(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
2004 年12 月9 日 (09.12.2004)

PCT

(10) 国際公開番号
WO 2004/107319 A1

(51) 国際特許分類[7]: G10L 21/02

(21) 国際出願番号: PCT/JP2004/007587

(22) 国際出願日: 2004 年5 月26 日 (26.05.2004)

(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2003-154964 2003 年5 月30 日 (30.05.2003) JP
特願2003-167118 2003 年6 月11 日 (11.06.2003) JP

(71) 出願人 (米国を除く全ての指定国について): 独立行政法人産業技術総合研究所 (NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY) [JP/JP]; 〒1008921 東京都千代田区霞が関１丁目３番１号 Tokyo (JP).

(72) 発明者; および

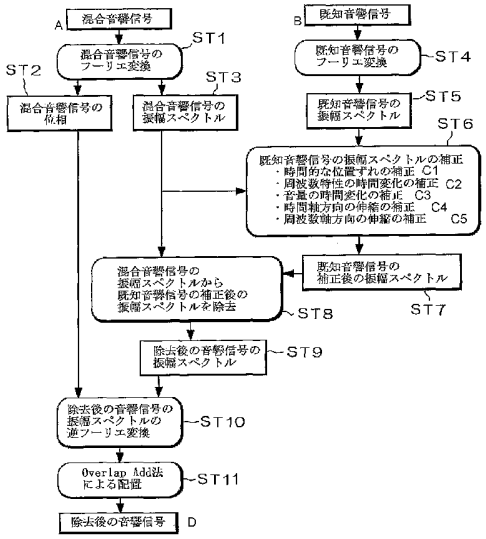
(75) 発明者/出願人 (米国についてのみ): 後藤 真孝 (GOTO, Masataka) [JP/JP]; 〒3058568 茨城県つくば市梅園１－１－１ 中央第２ 独立行政法人産業技術総合研究所内 Ibaraki (JP).

(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, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG,

[続葉有]

(54) Title: METHOD AND DEVICE FOR REMOVING KNOWN ACOUSTIC SIGNAL

(54) 発明の名称: 既知音響信号除去方法及び装置



A...MIXED ACOUSTIC SIGNAL
ST1...FOURIER TRANSFORM OF MIXED ACOUSTIC SIGNAL
ST2...PHASE OF MIXED ACOUSTIC SIGNAL
ST3...AMPLITUDE SPECTRUM OF MIXED ACOUSTIC SIGNAL
B...KNOWN ACOUSTIC SIGNAL
ST4...FOURIER TRANSFORM OF KNOWN ACOUSTIC SIGNAL
ST5...AMPLITUDE SPECTRUM OF KNOWN ACOUSTIC SIGNAL
ST6...CORRECTION OF AMPLITUDE SPECTRUM OF KNOWN ACOUSTIC SIGNAL
C1...CORRECTION OF POSITIONAL SHIFT BY TIME
C2...CORRECTION OF TEMPORAL CHANGE OF FREQUENCY CHARACTERISTIC
C3...CORRECTION OF TEMPORAL CHANGE OF SOUND VOLUME
C4...CORRECTION OF EXPANSION/CONTRACTION IN THE TEMPORAL AXIS DIRECTION
C5...CORRECTION OF EXPANSION/CONTRACTION IN THE FREQUENCY AXIS DIRECTION
ST8...REMOVE AMPLITUDE SPECTRUM AFTER CORRECTION OF THE KNOWN ACOUSTIC SIGNAL FROM THE AMPLITUDE SPECTRUM OF MIXED ACOUSTIC SIGNAL
ST7...AMPLITUDE SPECTRUM OF KNOWN ACOUSTIC SIGNAL AFTER CORRECTION
ST9...AMPLITUDE SPECTRUM OF ACOUSTIC SIGNAL AFTER THE REMOVAL
ST10...REVERSE FOURIER TRANSFORM OF AMPLITUDE SPECTRUM OF ACOUSTIC SIGNAL AFTER THE REMOVAL
ST11...ARRANGEMENT BY Overlap Add METHOD
D...ACOUSTIC SIGNAL AFTER THE REMOVAL

(57) Abstract: There is provided a known acoustic signal removal device which receives an acoustic signal as a mixture of a plurality of acoustic signals and can remove a known acoustic signal when the known acoustic signal similar to one of the acoustic signals is given. The known acoustic signal removal device converts an input mixed acoustic signal m(t) and a known acoustic signal b' (t) into an amplitude spectra M(ω, t) and B' (ω, t) in the respective time frequency areas and subtracts a component corresponding to the B' (ω, t) in the M(ω, t) so as to be removed, thereby obtaining an amplitude spectrum S(ω, t) after the removal. Here, the component corresponding to the B' (ω, t) in the M(ω, t) is deformed by various factors such as a positional shift by time, temporal change of frequency characteristic, and temporal change of the sound volume. Accordingly, they are corrected to obtain B(ω, t), which is subtracted. Lastly, by using the phase of the m(t) and S(ω, t), reverse conversion is performed in the temporal region so as to obtain an acoustic signal s(t) after a desired removal.

(57) 要約: 複数の音響信号が混合された音響信号を入力とし、そのうち一つの音響信号に類似した既知音響信号が与えられたときに、その既知音響信号を除去することを可能にする既知音響信号除去装置が提供される。この既知音響信号除去装置は、入力混合音響信号ｍ（ｔ）と既知音響信号ｂ'（ｔ）をそれぞれ時間周波数領域での振幅スペクトルＭ（ω，ｔ），Ｂ'（ω，ｔ）に変換し、Ｍ（ω，ｔ）中のＢ'（ω，ｔ）に対応する成分を減算して除去することにより、除去後の振幅スペクトルＳ（ω，ｔ）を得る。その際、Ｍ（ω，ｔ）中のＢ'（ω，ｔ）に対応する成分は、時間的な位置のずれ、周波数特性の時間変化、音量の時間変化等の要因により変形しているため、それらを補正したＢ（ω，ｔ）を減算する。最後に、ｍ（ｔ）の位相と、Ｓ（ω，ｔ）とを用いて時間領域に逆変換し、所望の除去後の音響信号ｓ（ｔ）を得る。

WO 2004/107319 A1

SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告書

２文字コード及び他の略語については、定期発行される各*PCT*ガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

明細書

既知音響信号除去方法及び装置

技術分野

本発明は、複数の音響信号が混合された混合音響信号の中から、既知の音響信号の成分を除去する既知音響信号除去方法および既知音響信号除去装置に関するものである。

背景技術

従来から、音響信号処理として、スペクトルサブトラクション法（非特許文献１）と呼ばれる方法が知られている。従来のスペクトルサブトラクション法は、定常雑音（スペクトルが時間的に変化せず、周波数特性や音量等がほぼ一定な雑音）と所望の音（ターゲット音）が混合された音響信号（混合音）から定常雑音を除去してターゲット音を得る方法である。この方法では、事前に定常なスペクトルの平均を求める等の簡易な方法で定常雑音のスペクトルを学習しておき、入力された混合音のスペクトルから定常雑音のスペクトルを引き去る処理を行う。つまり、雑音の平均を引き去る処理を行う。

一般に、音響信号除去に関しては、複数のマイクロホンからの入力を用いる方法が多数提案されている。また、スペクトルサブトラクション法には、特許文献１～７に開示されているように、様々な改良がなされている。

［非特許文献１］

Steven Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No. 2, April 1979.

[特許文献１]　特開２００２－１７５０９９号公報
[特許文献２]　特開２００２－０１４６９４号公報
[特許文献３]　特開２００１－２２８８９２号公報
[特許文献４]　特開２００１－２１５９９２号公報
[特許文献５]　特開平１１－００３０９４号公報
[特許文献６]　特開平１０－２４０２９４号公報
[特許文献７]　特開平０８－２２１０９２号公報

発明の開示

従来におけるスペクトルサブトラクション法は、定常雑音を前提としており、非定常雑音（スペクトルが時間的に大きく変化し、周波数特性や音量等も変化する雑音）には適用できなかった。例えば、バックグラウンドミュージック（ＢＧＭ）として使用されている音楽のような時間的に大きく変化する非定常雑音を除去することは不可能であった。これは、非定常雑音のスペクトルの変化が大きすぎて学習ができないからである。

また、仮に、従来の方法により非定常雑音が事前に与えられた条件を扱おうとしても、非定常雑音の周波数特性、音量、振幅スペクトルの時間軸方向の伸縮及び周波数軸方向の伸縮等の変化の影響で、引き去る処理を適切に行うことはできなかった。複数のマイクロホンからの入力を用いる方法では、モノラル音響信号には適用することができなかった。改良された従来のスペクトルサブトラクション法のいずれの方法も、主に音声認識の前処理を目的としているため、非定常雑音が事前に与えられ、その非定常雑音を除去する用途には利用できなかった。

したがって、本発明の目的は、複数の音響信号が混合された混合音響信号の中から、既知の音響信号（非定常でも定常でもよい）の成分を、それに対応する元音源からの既知音響信号を用いて除去することができる既知音響信号

除去方法及び既知音響信号除去装置並びに該装置に用いるプログラムを提供することにある。

また、本発明の他の目的は、例えば、既知の音響信号が音楽であり、その音楽音響信号が、人間の音声や物音に対するバックグラウンドミュージック（ＢＧＭ）として使用されている混合音から、既知の音響信号に対応する元音源である既知音響信号（例えばＣＤやレコード等から同一音楽の音響信号を別途入手したもの）を用いてＢＧＭを除去することができる既知音響信号除去方法及び装置並びに該装置に用いるプログラムを提供することにある。

本発明の更に別の目的は、複数の音響信号が混合された音響信号（混合音）の中から、既知音響信号の成分を除去する際に、混合音中での既知音響信号の正確な位置を自動推定し、その位置の既知の音響信号を除去することができる既知音響信号除去方法及び装置並びに該装置に用いるプログラムを提供することにある。

本発明の更に別の目的は、複数の音響信号が混合された音響信号（混合音）の中から、既知音響信号の成分を除去する際に、混合音中での既知音響信号の正確な位置を人間が指定できるインタフェースを備えた既知音響信号除去装置を提供することにある。

本発明の更に別の目的は、複数の音響信号が混合された音響信号（混合音）の中から、既知音響信号の成分を除去する際に、混合音中で既知音響信号の周波数特性や音量が時間的に変化しているときに、それらの変化を自動推定して補正しながら除去することができる既知音響信号除去方法及び装置並びに該装置に用いるプログラムを提供することにある。

本発明の更に別の目的は、複数の音響信号が混合された音響信号（混合音）の中から、既知音響信号の成分を除去する際に、混合音中で既知音響信号の周波数特性や音量が時間的に変化しているときに、それらの変化を人間が指定することができるインタフェースを備えた既知音響信号除去装置を提供することにある。

本発明の更に別の目的は、複数の音響信号が混合された音響信号（混合音）の中から、既知音響信号の成分を除去する際に、混合音中で既知の音響信号が時間軸あるいは周波数軸方向に伸縮しているときに、それらの伸縮を自動推定して補正しながら除去することができる既知音響信号除去方法及び装置並びに該装置に用いるプログラムを提供することにある。

本発明の更に別の目的は、複数の音響信号が混合された音響信号（混合音）の中から、既知音響信号の成分を除去する際に、混合音中で既知の音響信号が時間軸あるいは周波数軸方向に伸縮しているときに、それらの伸縮を人間が指定することができるインタフェースを備えた既知音響信号除去装置を提供することにある。

本発明の更に別の目的は、複数の音響信号が混合された音響信号の中から、複数の既知音響信号の成分を除去する際に、既知の音響信号を一つずつ繰り返し除去できるようにした既知音響信号除去方法及び装置並びに該装置に用いるプログラムを提供することにある。

本発明による既知音響信号除去方法においては、複数の音響信号が混合された混合音響信号から、既知の音響信号（非定常でも定常でもよい）の成分を、それに対応する元音源からの既知音響信号を用いて除去する。

このため、本発明の既知音響信号除去方法では、まず、混合音響信号を時間周波数表現に変換して混合音響信号の振幅スペクトルと混合音響信号の位相とを求める（混合音響信号変換ステップ）。ここでの音響信号を時間周波数表現に変換する方法としては、フーリエ変換やウェーブレット変換など公知の変換方法を用いる。

次に、混合音響信号中に含まれている既知の音響信号に対応（類似）している既知音響信号（ＣＤやレコード等から同一音楽の音響信号を別途入手したもの）を時間周波数表現に変換して既知音響信号の振幅スペクトルを求める（既知音響信号変換ステップ）。

続いて、求めた混合音響信号の振幅スペクトルに基づいて、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した前記既知音響信号の補正振幅スペクトルを求める（補正ステップ）。

次に、混合音響信号の振幅スペクトルから既知音響信号の補正振幅スペクトルを除去する（除去ステップ）。この除去ステップにより得た除去後振幅スペクトルと混合音響信号の位相とに基づいて時間表現に逆変換を行って単位波形を求める（逆変換ステップ）。

最後に、単位波形をオーバーラップ・アド法等の合成方法を用いて合成して既知音響信号の成分を除去した音響信号を得る（合成ステップ）。

また、本発明の既知音響信号除去方法においては、次のような補正ステップを実行することにより、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した既知音響信号の補正振幅スペクトルを求め、この補正振幅スペクトルを混合音響信号の振幅スペクトルから除去する。このため、混合音響信号中に非定常雑音として含まれている既知音響信号を高い精度で除去することができる。

原則的には、既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の中で、実際に混合音響信号中でその現象または変化が起きていたものを全て補正するのが好ましい。

しかしながら、何も補正しない場合よりも、実際に混合音響信号中でその現象または変化が起きているものの１つでも補正すれば、既知音響信号の除去精度を高めることができるので、補正のすべてを行わなくてもよい。もちろん必要な補正のすべてを行ってもよい。

