

(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
2020年10月15日(15.10.2020)



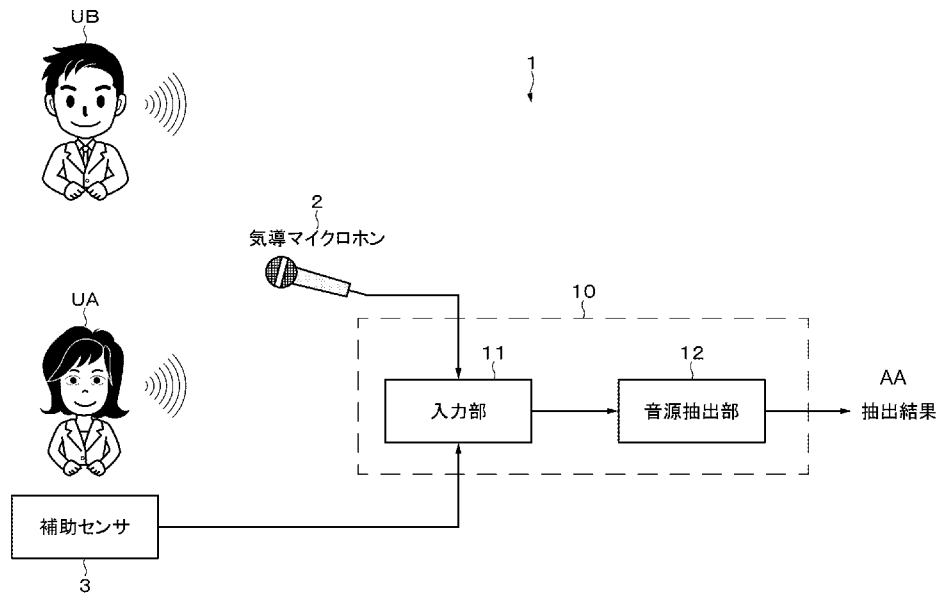
(10) 国際公開番号

WO 2020/208926 A1

- (51) 国際特許分類:
H04R 1/10 (2006.01) G10L 15/20 (2006.01)
H04R 3/00 (2006.01) G10L 25/84 (2013.01)
G10L 21/0272 (2013.01)
- (71) 出願人: ソニー株式会社 (SONY CORPORATION) [JP/JP]; 〒1080075 東京都港区港南1丁目7番1号 Tokyo (JP).
- (21) 国際出願番号: PCT/JP2020/005061
- (72) 発明者: 廣江 厚夫 (HIROE, Atsuo); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP).
- (22) 国際出願日: 2020年2月10日(10.02.2020)
- (25) 国際出願の言語: 日本語
- (74) 代理人: 杉浦 正知, 外 (SUGIURA, Masatomo et al.); 〒1710022 東京都豊島区南池袋1-1-11 カドラービル402 Tokyo (JP).
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2019-073542 2019年4月8日(08.04.2019) 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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,

(54) Title: SIGNAL PROCESSING DEVICE, SIGNAL PROCESSING METHOD, AND PROGRAM

(54) 発明の名称: 信号処理装置、信号処理方法及びプログラム



- 2 Air conduction microphone
- 3 Supplementary sensor
- 11 Input unit
- 12 Sound source extraction unit
- AA Extraction result

(57) Abstract: A signal processing device having: an input unit to which a microphone signal and a one-dimensional time series signal are inputted, said microphone signal including a mixed sound in which a target sound and a sound other than the target sound are mixed, and said one-dimensional time series signal being synchronized with the target sound and having been acquired by a supplementary sensor; and a sound source extraction unit that extracts, from the microphone signal, a target sound signal corresponding to the target sound, on the basis of the one-dimensional time series signal.



WO 2020/208926 A1

HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH,
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, 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, ST, 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, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

- 一 国際調査報告 (条約第21条(3))

(57) 要約：目的音と目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、目的音と同期している1次元の時系列信号とが入力される入力部と、1次元の時系列信号に基づいて、マイクロホン信号から目的音に対応する目的音信号を抽出する音源抽出部とを有する信号処理装置である。

明 細 書

発明の名称： 信号処理装置、信号処理方法及びプログラム

技術分野

[0001] 本開示は、信号処理装置、信号処理方法及びプログラムに関する。

背景技術

[0002] ユーザが発話した音声とそれ以外の音声（例えば、周囲の騒音）とが混じった混合音から、ユーザが発話した音声を抽出する技術が開発されている（例えば、非特許文献1及び非特許文献2の参照のこと）。

