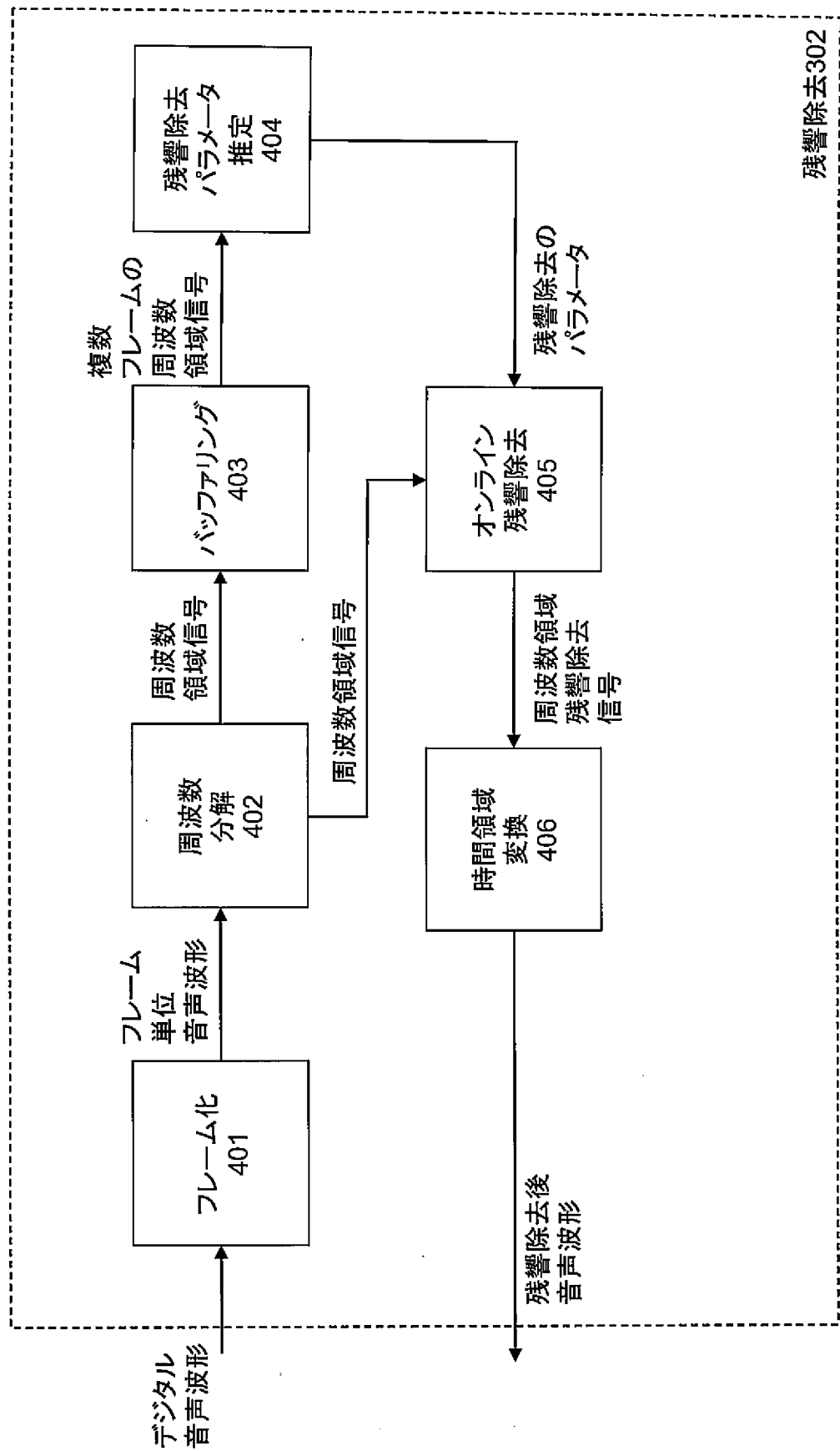
(iii) 残響除去後の信号



周波数
↑
時間 →

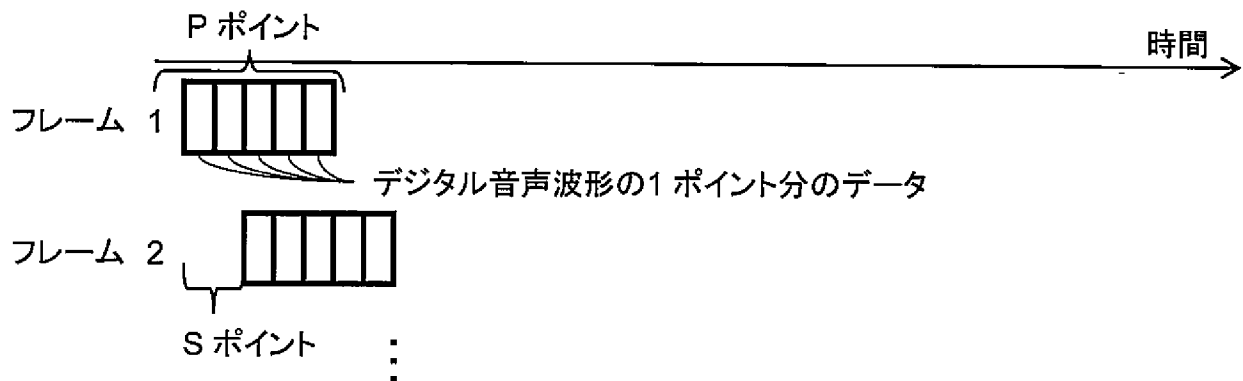
[図8]

図 8



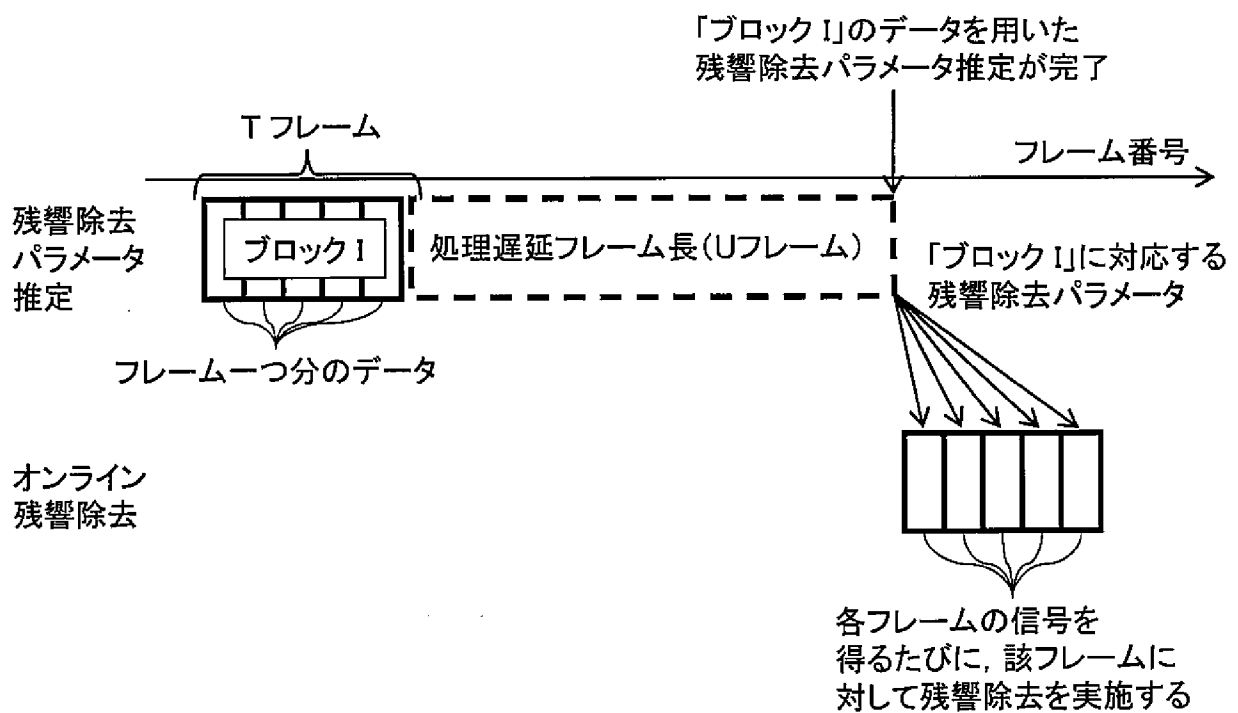
[図9]