補正ステップの実行では、例えば、混合音響信号に含まれる既知音響信号の時間的な位置を推定し、推定した時間的な位置に基づいて既知音響信号の振幅スペクトルの時間的な位置のずれを補正する。推定方法は、例えば、混合音響信号の振幅スペクトルの所定の区間と既知音響信号の振幅スペクトルの所定の区間の距離（類似度）を求め、距離が最も近い区間を混合音響信号に含まれる既知音響信号の時間的な位置と推定する。

また、補正ステップの実行では、例えば、混合音響信号に含まれる既知音響信号の周波数特性の変化を推定し、推定した周波数特性の時間変化に基づいて既知音響信号の振幅スペクトルの周波数特性の時間変化を補正する。周波数特性の変化の推定は、例えば、混合音響信号中の既知の音響信号だけが含まれている区間を特定し、この区間の周波数特性とこの区間に対応する既知音響信号の周波数特性との対比から、混合音響信号に含まれる既知音響信号の周波数特性の変化を推定する。

また、補正ステップの実行では、例えば、混合音響信号に含まれる既知音響信号の音量の時間変化を推定し、推定した音量の時間変化に基づいて既知音響信号の振幅スペクトルの音量の時間変化を補正する。音量の時間変化の推定は、周波数特性の補正を行った後に、例えば、混合音響信号に含まれる既知音響信号に相当する振幅を持つ周波数帯域を各時刻において特定し、その周波数帯域における混合音響信号の振幅と既知音響信号の振幅との対比から推定する。

また、補正ステップの実行では、例えば、混合音響信号に含まれる既知音響信号の時間軸方向の伸縮を推定し、推定した時間軸方向の伸縮に基づいて既知音響信号の振幅スペクトルの時間軸方向の伸縮を補正する。時間軸方向の伸縮の推定には、例えば、混合音響信号中の既知の音響信号だけが含まれている区間を特定し、この区間に対応する既知音響信号の区間との時間軸の対比により、時間軸方向の伸縮を推定する。あるいは、時間軸を短い区間に分割した

全区間の対比によって推定する。

また、補正ステップの実行では、例えば、混合音響信号に含まれる既知音響信号の周波数軸方向の伸縮を推定し、推定した周波数軸方向の伸縮に基づいて既知音響信号の振幅スペクトルの周波数軸方向の伸縮を補正する。周波数軸方向の伸縮の推定には、例えば、混合音響信号中の既知の音響信号だけが含まれている区間を特定し、この区間に対応する既知音響信号の区間との周波数軸の対比により、周波数軸方向の伸縮を推定する。

また、本発明の既知音響信号除去方法においては、混合音響信号の振幅スペクトルと既知音響信号の振幅スペクトルを視覚により認識できるように画像表示する画像表示ステップを、更に実行するようにしても良い。この場合には、画像表示に基づいて人間が混合音響信号中における既知の音響信号が含まれている区間を定め、この区間について補正ステップ、除去ステップ、逆変換ステップまたは合成ステップを実行する。

また、本発明の既知音響信号除去方法においては、混合音響信号、既知音響信号及び合成ステップの出力信号を音響として再生する音響再生ステップを、更に実行するようにしても良い。この場合には、音響再生ステップからの再生音に基づいて人間が混合音響信号中における既知の音響信号が含まれている区間を定め、この区間について補正ステップ、除去ステップ、逆変換ステップ及び合成ステップを実行する。

また、本発明の既知音響信号除去方法においては、混合音響信号の振幅スペクトルに基づいて、混合音響信号中における既知の音響信号が含まれている区間を自動推定し、この区間について補正ステップ、除去ステップ、逆変換ステップ及び合成ステップを実行するようにしても良い。混合音響信号中に比較的はっきりと既知の音響信号が含まれている場合（例えば、混合音響信号中で既知の音響信号が単独で鳴っている区間がある場合）には、自動推定により区間を特定することができ、自動推定を利用することにより、既知音響信号の

除去作業を速く実施できる。なお、混合音響信号中に含まれる既知音響信号の存在があまりはっきりとしていない場合においては、人間が区間を指定する。

更に、本発明の既知音響信号除去方法においては、混合音響信号中に含まれている音響信号に相当する既知音響信号が複数種類存在する場合には、それら複数の既知音響信号のすべてに関して、既知音響信号変換ステップ及び補正ステップを実行し、混合音響信号の振幅スペクトルから複数の既知音響信号の補正振幅スペクトルをすべて除去する除去ステップを実行して得た除去後振幅スペクトルを用いて、逆変換ステップ及び合成ステップを実行する。これにより、混合音響信号中から複数種類のすべての既知音響信号を除去することができる。

また、補正ステップを実行する際、時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つの補正を人間が手作業で指定することができるインタフェース処理を行うグラフィックユーザインタフェース（ＧＵＩ）を用いる。

インタフェース処理を行う処理モジュールは、複数の音響信号が混合された混合音響信号の中から、既知音響信号の成分を除去する際に、混合音響信号中での既知音響信号の正確な位置を人間が指定できるように構成される。

また、インタフェース処理を行う処理モジュールは、混合音響信号中で既知音響信号の周波数特性が時間的に変化しているときに、それらの変化を人間が指定できるように構成される。また、インタフェース処理を行う処理モジュールは、混合音響信号中で既知音響信号の音量が時間的に変化しているときに、それらの変化を人間が指定できるように構成される。

更に、インタフェース処理を行う処理モジュールは、混合音響信号中で既知の音響信号が時間軸または周波数軸方向に伸縮しているときに、それらの伸縮を人間が指定できるように構成される。

また、インタフェース処理を行う処理モジュールは、混合音響信号と既知音響信号の対応する区間を人間が指定できるように構成される。

また、本発明による既知音響信号除去装置は、混合音響信号を時間周波数表現に変換して混合音響信号の振幅スペクトルと混合音響信号の位相とを求める混合音響信号変換手段と、混合音響信号中に含まれている音響信号に相当する既知音響信号を時間周波数表現に変換して既知音響信号の振幅スペクトルを求める既知音響信号変換手段と、混合音響信号の振幅スペクトルに基づいて、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した既知音響信号の補正振幅スペクトルを求める補正手段と、混合音響信号の振幅スペクトルから既知音響信号の補正振幅スペクトルを除去する除去手段と、除去手段により得た除去後振幅スペクトルと混合音響信号の位相とに基づいて時間表現に逆変換を行って単位波形を求める逆変換手段と、単位波形を合成して既知音響信号の成分を除去した音響信号を得る合成手段とから構成される。

ここでの補正手段には、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つの補正の指定を人間が手作業で行うことができるインタフェース処理を行う処理モジュールを設ける。

インタフェース処理を行う処理モジュールは、混合音響信号の振幅スペクトルと既知音響信号の振幅スペクトルとを視覚により対比できるように画像表示する画像表示部と、混合音響信号、既知音響信号及び合成手段の出力信号を音響として再生する音響再生部とを備える。

インタフェース処理を行う処理モジュールを用いると、画像表示部に表示された混合音響信号の振幅スペクトル及び既知音響信号の振幅スペクトルの画像表示及び／または音響再生部からの再生音に基づいて、混合音響信号中に含まれている既知音響信号の区間を人間が指定できるだけでなく、この区間について人間が手作業で既知音響信号の振幅スペクトルの時間的な位置のずれ、

周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つの補正を指定できる。その結果、混合音響信号中に含まれている既知音響信号の態様が多少複雑であっても、高い除去精度で既知音響信号を除去することができる。

なお、画像表示部は、既知の音響信号が含まれている混合音響信号中の区間の振幅スペクトルと、既知音響信号の対応区間の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した補正振幅スペクトルとを時間軸上で位置を合わせて表示できるように構成されていることが望ましい。

このような構成にすると、補正振幅スペクトルの状態を視覚で確認できるので、補正スペクトルをどのようにすれば、除去精度を高めることができるのかを、画像を見ながら推測することができるので、除去作業が速くなる。

また、画像表示部は、前記混合音響信号の前記振幅スペクトルから前記補正振幅スペクトルを除去した音響信号の振幅スペクトルを画像表示できるように構成することが望ましい。このような構成にすると、補正の効果を画像で確認できるので、カットアンドトライ方式で補正を行いながら、混合音響信号中から既知音響信号を最大限除去することができる。

また、本発明による既知音響信号除去プログラムは、混合音響信号を時間周波数表現に変換して混合音響信号の振幅スペクトルと混合音響信号の位相とを求める混合音響信号変換ステップと、混合音響信号中に含まれている音響信号に相当する既知音響信号を時間周波数表現に変換して既知音響信号の振幅スペクトルを求める既知音響信号変換ステップと、混合音響信号の振幅スペクトルに基づいて、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した前記既知音響信号の補正振幅スペクトルを求める補正ステップと、混合音響信号の振幅スペクトルから既知音響信号の補正振幅スペクトルを除去する除去ステップ

と、除去ステップにより得た除去後振幅スペクトルと混合音響信号の位相とに基づいて時間表現に逆変換を行って単位波形を求める逆変換ステップと、単位波形を合成して既知音響信号の成分を除去した音響信号を得る合成ステップとの処理をコンピュータにより実行させるように構成されている。

本発明の既知音響信号除去方法によれば、補正ステップにより、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した既知音響信号の補正振幅スペクトルを求め、この補正振幅スペクトルを混合音響信号の振幅スペクトルから除去するため、混合音響信号中に非定常な雑音として含まれている既知音響信号を高い精度で除去することができる利点が得られる。

また、本発明の既知音響信号除去方法によれば、例えば、人間の声や物音の背景にＢＧＭが鳴っているテレビ番組や映画等の音響信号を入力とすると、別途用意したＢＧＭの音楽音響信号を用いて番組中のＢＧＭを除去し、人間の声や物音だけの音響信号を得ることが可能となる。更に、ＢＧＭ除去後の音響信号に、別の音楽をＢＧＭとして付与することにより、テレビ番組や映画等の音楽を差し換えた再利用が可能となる。

ここでの既知音響信号は、任意の音響信号でよいため、音楽のジャンルを問わず、ボーカルの有無を問わず、伴奏の有無を問わずに適用できる。また、音楽に限らず、定常雑音及び非定常雑音を含めた、任意の既知の雑音に適用できる。

また、既知音響信号除去装置におけるユーザインタフェースを使用して、人間が手作業で修正することで、実務の現場でより高品質な除去作業が実現できる。

図面の簡単な説明

第１図は、本発明の既知音響信号除去装置の実施の形態の一例の構成を

示すブロック図である。

第２図は、本発明の既知音響信号除去方法を実施する場合のステップを示すブロック図である。

第３図は、本発明の既知音響信号除去装置の主要部を、コンピュータを用いて実現する場合に用いるプログラムのアルゴリズムの一例を示すフローチャートである。

第４図は、第３図のステップＳＴ１０３内の詳細な処理を示すフローチャートである。

第５図は、人間がかかわる推定と自動推定のいずれでも推定動作をする場合のステップの詳細を示すフローチャートである。

第６図は、エディタのインタフェースの画面構成を示す図である。

第７図は、混合音響信号のパワーの時間変化を示す図である。

第８図は、混合音響信号の振幅スペクトルの時間変化を示す図である。

第９図は、ＢＧＭの元となる音源の既知音響信号のパワーの時間変化を示す図である。

第１０図は、ＢＧＭの元となる音源の既知音響信号の振幅スペクトルの時間変化を示す図である。

第１１図は、既知音響信号除去後の所望の音響信号のパワーの時間変化を示す図である。

第１２図は、既知音響信号除去後の所望の音響信号の振幅スペクトルの時間変化を示す図である。

発明を実施するための最良の形態

以下、図面を参照して本発明の実施の形態の一例を詳細に説明する。第１図は、本発明の既知音響信号除去方法を実施する既知音響信号除去装置の一実施の形態の構成を示すブロックである。

既知音響信号除去装置は、システム構成としては、混合音響信号変換手

段１と、既知音響信号変換手段２と、補正手段３と、インタフェース４と、除去手段５と、逆変換手段６と、合成手段７とから構成される。

混合音響信号変換手段１は、所望の音声や物音等の音響信号ｓ（ｔ）（ｔは時間軸）に、ＢＧＭ等の音響信号ｂ（ｔ）が混合された混合音響信号ｍ（ｔ）を（この時点ではｓ（ｔ）とｂ（ｔ）は未知でありｍ（ｔ）のみが入力される）、時間周波数表現に変換して混合音響信号の振幅スペクトルＭ（ω，ｔ）と混合音響信号の位相θｍ（ω，ｔ）とを求める。

また、既知音響信号変換手段２は、除去すべき音響信号ｂ（ｔ）の元となる音源の既知音響信号ｂ’（ｔ）を時間周波数表現に変換して既知音響信号の振幅スペクトルＢ’（ω，ｔ）を求める。

そして、補正手段３は、混合音響信号の振幅スペクトルＭ（ω，ｔ）に基づいて、混合音響信号の振幅スペクトルＭ（ω，ｔ）に対する既知音響信号の振幅スペクトルＢ’（ω，ｔ）の時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮を補正した既知音響信号の補正振幅スペクトルＢ（ω，ｔ）を求める。自動化のためには、自動で位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮のすべてを自動で推定して補正するように補正手段３を構成することができる。

この実施の形態では、補正手段３は、時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮のすべての補正を、インタフェース４を用いて人間が手作業で指定することができるように構成されている。

このインタフェース４は、後に詳しく説明するように、混合音響信号の振幅スペクトルと既知音響信号の振幅スペクトルとを視覚により対比できるように画像表示する画像表示部を備えており、グラフィックユーザインタフェース（ＧＵＩ）によりインタフェース処理を行う処理モジュールである。

インタフェース４は、画面表示された入力部により混合音響信号の振幅

スペクトルと既知音響信号の振幅スペクトルとに基づいて混合音響信号中に含まれている既知音響信号の区間を人間が指定でき且つ前述の補正を指定できるように構成されている。