先行技術文献

非特許文献

[0003] 非特許文献1：A. Ephrat, I. Mosseri, O. Lang, T. Dekel, K. Wilson, A. Hassidim, W. Freeman, M. Rubinstein, “Looking to Listen at the Cocktail Party: A Speaker-Independent Audio-Visual Model for Speech Separation”, [online], 2018年8月9日, [2019年4月5日検索]、インターネット<URL : <https://arxiv.org/abs/1804.03619>>

[0004] 非特許文献2：M. Delcroix, K. Zmolikova, K. Kinoshita, A. Ogawa, T. Nakatani, “Single Channel Target Speaker Extraction and Recognition with Speaker Beam”, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), p.5554-5558, 2018

発明の概要

発明が解決しようとする課題

[0005] この分野では、抽出対象である音（以下、目的音と適宜、称する）を、目的音及び目的音以外の音が混合された混合音から適切に抽出できることが望まれている。

[0006] 本開示は、上述した点に鑑みてなされたものであり、目的音及び目的音以外の音が混合された混合音から目的音を適切に抽出することができる信号処理装置、信号処理方法及びプログラムに関する。

課題を解決するための手段

[0007] 本開示は、例えば、

目的音と目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、目的音と同期している1次元の時系列信号とが入力される入力部と、

1次元の時系列信号に基づいて、マイクロホン信号から目的音に対応する目的音信号を抽出する音源抽出部と

を有する信号処理装置である。

[0008] また、本開示は、例えば、

目的音と目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、目的音と同期している1次元の時系列信号とが入力部に入力され、

1次元の時系列信号に基づいて、マイクロホン信号から目的音に対応する目的音信号が音源抽出部により抽出される

信号処理方法である。

[0009] また、本開示は、例えば、

目的音と目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、目的音と同期している1次元の時系列信号とが入力部に入力され、

1次元の時系列信号に基づいて、マイクロホン信号から目的音に対応する目的音信号が音源抽出部により抽出される

信号処理方法をコンピュータに実行させるプログラムである。

図面の簡単な説明

[0010] [図1]図1は、実施の形態にかかる信号処理システムの構成例を説明するための図である。

[図2]図2A～図2Dは、実施の形態にかかる信号処理装置で行われる処理の概略を説明する際に参照される図である。

[図3]図3は、実施の形態にかかる信号処理装置の構成例を説明するための図

である。

[図4]図4は、実施の形態にかかる信号処理装置の一態様を説明するための図である。

[図5]図5は、実施の形態にかかる信号処理装置の他の態様を説明するための図である。

[図6]図6は、実施の形態にかかる信号処理装置の他の態様を説明するための図である。

[図7]図7は、実施の形態にかかる音源抽出部の詳細な構成例を説明するための図である。

[図8]図8は、実施の形態にかかる特徴量生成部の詳細な構成例を説明するための図である。

[図9]図9A～図9Cは、実施の形態にかかる短時間フーリエ変換部で行われる処理を説明する際に参照される図である。

[図10]図10は、実施の形態にかかる抽出モデル部の詳細な構成例を説明するための図である。

[図11]図11は、実施の形態にかかる再構成部の詳細な構成例を説明するための図である。

[図12]図12は、実施の形態にかかる学習系を説明する際に参照される図である。

[図13]図13は、実施の形態にかかる学習データを示す図である。

[図14]図14は、実施の形態にかかる気導マイクロホン及び補助センサの具体例を説明する際に参照される図である。

[図15]図15は、実施の形態にかかる気導マイクロホン及び補助センサの他の具体例を説明する際に参照される図である。

[図16]図16は、実施の形態にかかる信号処理装置で行われる全体の処理の流れを示すフローチャートである。

[図17]図17は、実施の形態にかかる音源抽出部で行われる処理の流れを示すフローチャートである。

[図18]図18は、変形例を説明する際に参照される図である。

[図19]図19は、変形例を説明する際に参照される図である。

[図20]図20は、変形例を説明する際に参照される図である。

[図21]図21は、変形例を説明する際に参照される図である。

[図22]図22は、変形例を説明する際に参照される図である。

発明を実施するための形態

[0011] 以下、本開示の実施の形態等について図面を参照しながら説明する。なお、説明は以下の順序で行う。

<1. 実施の形態>

<2. 変形例>

以下に説明する実施の形態等は本開示の好適な具体例であり、本開示の内容がこれらの実施の形態等に限定されるものではない。

[0012] <1. 実施の形態>

[本開示の概要]