図 9



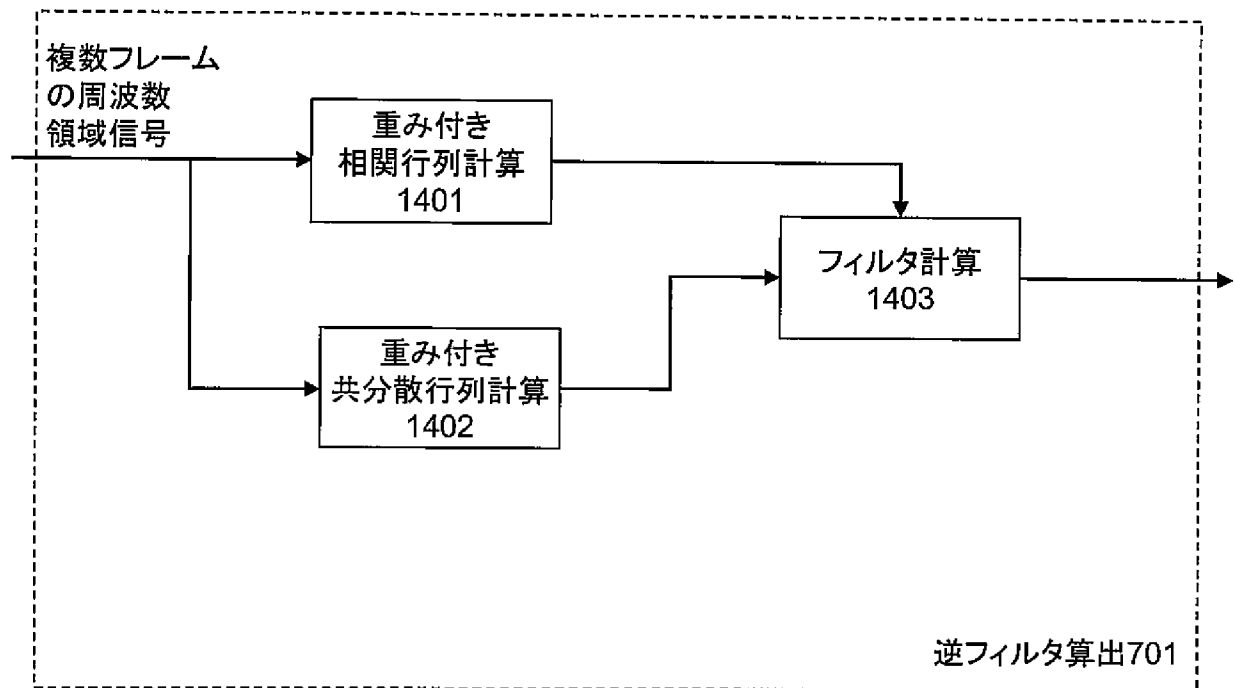
[図10]

図 1 0



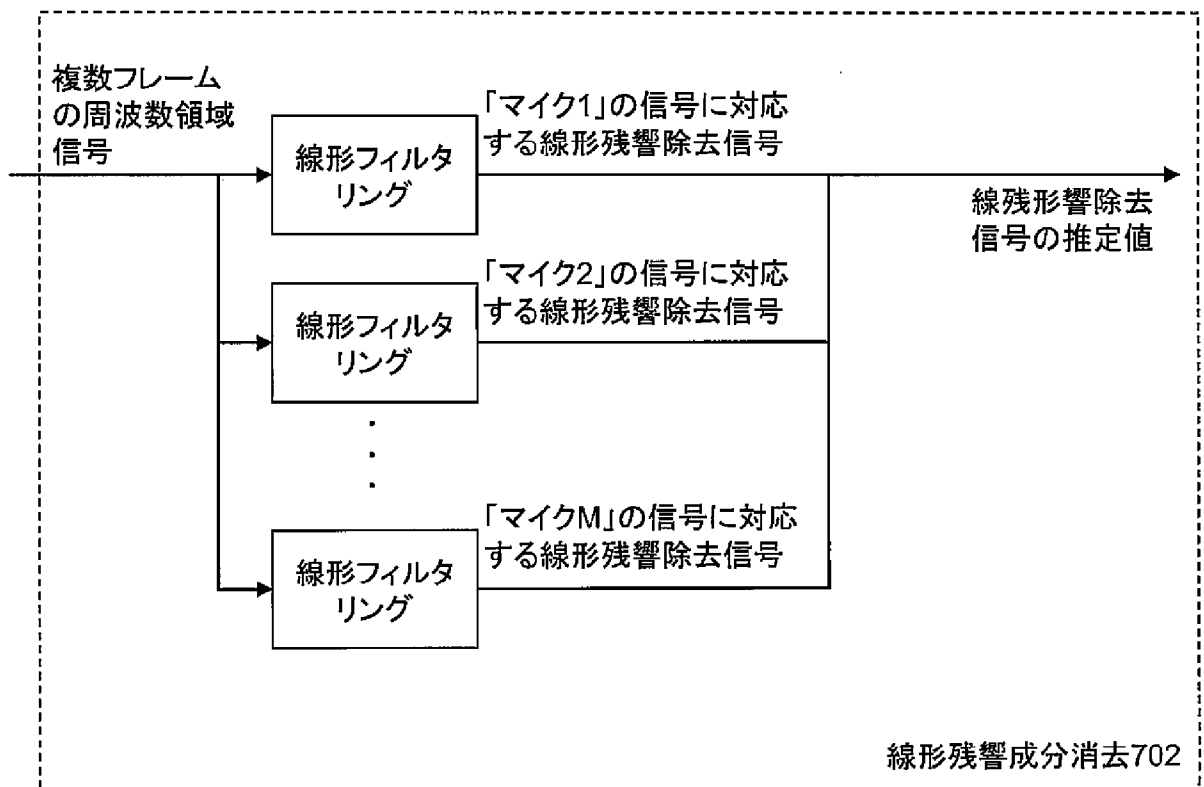
[図12]