除去手段５は、混合音響信号の振幅スペクトルＭ（ω，ｔ）から既知音響信号の補正振幅スペクトルＢ（ω，ｔ）を除去する。そして逆変換手段６は、除去手段５により得た除去後振幅スペクトルＳ（ω，ｔ）と混合音響信号の位相θｍ（ω，ｔ）とに基づいて時間表現に逆変換を行って単位波形ｓ’（ｔ）を求める。

最後に、合成手段７は、逆変換手段６から出力される単位波形ｓ’（ｔ）を合成して既知音響信号の成分を除去した音響信号ｓ（ｔ）を得る。インタフェース４は、除去手段５から出力された除去後振幅スペクトルＳ（ω，ｔ）を画像表示部（第６図参照）に表示する。また、インタフェース４は音響再生部を内蔵しており、混合音響信号、既知音響信号及び合成手段７から出力された合成された音響信号を再生する。

この構成によれば、補正の効果を画像表示部で視覚により確認し、また内蔵された音響再生部で聴覚によっても確認できるので、カットアンドトライ方式で補正を行いながら、インタフェース４の画像表示部の画面表示を見ながら、人間が必要な補正を指定することにより、混合音響信号中から既知音響信号を最大限除去することができる。

次に、第２図及び第３図を用いて、本発明の既知音響信号除去装置の詳細な実施の形態の一例を説明する。第２図は、本発明の既知音響信号除去方法を実施する場合のステップを示すブロック図であり、第３図は本発明の既知音響信号除去装置の主要部を、コンピュータを用いて実現する場合に用いるプログラムのアルゴリズムの一例を示すフローチャートである。

第４図は、第３図のステップＳＴ１０３内の詳細な処理を示すフローチャートである。また、第５図は、人間がかかわる推定と自動推定のいずれでも

推定処理を実行する場合のステップの詳細を示すフローチャートである。以下これらの第１図乃至第５図を参照しながら、本発明の既知音響信号除去方法及び装置における既知音響信号除去の動作を説明する。

まず、以下の説明では、所望の音声や物音等の音響信号ｓ（ｔ）（ｔは時間軸）に、除去する既知音響信号であるＢＧＭ等の音響信号ｂ（ｔ）が混合された混合音響信号ｍ（ｔ）が観測されるものとする。

$$m(t) = s(t) + b(t) \qquad \cdots\cdots\cdots(1)$$

ここでは、ｂ（ｔ）の元となる音源の音響信号ｂ’（ｔ）が既知という条件下で、ｍ（ｔ）が与えられたときに、未知のｓ（ｔ）を求める問題を解く。例えば、人間の声や物音と共にＢＧＭが鳴っているテレビ番組等の音響信号ｍ（ｔ）を入力とし、そのＢＧＭの楽曲が既知であり、その音響信号ｂ’（ｔ）が別途用意できるときに、そのＢＧＭの音楽音響信号を用いて番組中のＢＧＭを除去し、人間の声や物音だけの音響信号ｓ（ｔ）を得る処理を実現する。ここで、ｂ（ｔ）とｂ’（ｔ）は完全には一致しないため、

$$s(t) = m(t) - b(t) \qquad \cdots\cdots\cdots(2)$$

の減算に相当する処理では、ｂ’（ｔ）からｂ（ｔ）に相当する成分を推定して、ｓ（ｔ）を求める必要がある。具体的には、既知音響信号ｂ’（ｔ）は、混合音ｍ（ｔ）中では、以下のような変形を伴うことが多いため、補正することにより、ｂ（ｔ）に相当する成分を推定する。補正の対象は、主として以下に説明するように、時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸あるいは周波数軸方向の伸縮である。

（時間的な位置のずれ）

混合音ｍ（ｔ）中で既知音響信号ｂ’（ｔ）が鳴っている位置は先頭からとは限らない。そこで、既知音響信号ｂ’（ｔ）を時間軸方向にずらし、両

者の相対位置を合わせて、混合音から既知音響信号を減算する必要がある。

（周波数特性の時間変化）

混合音ｍ（ｔ）中で既知音響信号ｂ’（ｔ）が鳴る際には、グラフィックイコライザ等の影響で周波数特性が変化することが多い。例えば、低域や高域が強調・減衰されることがある。そこで、ｂ’（ｔ）の周波数特性を同様に変化させて補正し、混合音から既知音響信号を減算する必要がある。

（音量の時間変化）

混合音ｍ（ｔ）中で既知音響信号ｂ’（ｔ）が鳴る際には、混合音作成時のミキサーのフェーダー等の操作で混合比率が変更され、音量が時間変化することが多い。そこで、ｂ’（ｔ）の音量を同様に時間変化させて補正し、混合音から既知音響信号を減算する必要がある。

（時間軸あるいは周波数軸方向の伸縮）

混合音ｍ（ｔ）中で既知音響信号ｂ’（ｔ）が鳴る際には、レコード等の回転数の違いにより、時間軸あるいは周波数軸方向に伸縮されることがある。そこで、ｂ’（ｔ）を時間軸あるいは周波数軸方向に伸縮して補正し、混合音から既知音響信号を減算する必要がある。

本発明の既知音響信号除去方法においては、基本的な処理として、第２図に示すように、ステップＳＴ１において、まず、混合音響信号をフーリエ変換して、混合音響信号の位相（ステップＳＴ２）と混合音響信号の振幅スペクトル（ステップＳＴ３）を求める（混合音響信号変換ステップ）と共に、ステップＳＴ４で混合音響信号中に含まれている音響信号に相当する既知音響信号をフーリエ変換して、既知音響信号の振幅スペクトル（ステップＳＴ５）を求める（既知音響信号変換ステップ）。

そして、ステップＳＴ６により、混合音響信号の振幅スペクトルに基づいて、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した既知音響信号の補正振幅スペクトル（ステップＳＴ７）を求める（補正ステップ）。次に、ステップＳＴ８で、混合音響信号の振幅スペクトルから既知音響信号の補正振幅スペクトルを除去して除去後振幅スペクトル（ステップＳＴ９）を求めて（除去ステップ）、次のステップＳＴ１０により、除去ステップにより得た除去後振幅スペクトルと混合音響信号の位相とに基づいて逆フーリエ変換を行って単位波形を求める（逆変換ステップ）。最後に、ステップＳＴ１１で、単位波形をオーバーラップ・アド法により合成して既知音響信号の成分を除去した音響信号を得る（合成ステップ）。

これらの処理をコンピュータを用いて実現する場合に用いるプログラムのアルゴリズムでは、第３図フローチャートに示すように、まず、ステップＳＴ１０１で、混合音響信号をフーリエ変換して混合音響信号の振幅スペクトルと混合音響信号の位相とを求める。次に、ステップＳＴ１０２で、混合音響信号中に含まれている音響信号に相当する既知音響信号をフーリエ変換して既知音響信号の振幅スペクトルを求める。

次のステップＳＴ１０３では、混合音響信号の振幅スペクトルに基づいて、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した既知音響信号の補正振幅スペクトルを求める。

その後、ステップＳＴ１０４で、混合音響信号の振幅スペクトルから既知音響信号の補正振幅スペクトルを除去して除去後振幅スペクトルを求める。次にステップＳＴ１０５で、ステップＳＴ１０４で得た除去後振幅スペクトル

と混合音響信号の位相とに基づいて逆フーリエ変換を行って単位波形を求め、ステップＳＴ１０６で単位波形をオーバーラップ・アド法により合成して既知音響信号の成分を除去した音響信号を得る。

その後、ステップＳＴ１０７で、除去後の音響信号をユーザが満足したと評価したか否かの判定が加わり、判定結果が不満足であれば、ステップＳＴ１０３へと戻って補正がやり直される。ユーザが満足するまでは、ステップＳＴ１０３からステップＳＴ１０７が繰り返される。

以下、更に、各ステップで実行される内容を詳細に説明する。本発明の実施の形態の方法では、時間領域で波形を減算処理をせずに、時間周波数領域での振幅スペクトル上で減算処理を行う。

例えば、音響信号ｍ（ｔ），ｂ’（ｔ）に対する窓関数ｈ（ｔ）を用いた時刻ｔにおける短時間フーリエ変換（ＳＴＦＴ）X_m（ω，ｔ），$X_{b'}$（ω，ｔ）が

$$X_m(\omega,t) = \int_{-\infty}^{\infty} m(\tau)h(\tau-t)e^{-j\omega\tau}d\tau \quad \cdots\cdots(3)$$

$$= \alpha_m(\omega,t) + j\beta_m(\omega,t) \quad \cdots\cdots(4)$$

$$X_{b'}(\omega,t) = \int_{-\infty}^{\infty} b'(\tau)h(\tau-t)e^{-j\omega\tau}d\tau \quad \cdots\cdots(5)$$

$$= \alpha_{b'}(\omega,t) + j\beta_{b'}(\omega,t) \quad \cdots\cdots(6)$$

で定義されるとき、それらの振幅スペクトルＭ（ω，ｔ），Ｂ’（ω，ｔ）は、

$$M(\omega,t) = |X_m(\omega,t)| \quad \cdots\cdots(7)$$

$$= \sqrt{\alpha_m^2(\omega,t) + \beta_m^2(\omega,t)} \quad \cdots\cdots(8)$$

$$B'(\omega,t) = |X_{b'}(\omega,t)| \quad \cdots\cdots (9)$$

$$= \sqrt{\alpha_{b'}^2(\omega,t) + \beta_{b'}^2(\omega,t)} \quad \cdots\cdots (10)$$

で求まる。

現在の実装では、音響信号を標本化周波数４４．１ｋＨｚ、量子化ビット数１６ｂｉｔでＡ／Ｄ変換し、窓関数ｈ（ｔ）として窓幅８１９２点のハニング窓を用いた短時間フーリエ変換（ＳＴＦＴ）を、高速フーリエ変換（ＦＦＴ）によって計算する。その際、高速フーリエ変換（ＦＦＴ）のフレームを４４１点ずつシフトするため、フレームシフト時間（１フレームシフト）は１０ｍｓとなる。このフレームシフトを、処理の時間単位とする。

既知音響信号除去後の所望の音響信号ｓ（ｔ）の振幅スペクトルＳ（ω，ｔ）は、振幅スペクトルＭ（ω，ｔ），Ｂ’（ω，ｔ）から以下の式によって求める。ここで、Ｂ（ω，ｔ）はＢ’（ω，ｔ）を補正した後の振幅スペクトルである。

$$S(\omega,t) = \begin{cases} c(\omega,t)\,(M(\omega,t) - B(\omega,t)) & \text{if } M(\omega,t) > B(\omega,t) \\ 0 & \text{otherwise} \end{cases} \quad \cdots\cdots (11)$$

$$B(\omega,t) = a(t)\,g(\omega,t)\,B'(p(\omega), q(t) + r(t)) \quad \cdots\cdots (12)$$

上記の式における各種パラメータ関数ａ（ｔ），ｇ（ω，ｔ），ｐ（ω），ｑ（ｔ），ｒ（ｔ），ｃ（ω，ｔ）を順に説明する。

ここでのａ（ｔ）は、混合音の振幅スペクトルから既知音響信号の振幅スペクトルに相当する成分を減算する分量を最終的に調整するための任意の形状の関数であり、通常、ａ（ｔ）≧１とする。これが大きいほど、減算量が大

きくなる。

g（ω，t）は、周波数特性の時間変化と音量の時間変化を補正するための関数であり、

$$g(\omega,t) = g_\omega(\omega,t)\, g_t(t) + g_r(t) \qquad \cdots\cdots\cdots (13)$$

のように定義する。ここで、g_ω（ω，t）は、周波数特性の時間変化を表し、周波数特性の変化がないときはg_ω（ω，t）＝1となる。一方、g_t（t）は、音量の時間変化を表し、音量の変化がないときは定数となる。M（ω，t）とB'（ω，t）との音量差は、基本的にg_t（t）で補正される。g_r（t）は、主にg（ω，t）の値を全体的に持ち上げるための関数で、補正時の微調整に使用される。使用しない場合には、g_r（t）＝0とする。

p（ω）は、周波数軸方向の伸縮を補正するための関数であり、振幅スペクトルB'（ω，t）の周波数軸ωを変換することで、周波数軸方向の線形・非線型な伸縮を可能にする。なお、B'（ω，t）は本来のωの定義域外では0をとり、離散化して実装する際には適宜補間することとする。

q（t）は、時間軸方向の伸縮を補正するための関数であり、振幅スペクトルB'（ω，t）の時間軸tを変換することで、時間軸方向の線形・非線型な伸縮を可能にする。なお、B'（ω，t）は本来のtの定義域外では0をとり、離散化して実装する際には適宜補間することとする。

r（t）は、時間的な位置のずれを補正するための関数であり、通常は定数を設定することで、一定のずれ幅を補正する。ずれ幅が時間変化するときには、各時刻での幅を補正する関数を設定する。なお、B'（ω，t）は本来のtの定義域外では0をとり、離散化して実装する際には適宜補間することとする。q（t）とr（t）を統合した一つの関数で表現することも可能だが、ここでは、q（t）は連続的な伸縮を表す目的で設定し、r（t）は不連続な位置のずれを表す目的で設定することとする。

c（ω，t）は、振幅スペクトルに対するイコライジング処理及びフェ

ーダー操作処理のための任意の形状の関数である。ω方向の形状により、グラフィックイコライザのように、既知音響信号除去後の周波数特性を調整することができる。また、t方向の形状により、ミキサーのボリュームフェーダー操作のように、既知音響信号除去後の音量変化を調整することができる。使用しない場合には、c（ω，t）＝1とする。