始めに、本開示の概要について説明する。本開示は、教示つき音源抽出の一種であり、混合音を取得するためのマイクロホン（気導マイクロホン）に加え、教示情報を取得するためのセンサ（補助センサ）を備える。補助センサの例として、以下のどれかあるいは2つ以上の組み合わせが考えられる。

（1）外耳道など、妨害音よりも目的音の方が優勢な状態で取得できる位置に設置（装着）されたもう一つの気導マイクロホン、（2）骨伝導マイクロホンや咽頭マイクロホンなど、大気中以外を伝播する音波を取得するマイクロホン、（3）音以外のモーダルであって、ユーザの発話と同期している信号を取得するセンサ。補助センサは、例えば、目的音の発生源に装着される。上記（3）の例においてはユーザの発話と同期した信号として、頬や喉付近の皮膚の振動や、顔付近の筋肉の動きなどが考えられる。それらを取得する補助センサの具体例については後述する。

[0013] 図1は、本開示の実施の形態にかかる信号処理システム（信号処理システム1）を示している。信号処理システム1は、信号処理装置10を有してい

る。信号処理装置10は、概略的には、入力部11及び音源抽出部12を有している。また、信号処理システム1は、音を收音する気導マイクロホン2及び補助センサ3を有している。気導マイクロホン2及び補助センサ3は、信号処理装置10の入力部11に対して接続されている。気導マイクロホン2及び補助センサ3は、入力部11に対して有線又は無線により接続されている。補助センサ3は、例えば、目的音の発生源に装着されるセンサである。本例における補助センサ3は、ユーザUAの近傍に配置されており、具体的には、ユーザUAの身体に装着されている。補助センサ3は、後述する目的音と同期した1次元の時系列信号を取得する。かかる時系列信号に基づいて教示情報が得られる。

[0014] 信号処理システム1において音源抽出部12による抽出対象となる目的音とは、ユーザUAが発声した音声である。目的音は、必ず音声であり、且つ、指向性音源である。妨害音音源は、目的音以外の妨害音を発する音源である。これは音声の場合も非音声の場合もあり得、さらに同一の音源から両方の信号が発生する場合もあり得る。妨害音音源は指向性音源または無指向性音源である。妨害音音源の個数は0または1個以上の整数である。図1に示す例では、妨害音の一例として、ユーザUBが発声した音声を示されている。勿論、騒音（例えば、ドアの開閉の際の音や、上空を旋回するヘリコプターの音や、多くの人が存在する場所の雑踏の音など）も妨害音になり得る。気導マイクロホン2は、大気中を伝わる音を収録するマイクロホンであり、目的音と妨害音との混合音を取得する。以下の説明では、取得された混合音のことをマイクロホン観測信号と適宜、称する。

[0015] 次に、図2A～図2Dを参照して、信号処理装置10で行われる処理の概略について説明する。図2A～図2Dでは、横軸が時間、縦軸が音量（またはパワー）をそれぞれ表わしている。

[0016] 図2Aは、マイクロホン観測信号のイメージ図である。マイクロホン観測信号は、目的音に由来する成分4Aと妨害音に由来する成分4Bとが混合した信号である。

[0017] 図2Bは、教示情報のイメージ図である。本例では、補助センサ3が気導マイクロホン2とは異なる位置に設置されたもう一つの気導マイクロホンである場合を想定している。従って、補助センサ3によって取得される1次元の時系列信号は、音信号である。かかる音信号が教示情報として使用される。図2Bは、目的音と妨害音とが混合しているという点は図1と同様だが、補助センサ3の装着位置がユーザの身体に装着されているため、妨害音に由来する成分4Bよりも目的音に由来する成分4Aの方が優勢な状態で観測される。

[0018] 図2Cは、教示情報の他のイメージ図である。本例では、補助センサ3が気導マイクロホン以外のセンサである場合を想定している。気導マイクロホン以外のセンサで取得される信号の例として、骨伝導マイクロホンや咽頭マイクロホンなどによって取得される、ユーザの体内を伝播した音波や、マイクロホン以外のセンサで取得される、ユーザの頬や喉などの皮膚表面の振動や、口付近の筋肉の筋電位および加速度などがある。これらの信号は大気中を伝播しないため、妨害音の影響を受けにくいと考えられる。そのため、教示情報は主に目的音に由来する成分4Aで構成される。すなわち、ユーザの発話開始と共に信号強度が立ち上がり、発話終了と共に立ち下がる。