図 1 2



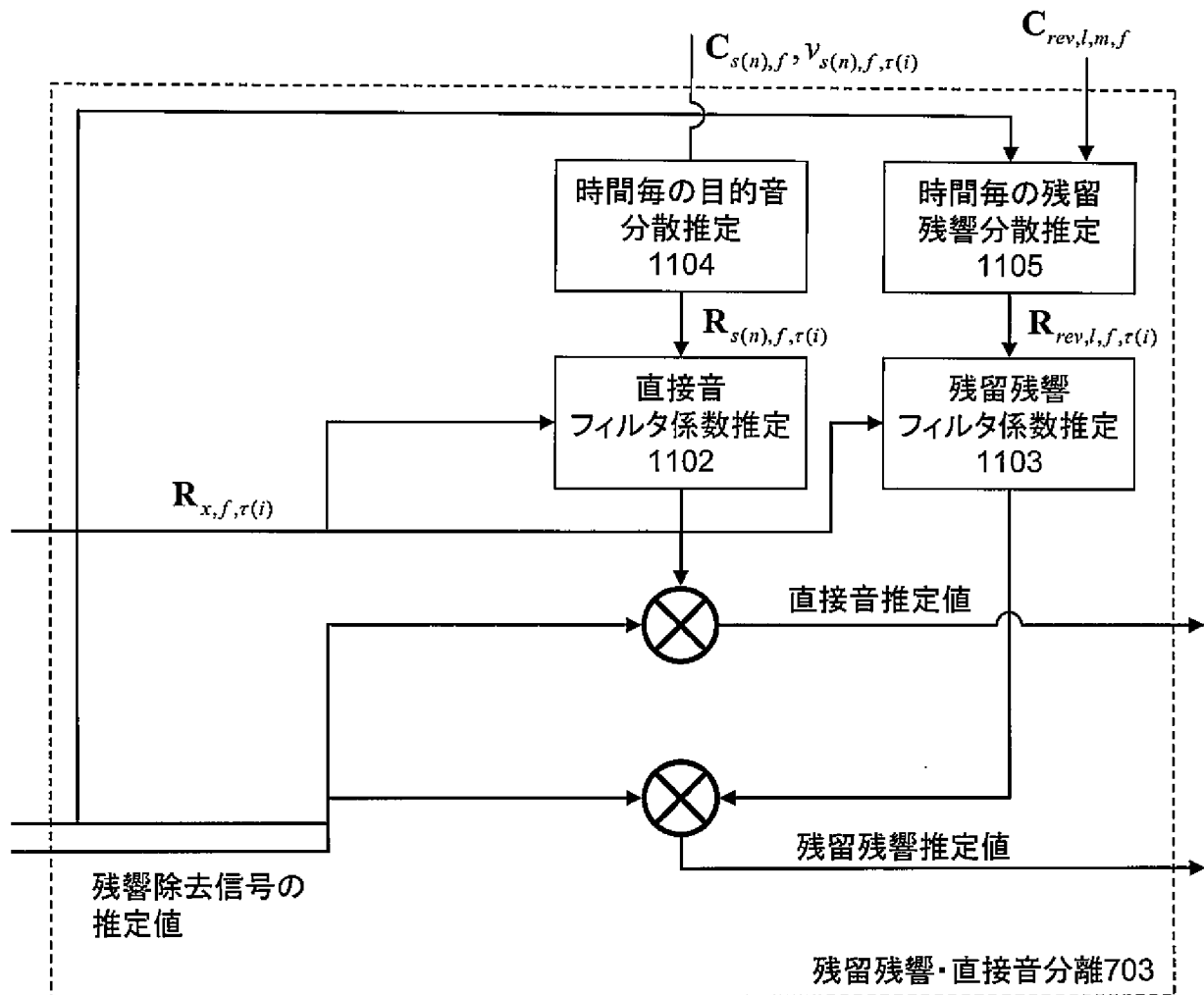
[図13]

図 1 3



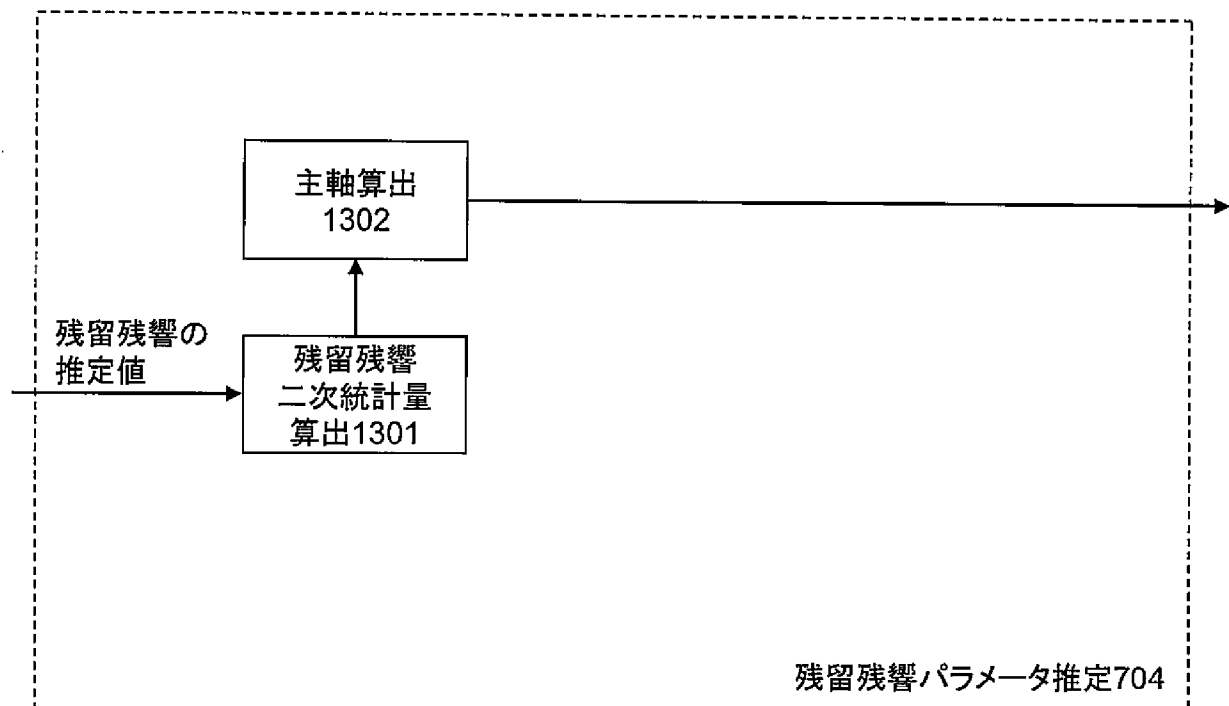
[図14]