こうして求めた振幅スペクトルS（ω，t）と、混合音m（t）の位相θ_m（ω，t）を用いてX_s（ω，t）を求め、それを逆フーリエ変換（IFFT）することで、単位波形s'（t）を得る。

$$\theta_m(\omega,t) = \arctan\left(\frac{\beta_m(\omega,t)}{\alpha_m(\omega,t)}\right) \quad \cdots\cdots (14)$$

$$X_s(\omega,t) = S(\omega,t)e^{j\theta_m(\omega,t)} \quad \cdots\cdots (15)$$

$$s'(t) = \frac{h(t)}{2\pi}\int_{-\infty}^{\infty} X_s(\omega,t)e^{j\omega t}d\omega \quad \cdots\cdots (16)$$

この単位波形s'（t）を、オーバーラップ・アド（Overlap Add）法によって配置することにより、既知音響信号除去後の所望の音響信号s（t）を合成する。

以上では、混合音響信号m（t）の中に、既知音響信号b'（t）が一種類含まれている場合を説明したが、b'$_1$（t），b'$_2$（t），…，b'$_N$（t）のように複数含まれている場合には、それらの振幅スペクトルB'$_1$（ω，t），B'$_2$（ω，t），…，B'$_N$（ω，t）からそれぞれに応じたパラメータ関数の設定で式（12）によってそれぞれ求めたB_1（ω，t），B_2（ω，t），…，B_N（ω，t）を用いて、

$$S(\omega,t)=\begin{cases} c(\omega,t)\,(M(\omega,t)-\sum_{n=1}^{N}B_n(\omega,t)) & \text{if } M(\omega,t)>\sum_{n=1}^{N}B_n(\omega,t) \\ 0 & \text{otherwise} \end{cases}$$

..... (17)

のようにＳ（ω，ｔ）を求める処理へ拡張できる。その際には、Ｂｎ（ω，ｔ）の各種パラメータ関数を順に設定するか、全体のバランスを取りながら、複数のＢｎ（ω，ｔ）の各種パラメータ関数を平行して設定する。

また、以上では、モノラル信号を対象に説明したが、ステレオ信号は、左右を混合してモノラル信号に変換して適用してもよいし、ステレオ信号の左右の各信号に対して適用してもよい。また、ステレオ信号中の音源方向を利用して、適用してもよい。

上記各種パラメータ関数の設定について説明する。本発明の方法を適用する際に、式（１１）、式（１２）、式（１３）の各種パラメータ関数ａ（ｔ），ｇ（ω，ｔ）（g_ω（ω，ｔ），g_t（ｔ），g_r（ｔ）），ｐ（ω），ｑ（ｔ），ｒ（ｔ），ｃ（ω，ｔ）の形状は、自動推定してもよいし、人間が手作業で設定してもよい。あるいは、自動推定後に人間が修正してもよい。以下では、具体的な自動推定方法と、人間の手作業による修正を可能にする既知音響信号除去装置におけるインタフェース４を用いる場合について説明する。

最初に、式（１１）、式（１２）、式（１３）の各種パラメータ関数ｇ（ω，ｔ）（g_ω（ω，ｔ），g_t（ｔ）），ｐ（ω），ｑ（ｔ），ｒ（ｔ）の形状を推定する方法を、第４図を用いて以下に説明する。まず、ステップＳＴ２０１でＢＧＭ区間ψの集合Ψの指定・自動推定を行い、ステップＳＴ２０２でｐ（ω），ｑ（ｔ）の自動推定を行い、ステップＳＴ２０３でg_ω（ω，ｔ），g_t（ｔ），ｒ（ｔ）の自動推定を行う。そして推定結果のパラメータ関数が収束するまでこれらのステップが継続される（ステップＳＴ２０４）。ステップ

ＳＴ２０５以降では、補正動作がインタフェース４を用いて実行される。

ｇ（ω，ｔ）の推定では、まず、周波数特性の時間変化g_ω（ω，ｔ）を推定し、次に、音量の時間変化g_t（ｔ）を推定する。ただし、ｇ（ω，ｔ）の推定に先立ち、ｐ（ω），ｑ（ｔ），ｒ（ｔ）は決定されている必要がある。ここでは、便宜上、Ｂ’（ｐ（ω），ｑ（ｔ）＋ｒ（ｔ））をＢ’（ω，ｔ）と記述する。

周波数特性の時間変化g_ω（ω，ｔ）の推定では、原則として、人間の声や物音だけの音響信号ｓ（ｔ）がほとんど含まれていない区間（以下、ＢＧＭ区間と呼ぶ）を用いる。ＢＧＭ区間は、複数用いてもよい。ＢＧＭ区間では、混合音ｍ（ｔ）の振幅スペクトルＭ（ω，ｔ）は、既知音響信号ｂ’（ｔ）によるＢＧＭに相当する振幅スペクトルＢ’（ω，ｔ）に由来の成分がほとんどとなる。そこで、周波数特性が時間変化せずに定常、すなわち、g_ω（ω，ｔ）＝g'_ω（ω）と仮定できるときには、g'_ω（ω）を

$$g'_\omega(\omega) = \frac{\int_{\psi\in\Psi} M(\omega,t)dt}{\int_{\psi\in\Psi} B'(\omega,t)dt} \quad \cdots\cdots\cdots (18)$$

により推定する。ただし、ψは一つのＢＧＭ区間（時間軸上の領域）を表し、Ψは、ψの集合とする。一方、周波数特性が時間変化していくときには、g_ω(ω，ｔ）の時刻ｔに近いＢＧＭ区間ψから

$$\frac{\int_{\psi} M(\omega,t)dt}{\int_{\psi} B'(\omega,t)dt} \quad \cdots\cdots\cdots (19)$$

を求め、補間（内挿あるいは外挿）することによりg_ω（ω，ｔ）を推定する（両側にＢＧＭ区間があるときには、両側から内挿する）。最後に、g_ω（ω，ｔ）を周波数軸方向に平滑化する。なお、平滑化幅は任意に設定でき、平滑化をし

なくてもよい。

音量の時間変化g_t（t）の推定では、振幅スペクトルM（ω，t）と、周波数特性補正後のg_ω（ω，t）B'（ω，t）の各時刻における振幅を比較する。しかし、振幅スペクトルM（ω，t）には、B'（ω，t）に由来の成分以外に、音響信号s（t）に由来の成分も含まれる。そこで、周波数軸ωを複数の周波数帯域Φに分割し、それぞれの帯域ϕ（$\phi \in \Phi$）ごとに

$$g'_t(\phi,t) = \frac{\int_\phi M(\omega,t)d\omega}{\int_\phi g_\omega(\omega,t)\,B'(\omega,t)d\omega} \quad (\phi \in \Phi) \qquad \cdots\cdots\cdots (20)$$

を求める（Φはϕの集合を表す）。Φとして任意の分割が適用できるが、例えば、音楽で用いる平均律の1オクターブごとに分割（対数周波数軸上で等間隔に分割）するとよい。そして、g_t（t）は、min（g'_t（ϕ，t））あるいは

$$\frac{\sum_{\phi\in\Phi} g'_t(\phi,t)}{\sum_{\phi\in\Phi} 1} \qquad \cdots\cdots\cdots (21)$$

により推定する。min（g'_t（ϕ，t））の場合には、M（ω，t）とg_ω（ω，t）B'（ω，t）が一番近い周波数帯域において振幅が比較されることになる。最後に、g_t（t）を時間軸方向に平滑化する。なお、平滑化幅は任意に設定でき、平滑化をしなくてもよい。

p（ω），q（t）の推定では、M（ω，t）とB（ω，t）との距離（例えば、対数スペクトル距離等）が最小となるように、p（ω）とq（t）を変更する。その際、B（ω，t）＝a（t）g（ω，t）B'（p（ω），q（t）＋r（t））の右辺のうち、a（t）＝1とし、

1．（推定途中の）ｐ（ω）とｑ（ｔ）を仮に固定した上で、ｇ（ω，ｔ）とｒ（ｔ）を推定する、

2．（推定途中の）ｇ（ω，ｔ）とｒ（ｔ）を仮に固定した上で、ｐ（ω）とｑ（ｔ）を推定する

との二つの推定を反復的に繰り返して、適切なｐ（ω），ｑ（ｔ）を推定する。これは、音響信号の全区間に対して一度に実行せず、時間軸を分割して、区分的に行うとよい。初期値は前後の区間の連続性を考慮して定める。また、ＢＧＭ区間ψの集合Ψを用いて、それらの複数の区間におけるＭ（ω，ｔ）とＢ（ω，ｔ）との対応関係の時間軸を合わせるように、ｐ（ω），ｑ（ｔ）を推定するとよい。

ｒ（ｔ）の推定では、原則として、ＢＧＭ区間ψの集合Ψを用いて、それらの区間におけるＭ（ω，ｔ）とＢ（ω，ｔ）との対応関係の時間軸を合わせるように、ｒ（ｔ）を求める。ｒ（ｔ）は定数であることが多いが、既知音響信号ｂ’（ｔ）の一部区間が使われずに、飛び飛びで使用されながら混合されていたとき等には、その区間を飛ばすようにｒ（ｔ）が不連続関数となる。

上記のｇ（ω，ｔ）やｒ（ｔ）等の推定では、ＢＧＭ区間ψの集合Ψを用いていた。これは、人間が手作業で指定してもよい。あるいは、手作業で指定したＢＧＭ区間の集合に自動推定で追加してもよい。第５図は、人間が手作業で指定する場合と自動推定する場合のいずれでも対応するプログラムのソフトウエアのアルゴリズムを示すフローチャートである。自動推定する場合には、第５図のステップＳＴ３０２～ＳＴ３１３を実行する。Ψの自動推定では、基本的に、どこか一箇所のＢＧＭ区間ψ１を手掛かりとして、残りのＢＧＭ区間の集合を求める。まず、最初のψ１は、人間が手作業で指定するか、音響信号の時間軸を細かく分割して、それらの短い分割区間の対応関係を判定して求める。人間が手作業で指定しない場合、Ｂ（ω，ｔ）を仮に計算し（ステップＳＴ３０２）、Ｍ（ω，ｔ）とＢ（ω，ｔ）を細かく分割した時間窓の振幅スペクトル間の距離（類似度に相当）を計算する（ステップＳＴ３０３）。

そして、その最小距離の時間窓の対応関係を調べ（ステップＳＴ３０４）、その結果を含む区間を$\psi 1$に設定して初期のΨとする（ステップＳＴ３０５）。次に、$\psi 1$を含むΨに基づいて、Ｂ（ω，ｔ）の各種パラメータ関数を推定し（ステップＳＴ３０６乃至ステップＳＴ３０９）、Ｂ（ω，ｔ）を計算する（ステップＳＴ３１０）。各パラメータの推定値が収束しているかを調べ、収束していない場合には、Ψの全区間に対して、Ｍ（ω，ｔ）とＢ（ω，ｔ）との振幅スペクトル間の距離（類似度に相当）を求める。ここでその最大値（もしくは平均値）の定数倍をＢＧＭ区間判定用閾値とする（ステップＳＴ３１２）。そして、ＢＧＭ区間判定用閾値以下の距離を持つ区間を検出し、新たにΨに追加する（ステップＳＴ３１３）。ただし、追加には上限を設けることもできる。この推定と追加を繰り返すことで、Ψが更新され、各種パラメータ関数が適切に求まっていく。ここで、Ｍ（ω，ｔ）とＢ（ω，ｔ）との距離としては、例えば、二乗平均対数スペクトル距離

$$\int (\log M(\omega,t) - \log B(\omega,t))^2 d\omega \qquad \cdots\cdots\cdots\cdots (22)$$

が有効である。

次に、既知音響信号除去エディタ上のインタフェースによる各種パラメータ関数の調整について説明する。

式（１１）～式（１３）のすべてのパラメータ関数ａ（ｔ），ｇ（ω，ｔ）（g_ω（ω，ｔ），g_t（ｔ），g_r（ｔ）），ｐ（ω），ｑ（ｔ），ｒ（ｔ），ｃ（ω，ｔ）の形状を、人間が手作業で設定するための既知音響信号除去装置のユーザインフェースであるエディタを以下に説明する。エディタのユーザは、最初から任意の関数形状を描いて指定してもよいし、最初はまず自動推定をして、その結果を修正してもよい。

エディタの画面構成を第６図に示す。このエディタは、大別して、混合

音響信号ｍ（ｔ）操作用のサブウィンドウＷ１、既知音響信号ｂ’（ｔ）操作用のサブウィンドウＷ２、既知音響信号除去後の所望の音響信号ｓ（ｔ）操作用のサブウィンドウＷ３の三つのサブウィンドウで構成されている。既知音響信号ｂ’（ｔ）が複数種類ある場合には、切り替えスイッチＷ２Ｓにより、サブウィンドウＷ２で操作する既知音響信号ｂ’（ｔ）を切り替えることができる。このインタフェースでは、第４図に示したステップＳＴ２０５からステップＳＴ２１９が実行される。

まず、全サブウィンドウに共通の機能を述べる。操作範囲スライダーＰ１は、音響信号中のどこを現在表示しているかを表す。カーソルＰ２は、現在の操作対象の時間軸上の位置を表す。アイコン化（折り畳み）ボタンＰ３は、これを押すと一時的にそのボタンの属するサブウィンドウが折り畳まれ、小さくなる。現在操作対象以外の未使用のサブウィンドウを隠して、狭い画面を有効活用できる。フロート化（拡大）ボタンＰ４は、これを押すと一時的にそのボタンの属するサブウィンドウが、親ウィンドウから切り離され(フロート化)、さらに拡大されて操作・編集が容易になる。フロート化（拡大）ボタンＰ４しか描かれていない場合には、このボタンを押すと、それに関連づけられたサブウィンドウがフロート化されて新たに出現する。