[0019] 教示情報は目的音の発話と同期して取得されるため、目的音に由来する成分4Aと目的音に由来する成分4Bとの立ち上がり・立ち下がりタイミングは、目的音に由来する成分4Aと同じである。

[0020] 図1に示すように、信号処理装置10の音源抽出部12は、気導マイクロホン2由来のマイクロホン観測信号と補助センサ3由来の教示情報とを入力し、マイクロホン観測信号から妨害音由来の成分を消して目的音由来の成分を残すことで、抽出結果を生成する。

[0021] 図2Dは、抽出結果のイメージ図である。理想的な抽出結果は、目的音に由来する成分4Aのみから構成される。このような抽出結果を生成するため、音源抽出部12はマイクロホン観測信号と教示情報とから抽出結果への対応付けを表わすモデルを有している。かかるモデルは大量のデータによって事

前に学習されている。

[0022] [信号処理装置の構成例]

(全体の構成例)

図3は、実施の形態にかかる信号処理装置10の構成例を説明するための図である。上述したように、気導マイクロホン2は、大気中を伝わってくる目的音と目的音以外の音（妨害音）とが混合された混合音を観測する。補助センサ3はユーザの身体に装着されていて、目的音と同期した1次元の時系列信号を教示情報として取得する。気導マイクロホン2に収録されたマイクロホン観測信号及び補助センサ3により取得された1次元の時系列信号が信号処理装置10の入力部11を介して音源抽出部12に入力される。また、信号処理装置10は、信号処理装置10を統括的に制御する制御部13を有している。音源抽出部12は、気導マイクロホン2により収録される混合音から目的音に対応する目的音信号を抽出し、出力する。具体的には、音源抽出部12は、1次元の時系列信号に基づいて生成される教示情報を使用して目的音信号を抽出する。目的音信号は、後処理部14に出力される。

[0023] 後処理部14の構成は、信号処理装置10が適用される機器に応じて異なる。図4は、後処理部14が、音再生部14Aにより構成される例を示している。音再生部14Aは、音信号を再生するための構成（アンプやスピーカ等）を有している。図示された例の場合には、目的音信号が音再生部14Aにより再生される。

[0024] 図5は、後処理部14が、通信部14Bにより構成される例を示している。通信部14Bは、インターネットや所定の通信網等のネットワークを介して、目的音信号を外部機器に送信するための構成を有している。図示された例の場合には、目的音信号が通信部14Bにより送信される。また、外部機器から送信された音声信号が通信部14Bにより受信される。本例の場合には、信号処理装置10は、例えば、通信機器に適用される。

[0025] 図6は、後処理部14が、発話区間推定部14C、音声認識部14D及びアプリケーション処理部14Eにより構成される例を示している。発話区間

推定部 14C によって、気導マイクロホン 2 から音源抽出部 12 までは途切れないストリームとして扱われていた信号が、発話という単位に分割される。発話区間推定（あるいは音声区間検出）の方法としては、公知の方法を適用することができる。さらに、発話区間推定部 14C の入力として、音源抽出部 12 の出力であるクリーンな目的音に加え、補助センサ 3 で取得された信号を使用しても良い（この場合における補助センサ 3 で取得された信号の流れが、図 6 では点線により示されている。）。即ち、音信号だけでなく、補助センサ 3 で取得された信号も併せて使用することで発話区間推定（検出）が行われても良い。かかる方法としても公知の方法を適用することができる。

[0026] 発話区間推定部 14C は、分割された音そのものを出力するという形態も可能だが、音の代わりに開始時刻や終了時刻といった区間を示す発話区間情報を出し、分割自体は発話区間情報を用いて音声認識部 14D で行うという形態も可能である。図 6 は、後者の形態を想定した例である。音声認識部 14D は、音源抽出部 12 の出力であるクリーンな目的音と発話区間推定部 14C の出力である区間情報とを入力とし、その区間に対応した単語列を音声認識結果として出力する。アプリケーション処理部 14E は、音声認識結果を利用する処理を担当するモジュールである。アプリケーション処理部 14E は、信号処理装置 10 が音声対話システムに適用される例であれば、応答生成や音声合成等を行うモジュールに対応する。また、信号処理装置 10 が音声翻訳システムに適用される例であれば、アプリケーション処理部 14E は、機械翻訳や音声合成等を行うモジュールに対応する。