図 1 4



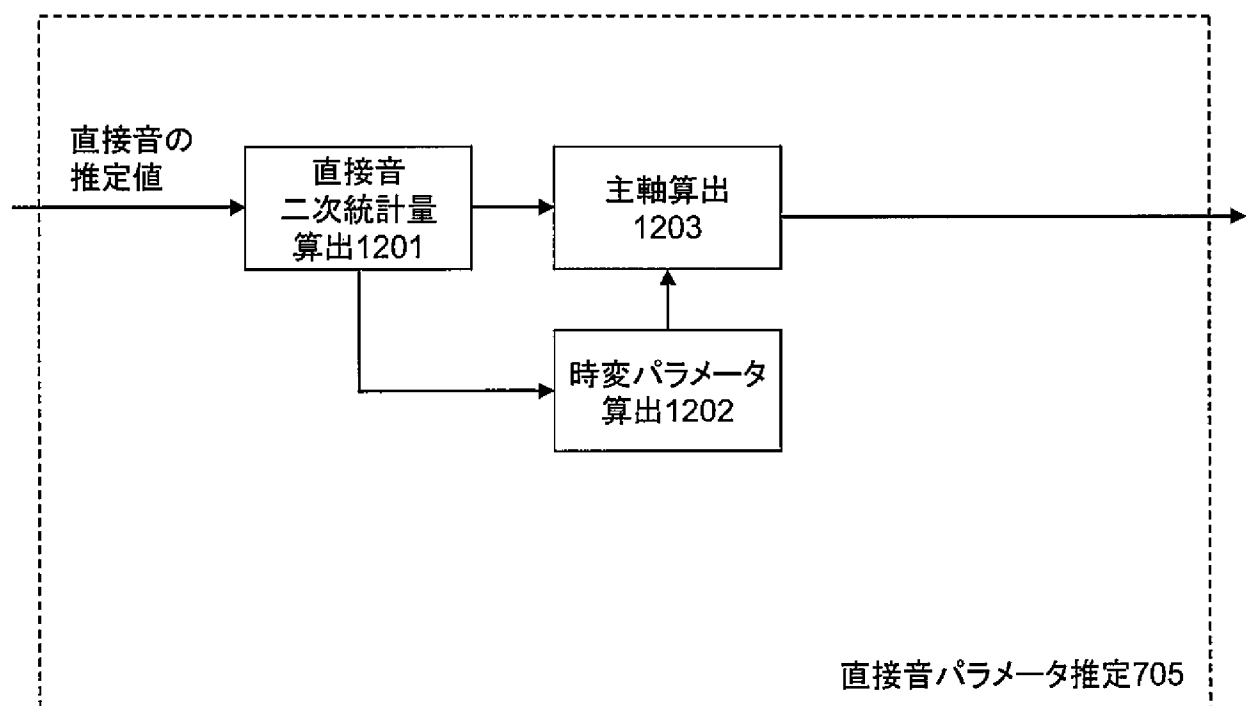
[図15]

図 1 5



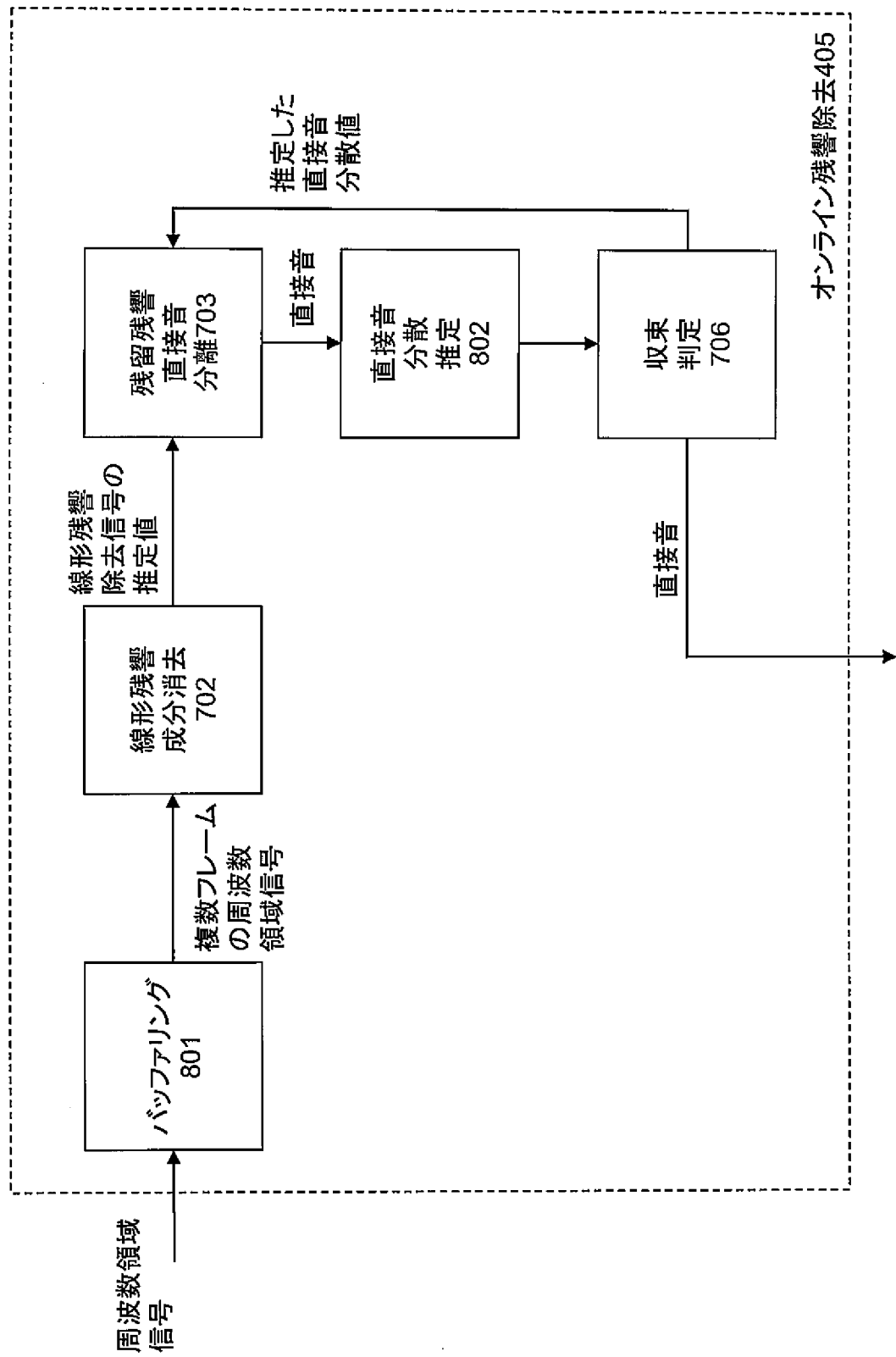
[図16]