サブウィンドウＷ１には、混合音響信号ｍ（ｔ）のパワーのグラフＥ１とその振幅スペクトルＭ（ω，ｔ）のグラフＥ２が表示されている。サブウィンドウＷ２には、既知音響信号ｂ’（ｔ）のパワーのグラフＥ３とその振幅スペクトルＢ’（ω，ｔ）のグラフＥ４が表示されている。サブウィンドウＷ３には、既知音響信号除去後の音響信号ｓ（ｔ）のパワーのグラフＥ５とその振幅スペクトルＳ（ω，ｔ）のグラフＥ６が表示されている。各振幅スペクトルのグラフ（Ｅ１，Ｅ２，Ｅ３）では、左側に濃淡で振幅が描かれ（横軸が時間軸、縦軸が周波数軸）、右側にカーソル位置での振幅が描かれている（横軸がパワー、縦軸が周波数軸）。

また、再生制御操作パネルＰ５１には、人間が聞いて確認するために、混合音響信号の再生、停止、早送り、早戻しが可能なボタン群が並んでいる。再生制御操作パネルＰ５１の操作により、インタフェース４は、内蔵する音響再生部によって混合音響信号を再生する。

既知音響信号ｂ’（ｔ）操作用のサブウィンドウＷ２が、操作の中心となるウィンドウであり、式（１２）、式（１３）のすべてのパラメータ関数ａ（ｔ），ｇ（ω，ｔ）（g_ω（ω，ｔ），g_t（ｔ），g_r（ｔ）），ｐ（ω），ｑ（ｔ），ｒ（ｔ）の形状を、自由に設定できる。以下、各操作パネルの説明を述べる。

１．周波数特性の時間変化の補正用操作パネルＣ１（Ｅ７の右側）

g_ω（ω，ｔ）を表示・操作するためのパネルで、カーソル位置の時刻ｔでのg_ω（ω，ｔ）が描かれている（横軸が大きさ、縦軸が周波数軸）。設定操作結果は、ｇ（ω，ｔ）の表示パネルＥ７に即座に反映される（ステップＳＴ２０５，ＳＴ２０６）。Ｅ７には、濃淡でｇ（ω，ｔ）の値の大きさが描かれている（横軸が時間軸、縦軸が周波数軸）。

２．音量の時間変化の補正用操作パネルＣ２（Ｅ７の下側）

g_t（ｔ）を表示・操作するためのパネルで、設定操作結果は、ｇ（ω，ｔ）の表示パネルＥ７に即座に反映される（ステップＳＴ２０７，ＳＴ２０８）。

３．ｇ（ω，ｔ）の値を全体的に持ち上げるための操作パネルＣ３（Ｅ７の下側）

g_r（ｔ）を表示・操作するためのパネルで、設定操作結果は、ｇ（ω，ｔ）の表示パネルＥ７に即座に反映される（ステップＳＴ２０９，ＳＴ２１０）。

４．混合音の振幅スペクトルから既知音響信号の振幅スペクトルに相当

する成分を減算する分量を最終的に調整するための操作パネルＣ４

ａ（ｔ）を表示・操作するためのパネルである。このパネルを操作するとａ（ｔ）の変更が即座に表示に反映する（ステップＳＴ２１１，ＳＴ２１２）。

５．周波数軸方向の伸縮を補正するための操作パネルＣ５

ｐ（ω）を表示・操作するためのパネルである。このパネルを操作するとｐ（ｔ）の変更が即座に表示に反映する（ステップＳＴ２１３，ＳＴ２１４）。

６．時間軸方向の伸縮を補正するための操作パネルＣ６

ｑ（ｔ）を表示・操作するためのパネルである。このパネルを操作するとｑ（ｔ）の変更が即座に表示に反映する（ステップＳＴ２１５，ＳＴ２１６）。

７．時間的な位置のずれを補正するための操作パネルＣ７

ｒ（ｔ）を表示・操作するためのパネルである。このパネルを操作するとｒ（ｔ）の変更が即座に表示に反映する（ステップＳＴ２１７，ＳＴ２１８）。

また、再生制御操作パネルＰ５２には、人間が聞いて確認するために、既知音響信号の再生、停止、早送り、早戻しが可能なボタン群が並んでいる。再生制御操作パネルＰ５２の操作により、インタフェース４は、内蔵する音響再生部によって既知音響信号を再生する。

次に、既知音響信号除去後の音響信号ｓ（ｔ）操作用のサブウィンドウＷ３では、式（１１）のパラメータ関数ｃ（ω，ｔ）の形状を、自由に設定できる。以下、各操作パネルを説明する。

１．グラフィックイコライザ（ＧＥＱ）操作パネルＣ８（Ｅ８の右側）

ｃ（ω，ｔ）のω方向の形状を表示・操作するためのパネルで、カーソル位

置の時刻ｔでのｃ（ω，ｔ）が描かれている（横軸が大きさ、縦軸が周波数軸）。設定操作結果は、ｃ（ω，ｔ）の表示パネルＥ８に即座に反映される。Ｅ８には、濃淡でｃ（ω，ｔ）の値の大きさが描かれている（横軸が時間軸、縦軸が周波数軸）。

２．ボリュームフェーダー操作パネルＣ９（Ｅ８の下側）

ｃ（ω，ｔ）のｔ方向の形状を表示・操作するためのパネルで、設定操作結果は、ｃ（ω，ｔ）の表示パネルＥ８に即座に反映される。

また再生制御操作パネルＰ５３には、人間が聞いて確認するために、合成した音響信号（合成手段７の出力）の再生、停止、早送り、早戻しが可能なボタン群が並んでいる。再生制御操作パネルＰ５３の操作により、インタフェース４は、内蔵する音響再生部によって合成した音響信号を再生する。

次に、本実施の形態の実装について説明する。まず、音声や物音等の音響信号ｓ（ｔ）にＢＧＭ等の音響信号ｂ（ｔ）が加えられている混合音響信号ｍ（ｔ）が観測されたときに、ｂ（ｔ）の元となる音源の音響信号ｂ’（ｔ）が既知という条件下で、未知のｓ（ｔ）を求めることが可能なプログラムを、各種オペレーティングシステム（Ｌｉｎｕｘ２．４，ＳＧＩ　ＩＲＩＸ６．５，Ｍｉｃｒｏｓｏｆｔ　Ｗｉｎｄｏｗｓ　ＸＰ：登録商標）上に実装した。本プログラムに、ｍ（ｔ）とｂ’（ｔ）が収録されたオーディオファイルを与えると、ｓ（ｔ）のオーディオファイルを得ることができる。

人間の音声や物音にバックグラウンドミュージック（ＢＧＭ）が加えられた様々な混合音に対して実験した結果、そのＢＧＭの原曲の音響信号を用いて、混合音中のＢＧＭを除去し、人間の音声や物音が得られることを確認した。ドラムスの鳴っている曲や鳴っていない曲、ポピュラー音楽やクラシック音楽等の様々なジャンルの曲がＢＧＭとして含まれていても、除去が可能であった。

実験結果の例として、二人の男女の対話のＢＧＭにクラシック音楽が鳴っている混合音を実際に処理した結果を、第７図～第１２図に示す。第７図、第８図に示す混合音響信号ｍ（ｔ）を入力として、第９図、第１０図に示す元音源の既知音響信号ｂ’（ｔ）を用いてＢＧＭ成分を除去した結果が、第１１図、第１２図に示す既知音響信号除去後の音響信号ｓ（ｔ）となる。この処理結果の例の混合音は、「ＲＷＣＰ音声対話データベース」から抜粋した二人の男女の対話の音響信号に、「ＲＷＣ研究用音楽データベース」から抜粋したクラシック音楽の音響信号が加えられたものである。

以上に説明したように、本発明によれば、特に、補正ステップにより、混合音響信号の振幅スペクトルに対する既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した既知音響信号の補正振幅スペクトルを求め、この補正振幅スペクトルを混合音響信号の振幅スペクトルから除去するため、混合音響信号中に非定常な雑音として含まれている既知音響信号を高い精度で除去することができる利点が得られる。

また、人間の声や物音の背景にＢＧＭが鳴っているテレビ番組や映画等の音響信号を入力とすると、別途用意したＢＧＭの音楽音響信号を用いて番組中のＢＧＭを除去し、人間の声や物音だけの音響信号を得ることが可能となる。

更に、ＢＧＭ除去後の音響信号に、別の音楽をＢＧＭとして付与することで、テレビ番組や映画等の音楽を差し換えた再利用が可能となる。

既知音響信号は、任意の音響信号でよいため、音楽のジャンルを問わず、ボーカルの有無を問わず、伴奏の有無を問わずに適用できる。また、音楽に限らず、定常雑音及び非定常雑音を含めた、任意の既知の雑音に適用できる。

請求の範囲

1． 複数の音響信号が混合された混合音響信号から、既知音響信号の成分を除去する既知音響信号除去方法であって、

前記混合音響信号を時間周波数表現に変換して前記混合音響信号の振幅スペクトルと前記混合音響信号の位相とを求める混合音響信号変換ステップと、

前記混合音響信号中に含まれている既知の音響信号に相当する既知音響信号を時間周波数表現に変換して前記既知音響信号の振幅スペクトルを求める既知音響信号変換ステップと、

前記混合音響信号の振幅スペクトルに基づいて、前記混合音響信号の振幅スペクトルに対する前記既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した前記既知音響信号の補正振幅スペクトルを求める補正ステップと、

前記混合音響信号の振幅スペクトルから前記既知音響信号の補正振幅スペクトルを除去する除去ステップと、

前記除去ステップにより得た除去後振幅スペクトルと前記混合音響信号の位相とに基づいて時間表現に逆変換を行って単位波形を求める逆変換ステップと、

前記単位波形を合成して前記既知音響信号の成分を除去した音響信号を得る合成ステップと

からなる既知音響信号除去方法。

2． 前記補正ステップでは、前記混合音響信号に含まれる前記既知音響信号の時間的な位置を推定し、

推定した前記時間的な位置に基づいて前記既知音響信号の振幅スペクトルの時間的な位置のずれを補正する

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

3． 前記補正ステップでは、前記混合音響信号に含まれる前記既知音響信号の周波数特性の時間変化を推定し、

推定した前記周波数特性の時間変化に基づいて前記既知音響信号の振幅スペクトルの周波数特性の時間変化を補正する

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

4． 前記補正ステップでは、前記混合音響信号に含まれる前記既知音響信号の音量の時間変化を推定し、

推定した前記音量の時間変化に基づいて前記既知音響信号の振幅スペクトルの音量の時間変化を補正する

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

5． 前記補正ステップでは、前記混合音響信号に含まれる前記既知音響信号の時間軸方向の伸縮を推定し、

推定した前記時間軸方向の伸縮に基づいて前記既知音響信号の振幅スペクトルの時間軸方向の伸縮を補正する

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

6． 前記補正ステップでは、前記混合音響信号に含まれる前記既知音響信号の周波数軸方向の伸縮を推定し、

推定した前記周波数軸方向の伸縮に基づいて前記既知音響信号の振幅スペクトルの周波数軸方向の伸縮を補正する

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

7． 前記混合音響信号の振幅スペクトルと前記既知音響信号の振幅スペクトルとを視覚により対比できるように画像表示する画像表示ステップと、

前記混合音響信号、前記既知音響信号及び前記合成ステップの出力信号を音として音響再生する音響再生ステップとを更に備え、

前記画像表示と前記音響再生とに基づいて人間が前記混合音響信号中における前記既知の音響信号が含まれている区間を定め、

前記区間について前記補正ステップ、前記除去ステップ、前記逆変換ステップ及び前記合成ステップを実行する

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

８． 前記混合音響信号の振幅スペクトルに基づいて前記混合音響信号中における前記既知の音響信号が含まれている区間を自動推定し、

前記区間について前記補正ステップ、前記除去ステップ、前記逆変換ステップ及び前記合成ステップを実行する

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

９． 前記混合音響信号中に含まれている前記既知の音響信号に相当する複数の前記既知音響信号が存在する場合に、

前記複数の既知音響信号のすべてに関して前記既知音響信号変換ステップ及び前記補正ステップを実行し、

前記混合音響信号の振幅スペクトルから前記複数の既知音響信号の補正振幅スペクトルをすべて除去する除去ステップによって得た除去後振幅スペクトルを用いて、前記逆変換ステップ及び前記合成ステップを実行する

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

１０． 前記補正ステップを実行する際に、前記時間的な位置のずれ、前記周波数特性の時間変化、前記音量の時間変化、前記時間軸方向の伸縮及び前記周波数軸方向の伸縮の少なくとも１つの補正を人間が手作業で指定するインタフェースを用いる

ことを特徴とする請求の範囲第１項に記載の既知音響信号除去方法。

１１．　前記インタフェースは、前記混合音響信号の振幅スペクトルと前記既知音響信号の振幅スペクトルを視覚により対比できるように画像表示する画像表示部を備え、

前記画像表示部の表示に基づいて前記補正を前記人間が手作業で指定することを特徴とする請求の範囲第１０項に記載の既知音響信号除去方法。

１２．　前記インタフェースは、前記混合音響信号、前記既知音響信号及び前記合成ステップの出力信号を音響として再生する音響再生部を備え、

前記音響再生部の表示に基づいて前記補正を前記人間が手作業で指定することを特徴とする請求の範囲第１０項に記載の既知音響信号除去方法。

１３．　前記インタフェースは、前記混合音響信号の振幅スペクトルと前記既知音響信号の振幅スペクトルを視覚により対比できるように画像表示する画像表示部と、前記混合音響信号、前記既知音響信号及び前記合成ステップの出力信号を音響として再生する音響再生部とを備え、

前記画像表示部の表示及び前記音響再生部の再生音に基づいて前記補正を前記人間が手作業で指定する

ことを特徴とする請求の範囲第１０項に記載の既知音響信号除去方法。

１４．　複数の音響信号が混合された混合音響信号から、既知音響信号の成分を除去する既知音響信号除去装置であって、

前記混合音響信号を時間周波数表現に変換して前記混合音響信号の振幅スペクトルと前記混合音響信号の位相とを求める混合音響信号変換手段と、

前記混合音響信号中に含まれている既知の音響信号に相当する既知音響信号を時間周波数表現に変換して前記既知音響信号の振幅スペクトルを求める既知

音響信号変換手段と、

前記混合音響信号の振幅スペクトルに基づいて、前記混合音響信号の振幅スペクトルに対する前記既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した前記既知音響信号の補正振幅スペクトルを求める補正手段と、