[0027] （音源抽出部について）

図 7 は、音源抽出部 12 の詳細な構成例を説明するためのブロック図である。音源抽出部 12 は、例えば、A/D (Analog to Digital) 変換部 12A と、特徴量生成部 12B と、抽出モデル部 12C と、再構成部 12D とを有している。

[0028] 音源抽出部 12 の入力は 2 種類ある。一方は気導マイクロホン 2 によって

取得されたマイクロホン観測信号であり、他方は補助センサ3によって取得された教示情報である。マイクロホン観測信号は、AD変換部12Aによってデジタル信号に変換された後、特徴量生成部12Bに送られる。教示情報は特徴量生成部12Bに送られる。図7では省略されているが、補助センサ3で取得された信号がアナログ信号である場合は、当該アナログ信号が、AD変換部12Aとは別のAD変換部によりデジタル信号に変換された後に、特徴量生成部12Bに入力される。このようにデジタル信号に変換されたものも、補助センサ3で取得された1次元の時系列信号に基づいて生成される教示情報の一つである。

[0029] 特徴量生成部12Bは、マイクロホン観測信号及び教示情報の両方を入力とし、抽出モデル部12Cに入力するための特徴量を生成する。また、抽出モデル部12Cの出力を波形に変換するために必要な情報の保持も行う。抽出モデル部12Cのモデルは、目的音と妨害音との混合信号であるマイクロホン観測信号と抽出すべき目的音のヒントとなる教示情報とのセットからクリーンな目的音への対応関係が予め学習されているモデルである。以降では、抽出モデル部12Cへの入力を入力特徴量、抽出モデル部12Cからの出力を出力特徴量と適宜、称する。

[0030] 再構成部12Dは、抽出モデル部12Cからの出力特徴量を音波形またはそれに類する信号に変換する。その際に、特徴量生成部12Bから波形生成のために必要な情報を受け取る。

[0031] (音源抽出部が有する各構成の詳細について)

「特徴量生成部の詳細について」

次に、特徴量生成部12Bの詳細について図8を参照して説明する。図8では、特徴量としてスペクトルまたはそれに類するものを想定しているが、それ以外の特徴量も使用可能である。特徴量生成部12Bは、短時間フーリエ変換部121Bと、教示情報変換部122Bと、特徴量バッファ一部123Bと、特徴量アライメント部124Bとを有している。

[0032] 特徴量生成部12Bの入力としては、2種類の信号が存在する。一方の入

力である、AD変換部12Aによりデジタル信号に変換されたマイクロホン観測信号が短時間フーリエ変換部121Bに入力される。そして、マイクロホン観測信号が、短時間フーリエ変換部121Bによって時間周波数領域の信号、即ち、スペクトルに変換される。

[0033] 他方の入力である補助センサ3からの教示情報に対しては、教示情報変換部122Bによって信号の種類に応じた変換が行われる。教示情報が音信号である場合は、マイクロホン観測信号と同様に短時間フーリエ変換が行われる。教示情報が音以外のモーダルである場合は、短時間フーリエ変換を行うことも無変換のまま使用することも可能である。

[0034] 短時間フーリエ変換部121B及び教示情報変換部122Bによって変換された信号は、所定の時間分だけ特徴量バッファ一部123Bで保存される。ここでは、時間情報と変換結果とが対応付けられた状態で保存されており、後段のモジュールから過去の特徴量の取得というリクエストがあった場合にその特徴量を出力することができる。また、マイクロホン観測信号の変換結果については、その情報が後段での波形生成で使用されるため、複素スペクトルの集合として保存される。

[0035] 特徴量バッファ一部123Bの出力は、2箇所、具体的には再構成部12D及び特徴量アライメント部124Bのそれぞれで使用される。特徴量アライメント部124Bは、マイクロホン観測信号由来の特徴量と教示情報由来の特徴量とで時間の粒度が異なる場合に、両者の粒度を合わせる処理を行う。