図 1 6

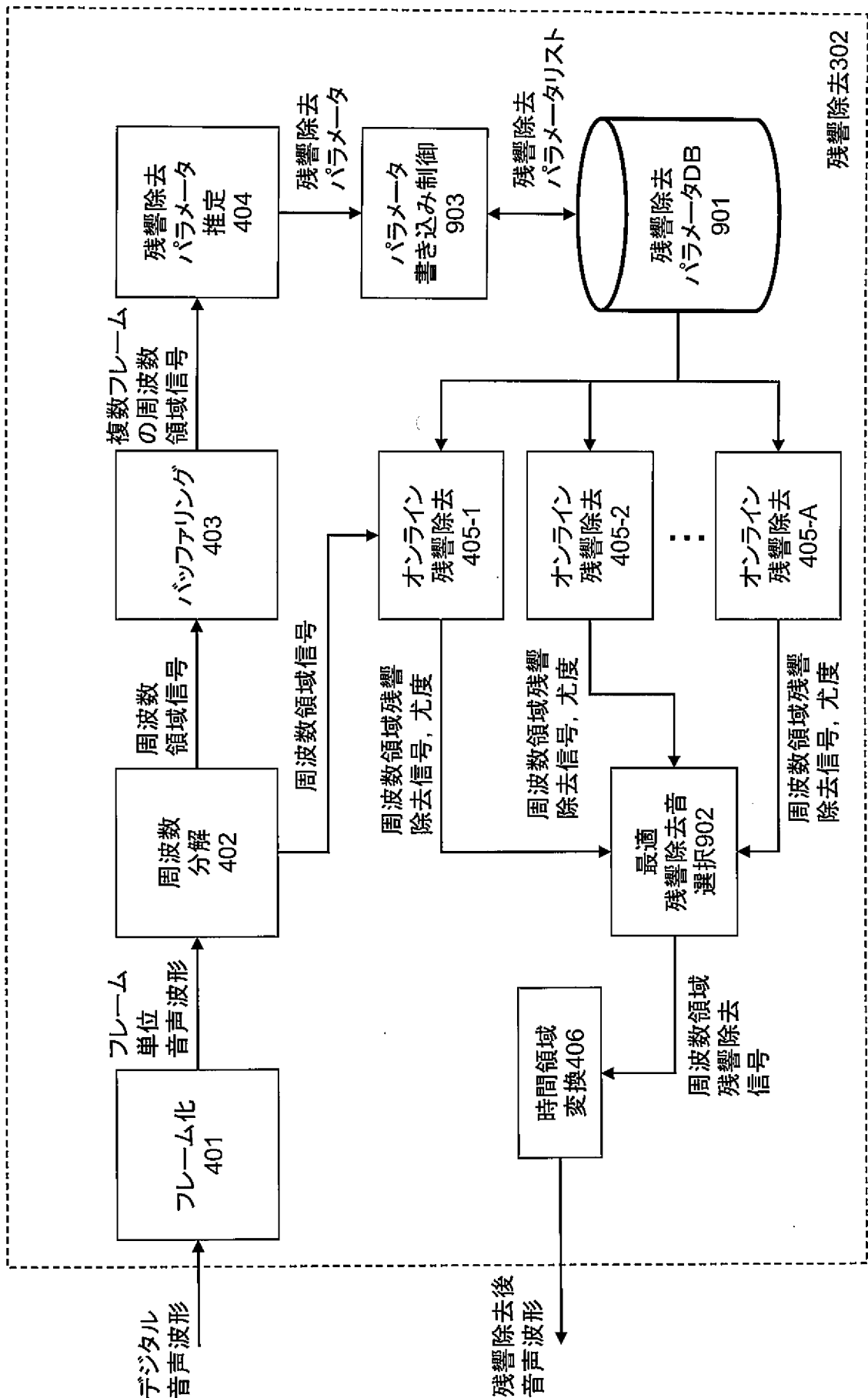


[図17]

図17

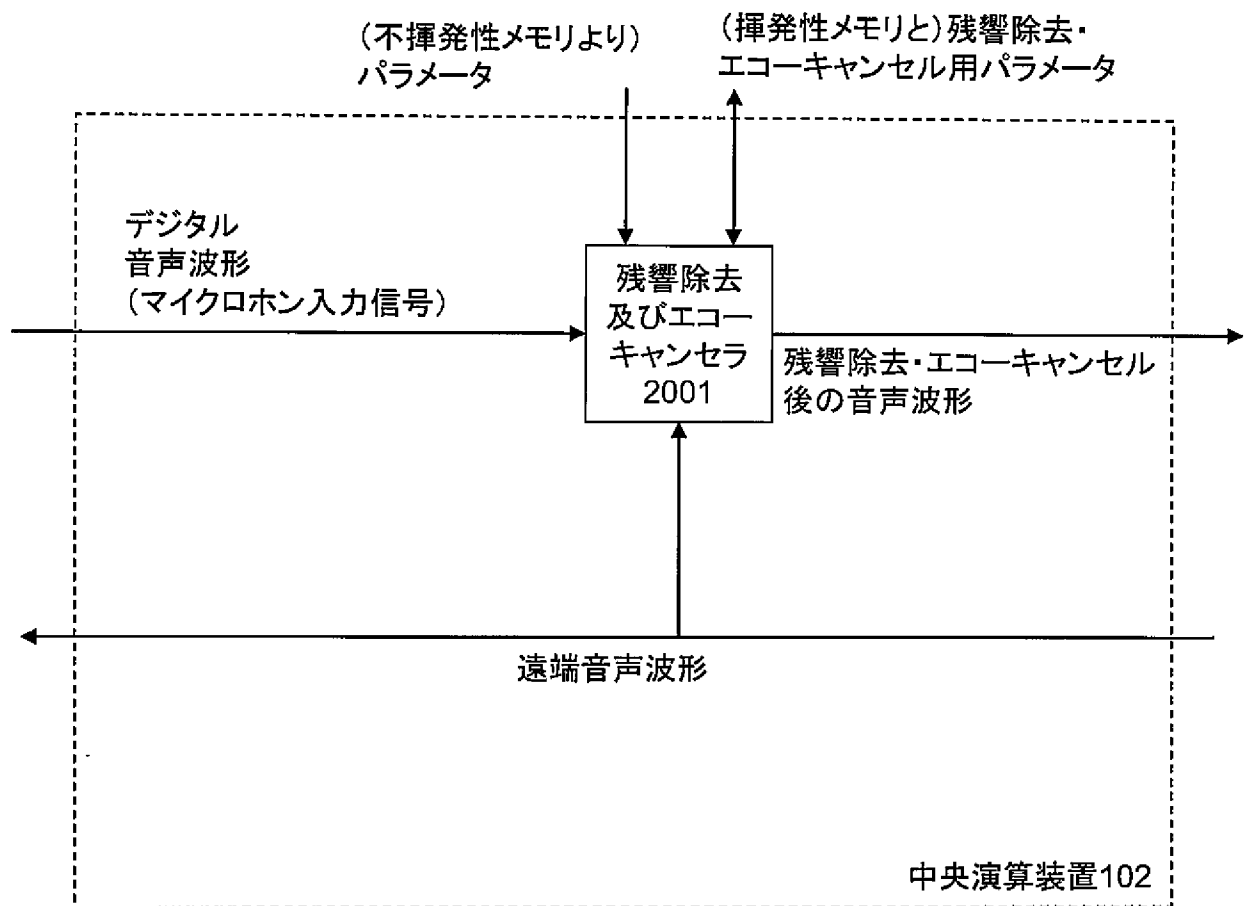


[図18]