前記混合音響信号の振幅スペクトルから前記既知音響信号の補正振幅スペクトルを除去する除去手段と、

前記除去手段により得た除去後振幅スペクトルと前記混合音響信号の位相とに基づいて時間表現に逆変換を行って単位波形を求める逆変換手段と、

前記単位波形を合成して前記既知音響信号の成分を除去した音響信号を得る合成手段とからなる既知音響信号除去装置。

１５．　前記補正手段は、前記時間的な位置のずれ、前記周波数特性の時間変化、前記音量の時間変化、前記時間軸方向の伸縮及び前記周波数軸方向の伸縮の少なくとも１つの補正を人間が手作業で指定することを可能にするインタフェースを備えている

ことを特徴とする請求の範囲第１４項に記載の既知音響信号除去装置。

１６．　前記インタフェースは、前記混合音響信号の振幅スペクトルと前記既知音響信号の振幅スペクトルとを視覚により対比できるように画像表示する画像表示部と、前記混合音響信号、前記既知音響信号及び前記合成手段の出力信号を音響として再生する音響再生部とを備え、

前記画像表示部に表示された前記混合音響信号の振幅スペクトルと前記既知音響信号の振幅スペクトルと、前記音響再生部からの再生音とに基づいて、前記混合音響信号中に含まれている前記既知音響信号の区間の指定と、前記既知音響信号の振幅スペクトルの前記時間的な位置のずれ、前記周波数特性の時間

変化、前記音量の時間変化、前記時間軸方向の伸縮及び前記周波数軸方向の伸縮の少なくとも１つの補正の指定を人間が手作業で行えるように構成されている

ことを特徴とする請求の範囲第１５項に記載の既知音響信号除去装置。

１７．　前記画像表示部は、前記既知の音響信号が含まれている前記混合音響信号中の区間の前記振幅スペクトルと、前記混合音響信号中に含まれている前記既知音響信号の対応区間の前記既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した補正振幅スペクトルとを時間軸上で位置を合わせて表示できる構成である

ことを特徴とする請求の範囲第１６項に記載の既知音響信号除去装置。

１８．　前記画像表示部は、前記混合音響信号の前記振幅スペクトルから前記補正振幅スペクトルを除去した音響信号の振幅スペクトルを画像表示できる構成である

ことを特徴とする請求の範囲第１６または第１７項に記載の既知音響信号除去装置。

１９．　複数の音響信号が混合された混合音響信号から既知音響信号の成分を除去する処理をコンピュータにより実行するためのプログラムであって、

前記混合音響信号を時間周波数表現に変換して前記混合音響信号の振幅スペクトルと前記混合音響信号の位相とを求める混合音響信号変換ステップと、

前記混合音響信号中に含まれている既知の音響信号に相当する既知音響信号を時間周波数表現に変換して前記既知音響信号の振幅スペクトルを求める既知音響信号変換ステップと、

前記混合音響信号の振幅スペクトルに基づいて、前記混合音響信号の振幅ス

ペクトルに対する前記既知音響信号の振幅スペクトルの時間的な位置のずれ、周波数特性の時間変化、音量の時間変化、時間軸方向の伸縮及び周波数軸方向の伸縮の少なくとも１つを補正した前記既知音響信号の補正振幅スペクトルを求める補正ステップと、

前記混合音響信号の振幅スペクトルから前記既知音響信号の補正振幅スペクトルを除去する除去ステップと、

前記除去ステップにより得た除去後振幅スペクトルと前記混合音響信号の位相とに基づいて時間表現に逆変換を行って単位波形を求める逆変換ステップと、

前記単位波形を合成して前記既知音響信号の成分を除去した音響信号を得る合成ステップとの処理をコンピュータにより実行させる

ことを特徴とするプログラム。

２０．　前記補正ステップで、前記混合音響信号に含まれる前記既知音響信号の時間的な位置を推定し、

推定した前記時間的な位置に基づいて前記既知音響信号の振幅スペクトルの時間的な位置のずれを補正することをコンピュータに実行させる

ことを特徴とする請求の範囲第１９項に記載のプログラム。

２１．　前記補正ステップで、前記混合音響信号に含まれる前記既知音響信号の周波数特性の時間変化を推定し、

推定した前記周波数特性の時間変化に基づいて前記既知音響信号の振幅スペクトルの周波数特性の時間変化を補正することをコンピュータに実行させる

ことを特徴とする請求の範囲第１９項に記載のプログラム。

２２．　前記補正ステップで、前記混合音響信号に含まれる前記既知音響信号の音量の時間変化を推定し、

推定した前記音量の時間変化に基づいて前記既知音響信号の振幅スペクトル

の音量の時間変化を補正することをコンピュータに実行させる
ことを特徴とする請求の範囲第１９項に記載のプログラム。

２３．　前記補正ステップで、前記混合音響信号に含まれる前記既知音響信号の時間軸方向の伸縮を推定し、

推定した前記時間軸方向の伸縮に基づいて前記既知音響信号の振幅スペクトルの時間軸方向の伸縮を補正することをコンピュータに実行させる
ことを特徴とする請求の範囲第１９項に記載のプログラム。

２４．　前記補正ステップで、前記混合音響信号に含まれる前記既知音響信号の周波数軸方向の伸縮を推定し、

推定した前記周波数軸方向の伸縮に基づいて前記既知音響信号の振幅スペクトルの周波数軸方向の伸縮を補正することをコンピュータに実行させる
ことを特徴とする請求の範囲第１９項に記載のプログラム。

２５．　前記混合音響信号の振幅スペクトルと前記既知音響信号の振幅スペクトルを視覚により対比できるように画像表示する画像表示ステップ
を更にコンピュータに実行させる
ことを特徴とする請求の範囲第１９項に記載のプログラム。

２６．　前記混合音響信号、前記既知音響信号及び前記合成ステップの出力信号を音響として再生する音響再生ステップ
を更にコンピュータに実行させる
ことを特徴とする請求の範囲第１９項に記載のプログラム。

２７．　前記混合音響信号の振幅スペクトルに基づいて前記混合音響信号中における前記既知の音響信号が含まれている区間を自動推定するステップの処理

をコンピュータに実行させ、

前記区間について前記補正ステップ、前記除去ステップ、前記逆変換ステップ及び前記合成ステップと

の処理をコンピュータに実行させる

ことを特徴とする請求の範囲第１９項に記載のプログラム。

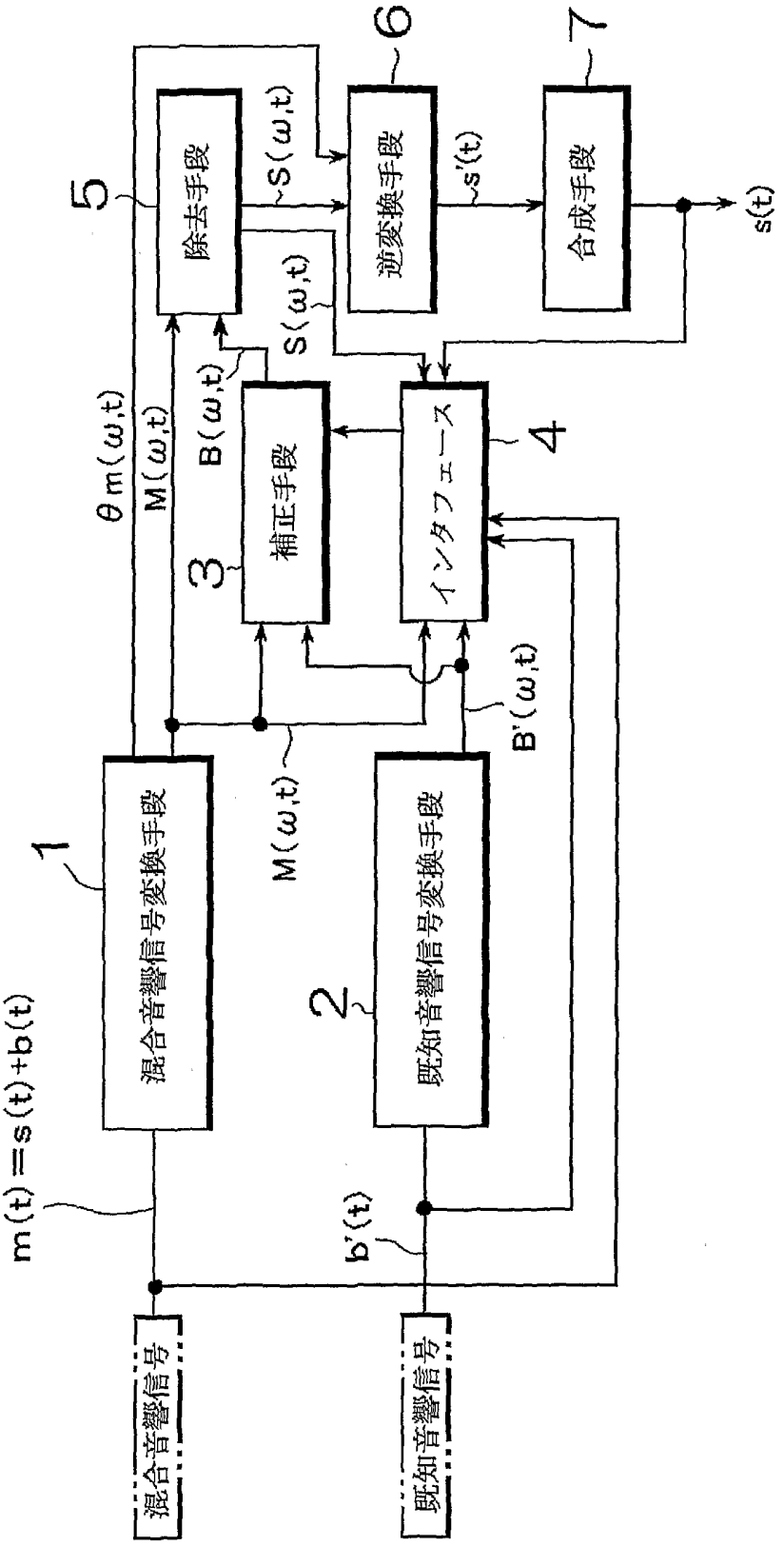
２８．　前記混合音響信号中に含まれている前記既知の音響信号に相当する複数の前記既知音響信号が存在する場合に、

前記複数の既知音響信号のすべてに関して前記既知音響信号変換ステップ及び前記補正ステップを前記コンピュータに実行させ、

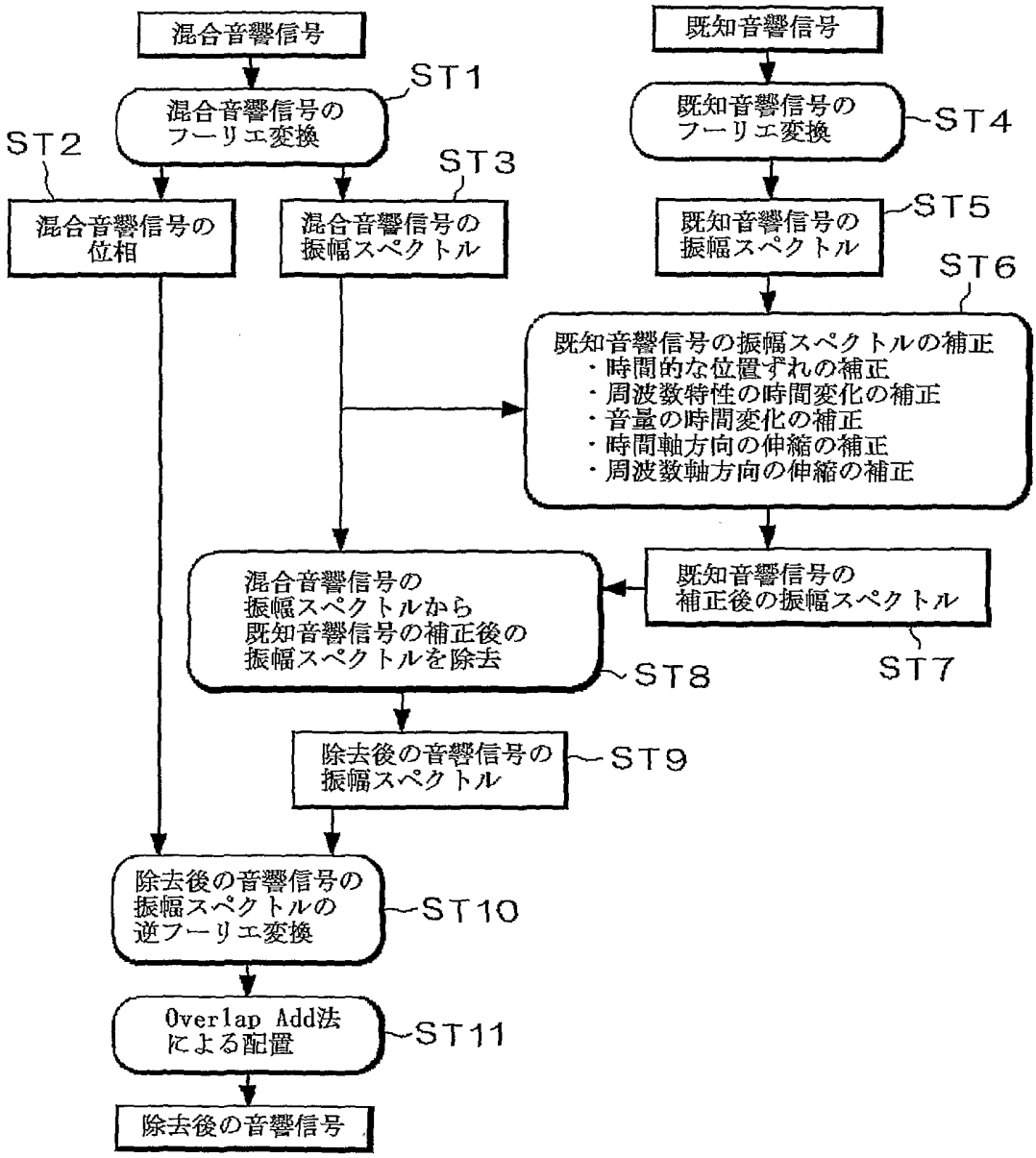
前記混合音響信号の振幅スペクトルから前記複数の既知音響信号の補正振幅スペクトルをすべて除去する除去ステップによって得た除去後振幅スペクトルを用いて、前記逆変換ステップ及び前記合成ステップを前記コンピュータに実行させる