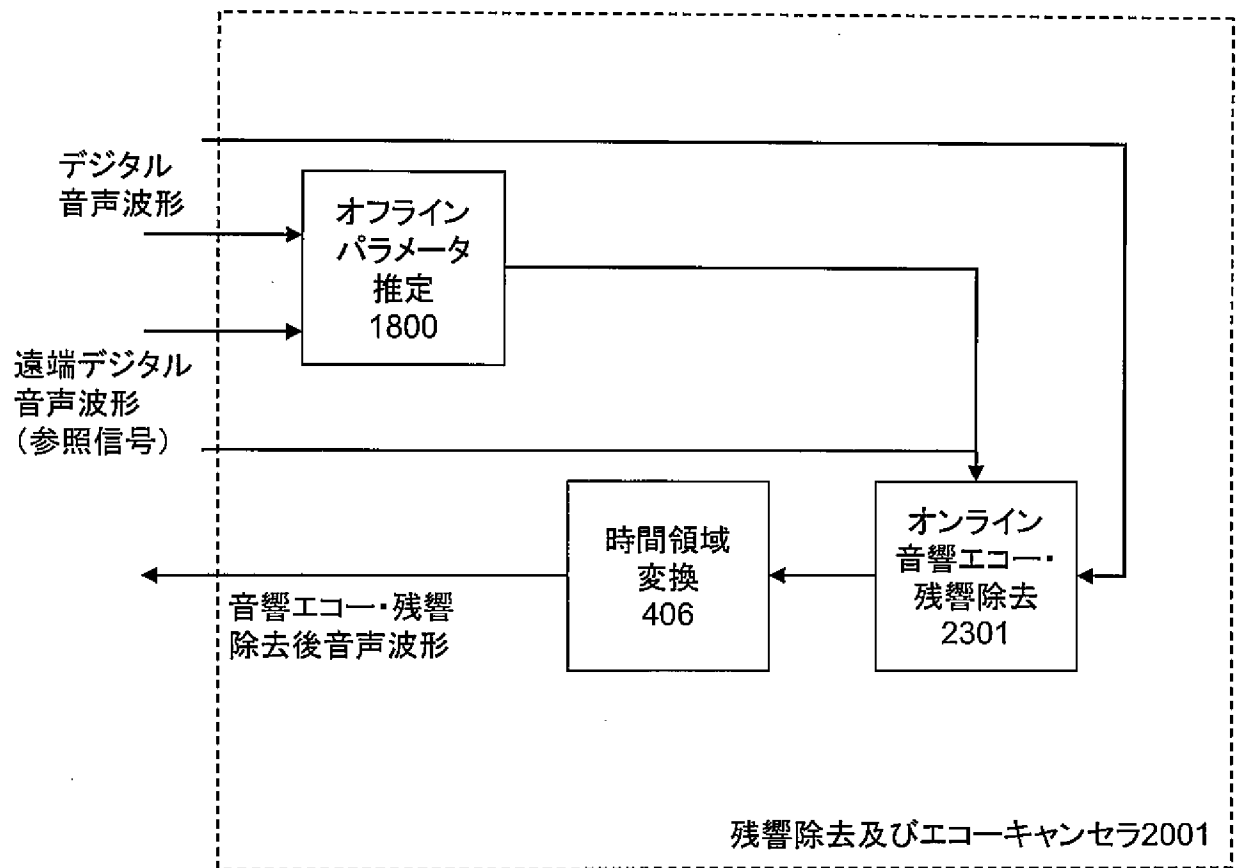
[図19]

図 1 9



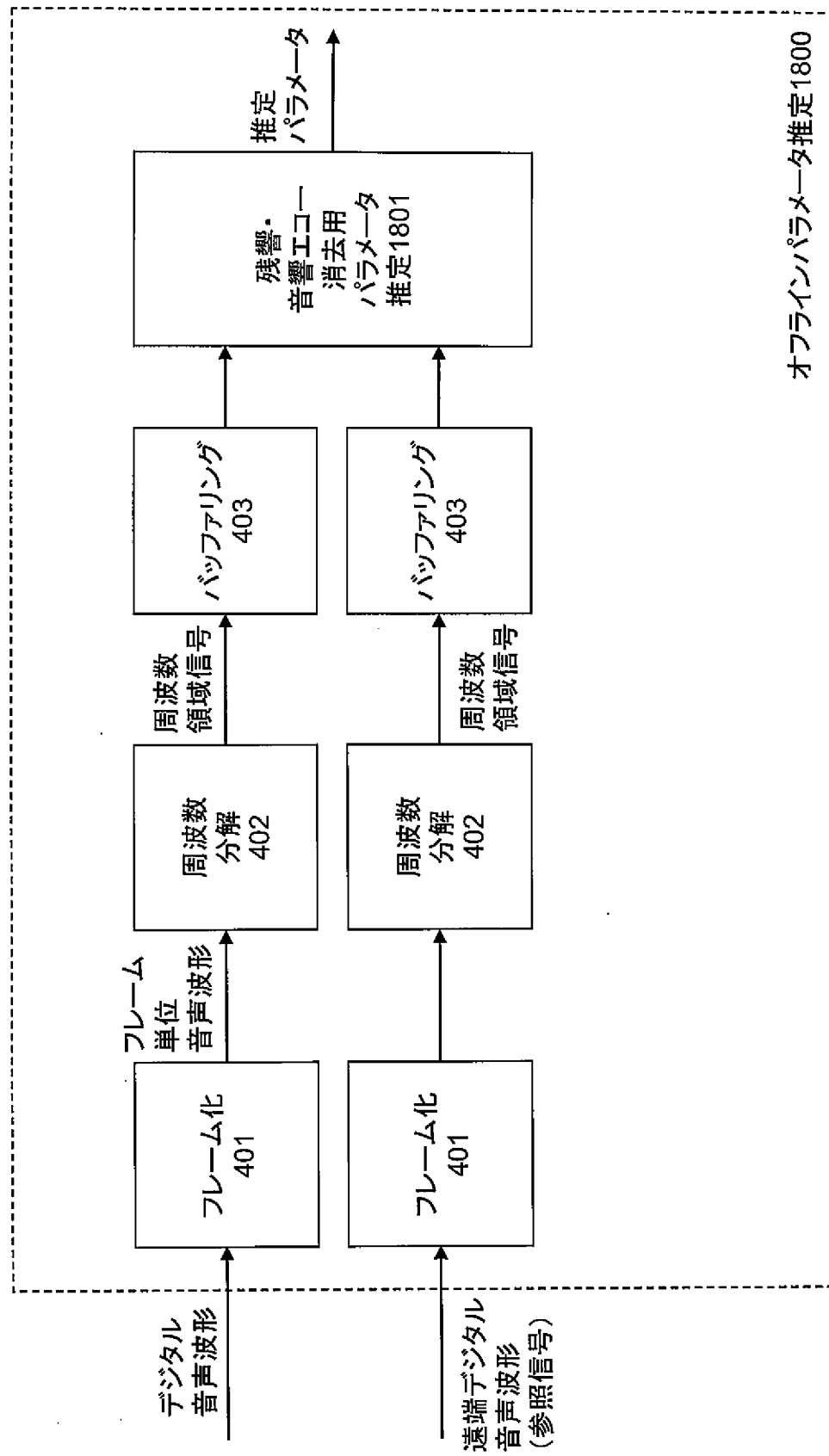
[図20]

図 20



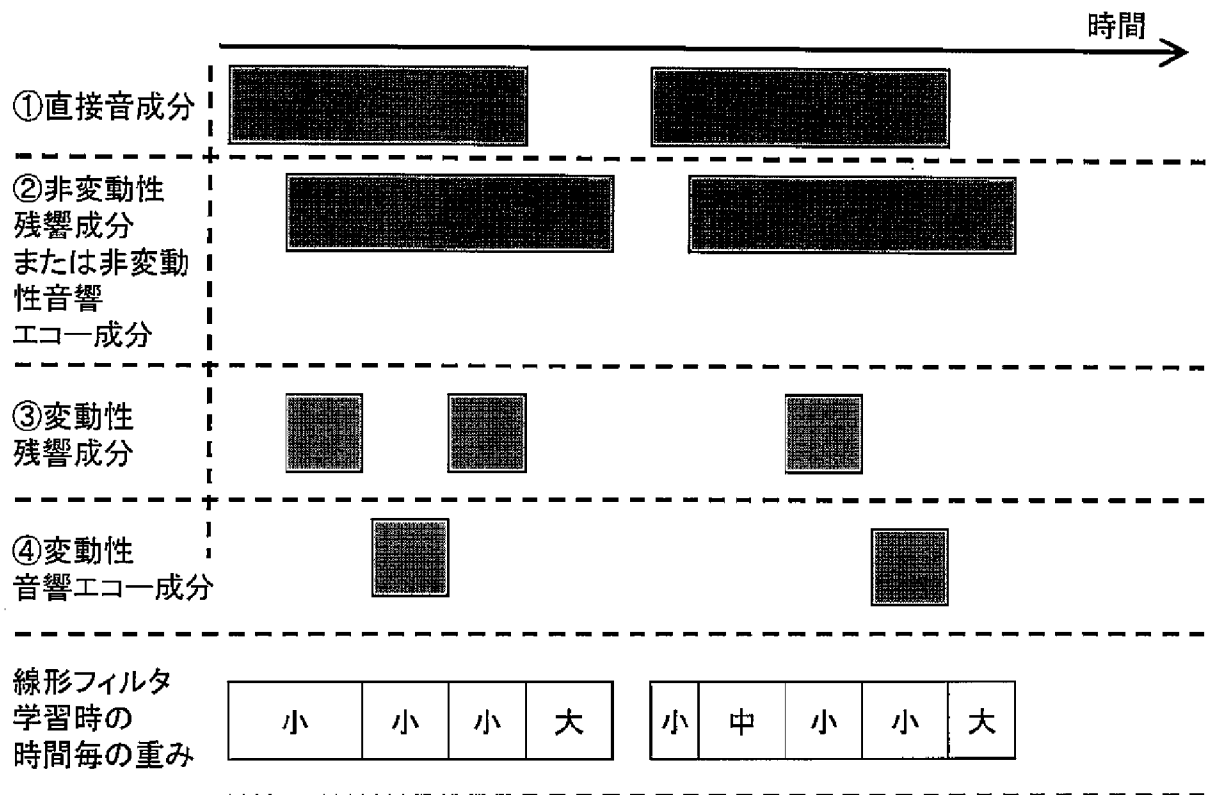
[図21]

図 21



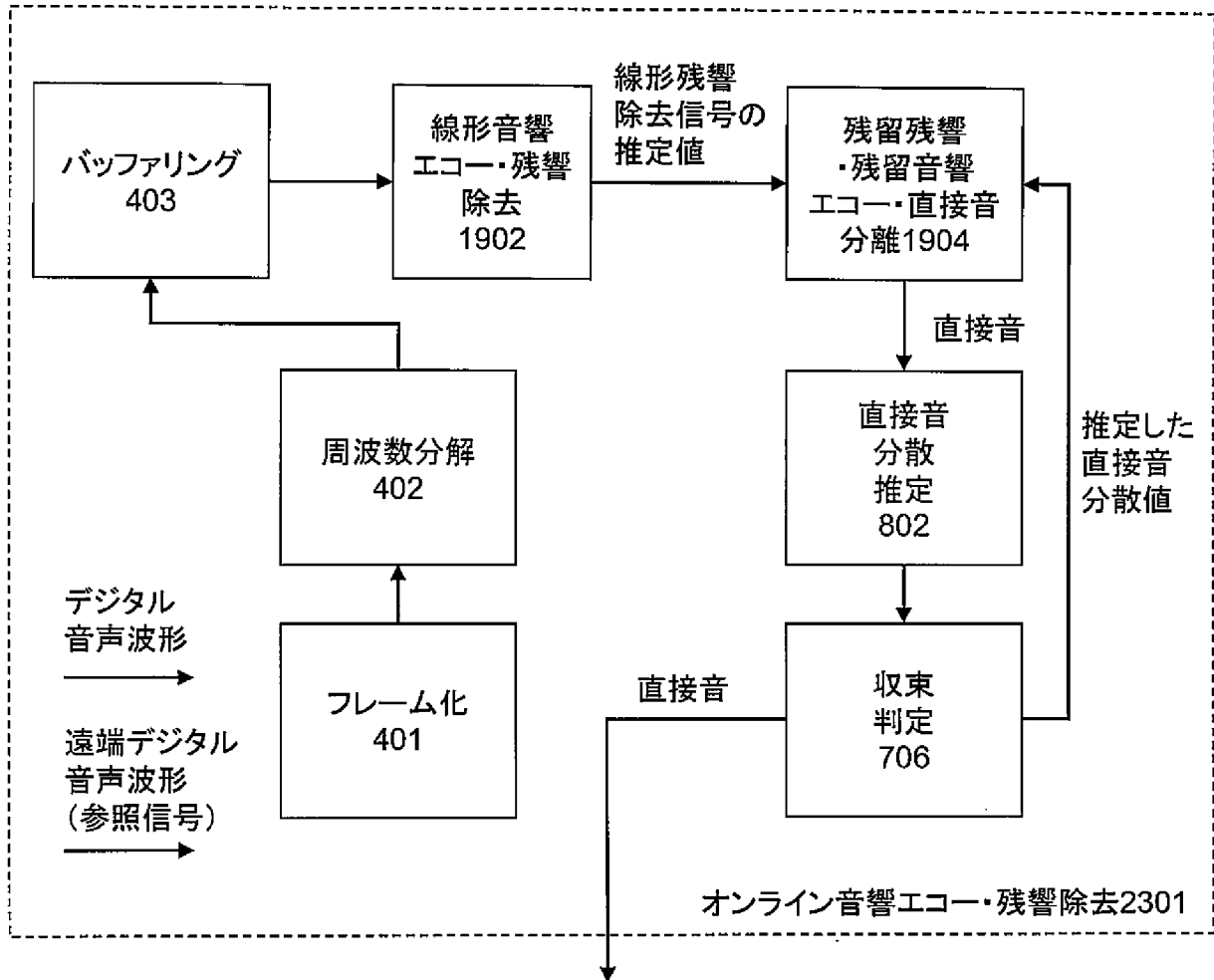
[図23]