ことを特徴とする請求の範囲第１９項に記載のプログラム。

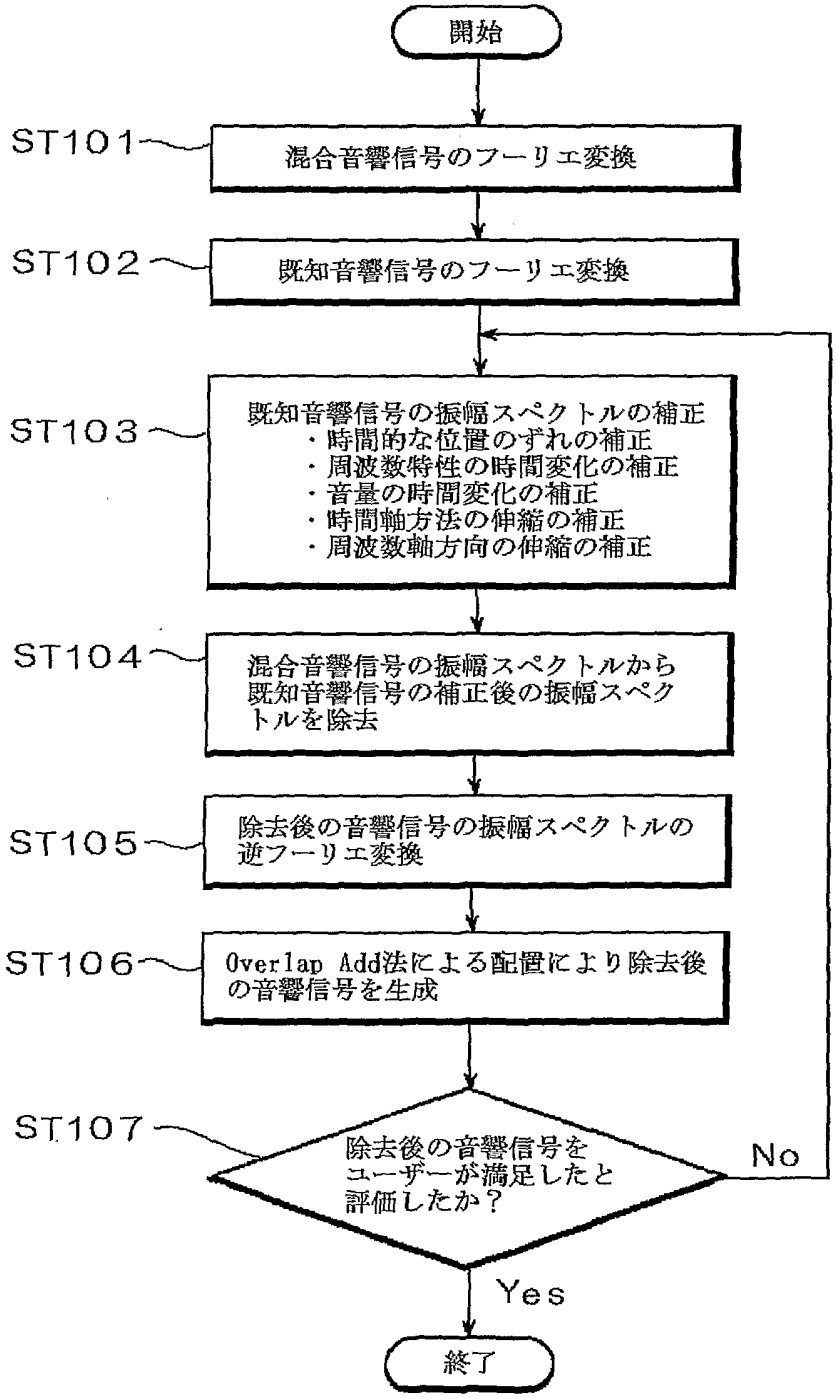
第１図


混合音響信号
既知音響信号
混合音響信号変換手段
既知音響信号変換手段
補正手段
インタフェース
除去手段
逆変換手段
合成手段
1
2
3
4
5
6
7
m(t)=s(t)+b(t)
b'(t)
θm(ω,t)
M(ω,t)
B(ω,t)
S(ω,t)
S(ω,t)
s'(t)
s(t)
B'(ω,t)

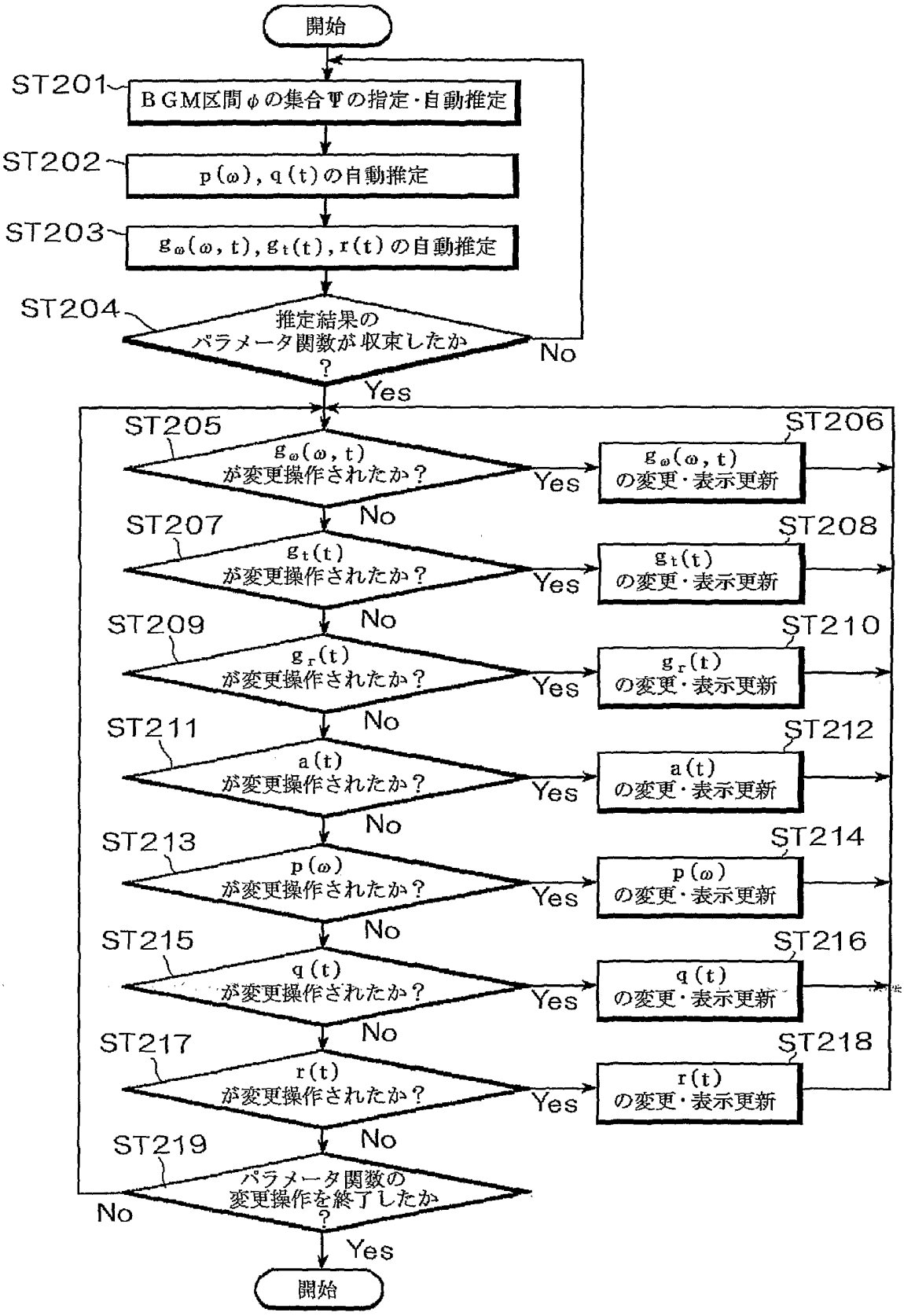
第2図


混合音響信号
既知音響信号
混合音響信号のフーリエ変換
ST1
既知音響信号のフーリエ変換
ST4
ST2
ST3
混合音響信号の位相
混合音響信号の振幅スペクトル
既知音響信号の振幅スペクトル
ST5
ST6
既知音響信号の振幅スペクトルの補正
・時間的な位置ずれの補正
・周波数特性の時間変化の補正
・音量の時間変化の補正
・時間軸方向の伸縮の補正
・周波数軸方向の伸縮の補正
混合音響信号の振幅スペクトルから既知音響信号の補正後の振幅スペクトルを除去
ST8
既知音響信号の補正後の振幅スペクトル
ST7
除去後の音響信号の振幅スペクトル
ST9
除去後の音響信号の振幅スペクトルの逆フーリエ変換
ST10
Overlap Add法による配置
ST11
除去後の音響信号

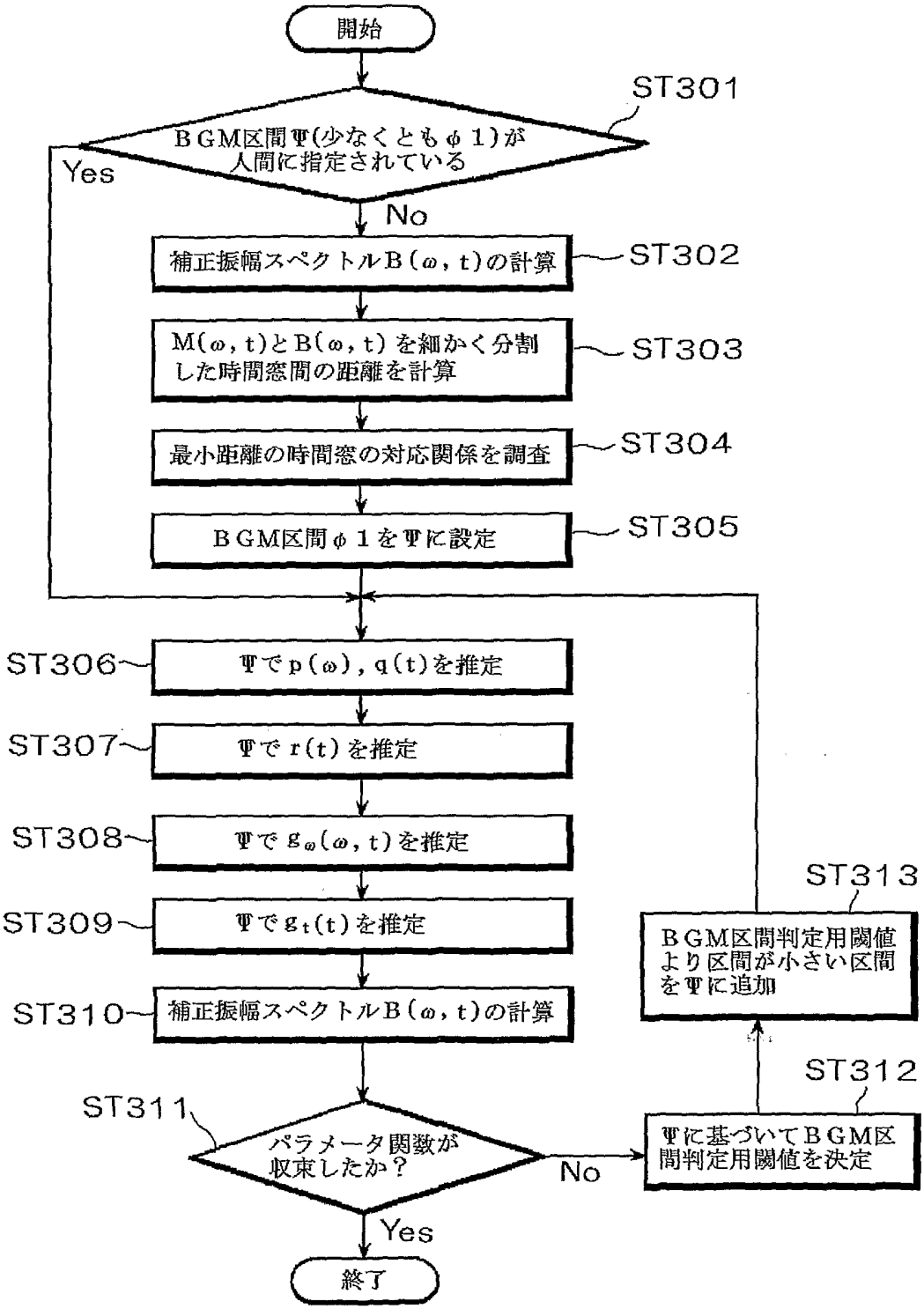
第３図


開始
ST101
混合音響信号のフーリエ変換
ST102
既知音響信号のフーリエ変換
ST103
既知音響信号の振幅スペクトルの補正
・時間的な位置のずれの補正
・周波数特性の時間変化の補正
・音量の時間変化の補正
・時間軸方法の伸縮の補正
・周波数軸方向の伸縮の補正
ST104
混合音響信号の振幅スペクトルから既知音響信号の補正後の振幅スペクトルを除去
ST105
除去後の音響信号の振幅スペクトルの逆フーリエ変換
ST106
Overlap Add法による配置により除去後の音響信号を生成
ST107
除去後の音響信号をユーザーが満足したと評価したか？
No
Yes
終了