図 2 3



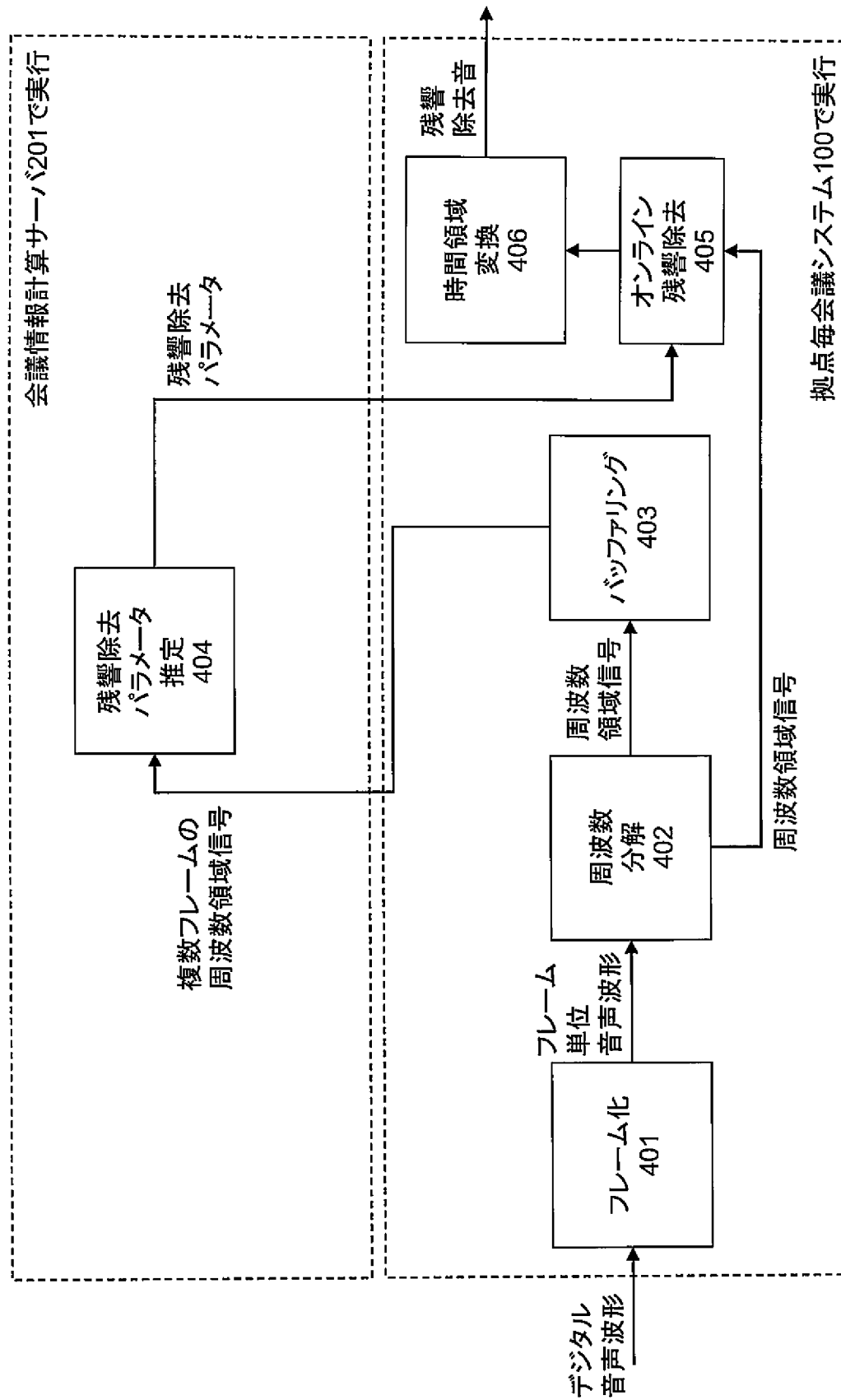
[図24]

図 2 4



[図25]

図 2 5



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/053645

A. CLASSIFICATION OF SUBJECT MATTER

G10L21/0208(2013.01)i, G10L21/0216(2013.01)i, H04B3/23(2006.01)i, H04R3/02(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L21/0208, G10L21/0216, H04B3/23, H04R3/02

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2013
Kokai Jitsuyo Shinan Koho	1971-2013	Toroku Jitsuyo Shinan Koho	1994-2013

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2006-157498 A (Matsushita Electric Industrial Co., Ltd.), 15 June 2006 (15.06.2006), entire text; all drawings (Family: none)	1-12
A	JP 62-130029 A (Matsushita Electric Industrial Co., Ltd.), 12 June 1987 (12.06.1987), entire text; all drawings (Family: none)	1-12
A	JP 63-19924 A (Matsushita Electric Industrial Co., Ltd.), 27 January 1988 (27.01.1988), entire text; all drawings (Family: none)	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
28 February, 2013 (28.02.13)

Date of mailing of the international search report
12 March, 2013 (12.03.13)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (I P C))

Int.Cl. G10L21/0208(2013.01)i, G10L21/0216(2013.01)i, H04B3/23(2006.01)i, H04R3/02(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (I P C))

Int.Cl. G10L21/0208, G10L21/0216, H04B3/23, H04R3/02

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 3 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 3 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 3 年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2006-157498 A (松下電器産業株式会社) 2006.06.15, 全文・全図 (ファミリーなし)	1 - 1 2
A	JP 62-130029 A (松下電器産業株式会社) 1987.06.12, 全文・全図 (ファミリーなし)	1 - 1 2
A	JP 63-19924 A (松下電器産業株式会社) 1988.01.27, 全文・全図 (ファミリーなし)	1 - 1 2

C 欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

2 8 . 0 2 . 2 0 1 3

国際調査報告の発送日

1 2 . 0 3 . 2 0 1 3

国際調査機関の名称及びあて先

日本国特許庁 (I S A / J P)
郵便番号 1 0 0 - 8 9 1 5
東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

安田 勇太

5 Z

4 0 6 6

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 9 1