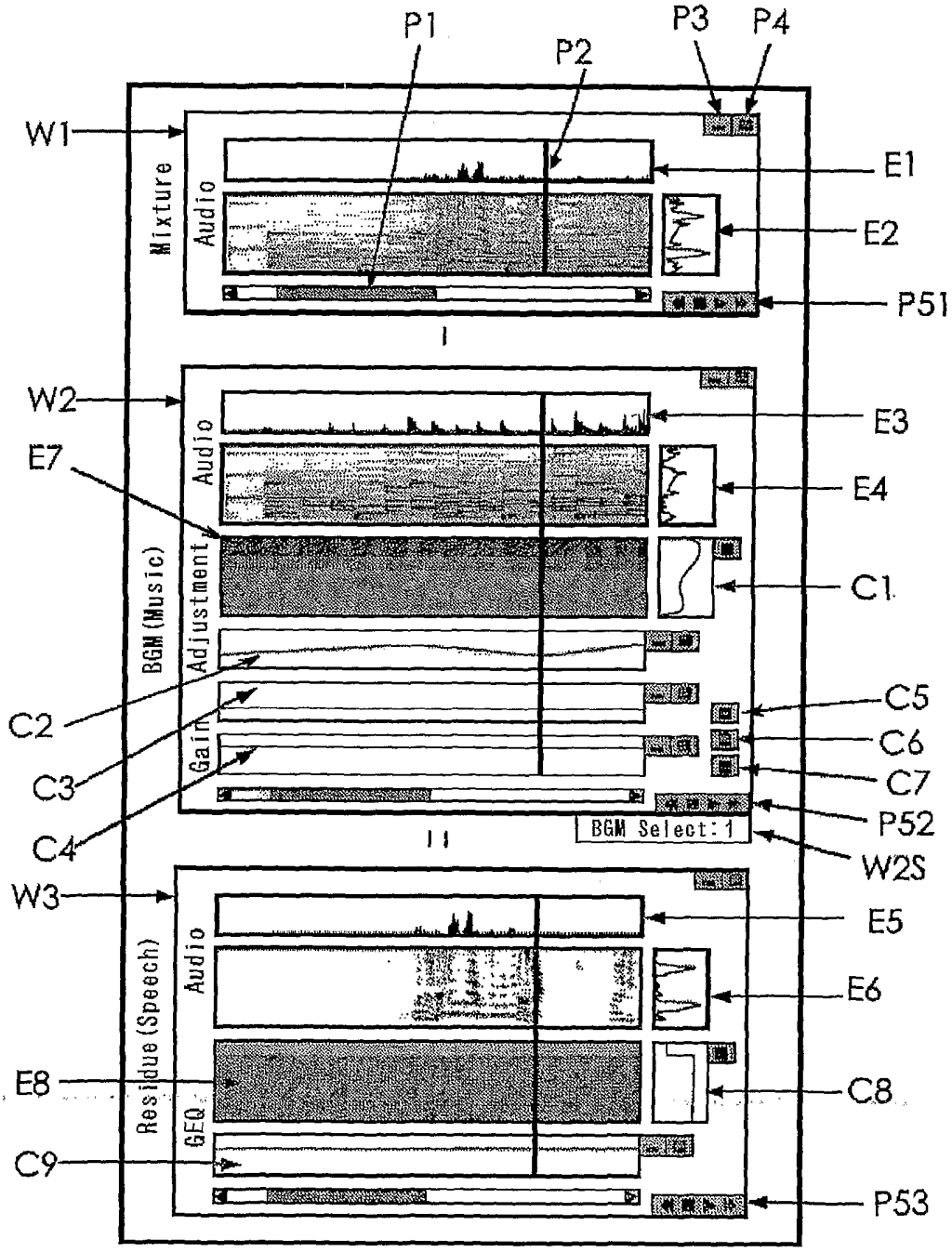
第４図


開始
ST201
BGM区間φの集合Ψの指定・自動推定
ST202
p(ω), q(t)の自動推定
ST203
$g_\omega(\omega, t)$, $g_t(t)$, r(t)の自動推定
ST204
推定結果のパラメータ関数が収束したか？
No
Yes
ST205
$g_\omega(\omega, t)$が変更操作されたか？
Yes
ST206
$g_\omega(\omega, t)$の変更・表示更新
No
ST207
$g_t(t)$が変更操作されたか？
Yes
ST208
$g_t(t)$の変更・表示更新
No
ST209
$g_r(t)$が変更操作されたか？
Yes
ST210
$g_r(t)$の変更・表示更新
No
ST211
a(t)が変更操作されたか？
Yes
ST212
a(t)の変更・表示更新
No
ST213
p(ω)が変更操作されたか？
Yes
ST214
p(ω)の変更・表示更新
No
ST215
q(t)が変更操作されたか？
Yes
ST216
q(t)の変更・表示更新
No
ST217
r(t)が変更操作されたか？
Yes
ST218
r(t)の変更・表示更新
No
ST219
パラメータ関数の変更操作を終了したか？
No
Yes
開始

第5図


開始
ST301
BGM区間Ψ(少なくともφ1)が人間に指定されている
Yes
No
補正振幅スペクトルB(ω, t)の計算
ST302
M(ω, t)とB(ω, t)を細かく分割した時間窓間の距離を計算
ST303
最小距離の時間窓の対応関係を調査
ST304
BGM区間φ1をΨに設定
ST305
ST306
Ψでp(ω), q(t)を推定
ST307
Ψでr(t)を推定
ST308
Ψでgω(ω, t)を推定
ST309
Ψでgt(t)を推定
ST310
補正振幅スペクトルB(ω, t)の計算
ST311
パラメータ関数が収束したか？
No
Yes
終了
ST312
Ψに基づいてBGM区間判定用閾値を決定
ST313
BGM区間判定用閾値より区間が小さい区間をΨに追加

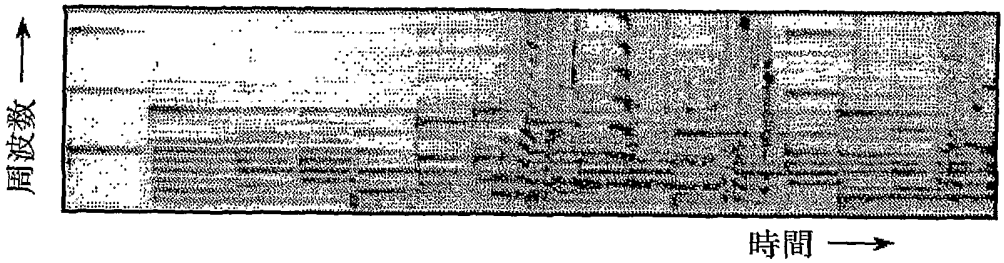
第6図


P1
P2
P3
P4
W1
Mixture
Audio
E1
E2
P51
I
W2
E7
BGM(Music)
Audio
Adjustment
Gain
E3
E4
C1
C2
C3
C4
C5
C6
C7
P52
BGM Select:1
W2S
II
W3
Residue(Speech)
Audio
GEQ
E5
E6
E8
C8
C9
P53

第７図



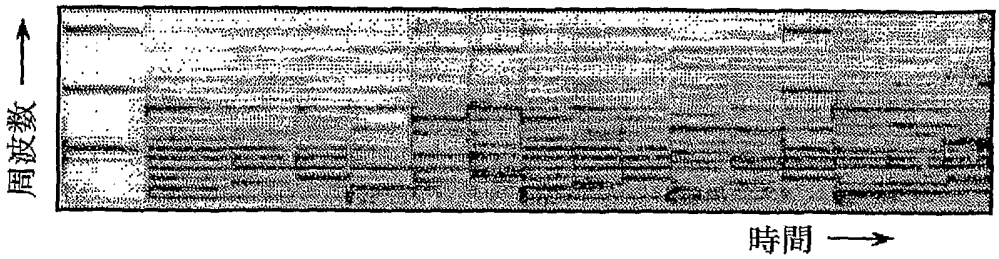
第８図



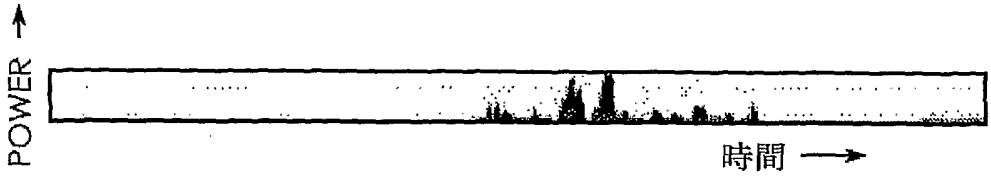
第９図



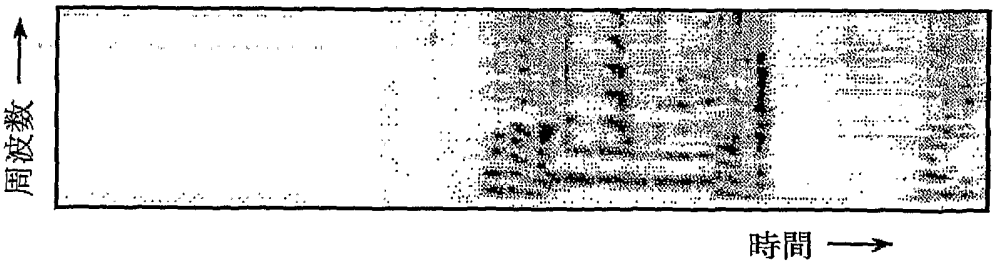
第10図



第11図



第12図



INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/007587

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl[7] G10L21/02

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)
Int.Cl[7] G10L15/20, G10L21/02, H04R3/00

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	1996-2004
Kokai Jitsuyo Shinan Koho	1971-2004	Toroku Jitsuyo Shinan Koho	1994-2004

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Shogo MATSUOKA, Shin'ichi TAKAGI, Akiyuki ODATE, Hideyoshi TOMINAGA, "Contents Haishinji no Tokutei Ongaku Joho Jokyo Shuho no Kento", Information Processing Society of Japan Kenkyu Hokoku, 07 March, 2003 (07.03.03), Vol.2003, No.24, pages 37 to 42	1-28
Y	JP 5-197385 A (Sanyo Electric Co., Ltd.), 06 August, 1993 (06.08.93), Full text; all drawings (Family: none)	1-28
Y	JP 7-199990 A (Ricoh Co., Ltd.), 04 August, 1995 (04.08.95), Full text; all drawings (Family: none)	1-28

☒ 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 17 August, 2004 (17.08.04)	Date of mailing of the international search report 07 September, 2004 (07.09.04)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/007587

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2003-99085 A (National Institute of Advanced Industrial Science and Technology), 04 April, 2003 (04.04.03), Full text; all drawings (Family: none)	1-28
Y	JP 2003-22100 A (Yamaha Corp.), 24 January, 2003 (24.01.03), Full text; all drawings (Family: none)	7,10-13, 15-18,25,26
A	JP 10-307595 A (Seiko Epson Corp.), 17 November, 1998 (17.11.98), Full text; all drawings (Family: none)	1-28
A	JP 8-107375 A (Hitachi, Ltd.), 23 April, 1996 (23.04.96), Full text; all drawings (Family: none)	1-28
A	JP 10-228296 A (Nippon Telegraph And Telephone Corp.), 25 August, 1998 (25.08.98), Full text; all drawings (Family: none)	1-28

国際調査報告	国際出願番号　PCT／JP2004／007587

A.　発明の属する分野の分類（国際特許分類（IPC））

Int. Cl⁷　G10L21／02

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（IPC））

Int. Cl⁷　G10L15／20, G10L21／02,
H04R　3／00

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	1922－1996
日本国公開実用新案公報	1971－2004
日本国実用新案登録公報	1996－2004
日本国登録実用新案公報	1994－2004

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）

JTSPlusファイル（JOIS）

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	関連する請求の範囲の番号
Y	松岡正悟，高木真一，小舘亮之，富永英義， コンテンツ配信時の特定音楽情報除去手法の検討， 情報処理学会研究報告，2003．03．07， Vol．2003，No．24，p．37－42	1－28
Y	JP 5－197385 A（三洋電機株式会社） 1993．08．06，全文，全図（ファミリーなし）	1－28

☒ 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	国際調査報告の発送日
17．08．2004	07. 9. 2004

国際調査機関の名称及びあて先	特許庁審査官（権限のある職員）	5C	3352
日本国特許庁（ISA／JP） 郵便番号100－8915 東京都千代田区霞が関三丁目4番3号	山下　剛史 電話番号　03－3581－1101　内線　3541		

C（続き）．　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	関連する請求の範囲の番号
Y	JP 7－199990 A（株式会社リコー） 1995．08．04，全文，全図（ファミリーなし）	1－28
Y	JP 2003－99085 A（独立行政法人産業技術総合研究所）　2003．04．04，全文，全図（ファミリーなし）	1－28
Y	JP 2003－22100 A（ヤマハ株式会社） 2003．01．24，全文，全図（ファミリーなし）	7， 10－13， 15－18， 25，26
A	JP 10－307595 A（セイコーエプソン株式会社） 1998．11．17，全文，全図（ファミリーなし）	1－28
A	JP 8－107375 A（株式会社日立製作所） 1996．04．23，全文，全図（ファミリーなし）	1－28
A	JP 10－228296 A（日本電信電話株式会社） 1998．08．25，全文，全図（ファミリーなし）	1－28