[0036] 例えば、マイクロホン観測信号のサンプリング周波数が16kHzであり、短時間フーリエ変換部121Bでのシフト幅が160サンプルとすると、マイクロホン観測信号由来の特徴量は1/100秒に1回の頻度で生成される。一方、教示情報由来の特徴量が1/200秒に1回の頻度で生成される場合は、マイクロホン観測信号由来の特徴量の1セット分と教示情報由来の特徴量の2セット分とを結合したデータを生成し、それを抽出モデル部12Cへの1回分の入力データとする。

[0037] 逆に、教示情報由来の特徴量が1 / 50秒に1回の頻度で生成される場合は、マイクロホン観測信号由来の特徴量の2セット分と教示情報由来の特徴量の1セット分とを結合したデータを生成する。更に、この段階において、複素スペクトルから振幅スペクトルへの変換等も必要に応じて行う。こうして生成された出力が抽出モデル部12Cに送られる。

[0038] ここで、図9を参照して、上述した短時間フーリエ変換部121Bで行われる処理について説明する。AD変換部12Aによって得られたマイクロホン観測信号の波形(図9A参照)から一定長を切り出し、それらにハニング窓やハミング窓等の窓関数を適用する。この切り出した単位をフレームと呼ぶ。1フレーム分のデータに短時間フーリエ変換を適用することにより、時間周波数領域の観測信号として、例えば、 $X(1, t)$ から $X(K, t)$ を得る(図9B参照)。ただし、 t はフレーム番号、 K は周波数ビンの総数を表わす。切り出すフレームの間には重複があってもよく、そうすることで連続するフレーム間で時間周波数領域の信号の変化が滑らかになる。1フレーム分のデータである $X(1, t)$ から $X(K, t)$ までのセットをスペクトルと呼び、複数のスペクトルを時間方向に並べたデータ構造をスペクトログラム(図9C参照)と呼ぶ。図9Cのスペクトログラムでは、横軸がフレーム番号を、縦軸が周波数ビン番号を表わし、図9Aから3本のスペクトル($X(1, t-1)$ から $X(K, t-1)$ まで、 $X(1, t)$ から $X(K, t)$ まで、 $X(1, t+1)$ から $X(K, t+1)$ まで)がそれぞれ生成されている。

[0039] 「抽出モデル部の詳細について」

次に、図10を参照して、抽出モデル部12Cの詳細について説明する。抽出モデル部12Cは特徴量生成部12Bの出力を入力とする。特徴量生成部12Bの出力には2種類のデータが含まれる。一方はマイクロホン観測信号由来の特徴量であり、他方は教示情報由来の特徴量である。以降では、マイクロホン観測信号由来の特徴量を第1特徴量、教示情報由来の特徴量を第2特徴量と適宜、称する。

[0040] 抽出モデル部12Cは、例えば、入力層121Cと、入力層122Cと、

中間層 1～中間層 nを含む中間層 1 2 3 Cと、出力層 1 2 4 Cとを有している。図 1 0 に示される抽出モデル部 1 2 Cは、所謂ニューラルネットワークを表わしている。入力層が入力層 1 2 1 C及び入力層 1 2 2 Cの2つに分かれている理由は、2種類の特徴量をそれぞれに入力するためである。

[0041] 図 1 0 に示した例において、入力層 1 2 1 Cは第 1 特徴量が入力される入力層であり、入力層 1 2 2 Cは第 2 特徴量が入力される入力層である。ニューラルネットワークの種類や構造（層の数）は任意に設定可能であり、後述する学習系により、第 1 特徴量と第 2 特徴量とのセットからクリーンな目的音への対応関係が予め学習されている。

[0042] 抽出モデル部 1 2 Cは、第 1 特徴量を入力層 1 2 1 Cに、第 2 特徴量を入力層 1 2 2 Cにそれぞれ入力し、所定の順方向伝播処理（forward propagation）を行うことで、出力データであるクリーンな目的音の目的音信号に相当する出力特徴量を生成する。出力特徴量の種類として、クリーンな目的音に対応した振幅スペクトルや、マイクロホン観測信号のスペクトルからクリーンな目的音のスペクトルを生成するための時間周波数マスクなどが使用可能である。

[0043] なお、図 1 0 では、2種類の入力データは直後の中間層である（中間層 1）で合流されているが、それよりも出力層 1 2 4 Cに近い側の中間層で合流させても良い。その場合、各入力層から合流地点までの層の個数はそれぞれ異なっても良く、一つの例として、入力データの一方が中間層から入力されるようなネットワーク構造を用いても構わない。2種類のデータを中間層で合流させる方法は、以下のように複数通りが考えられる。一つは、直前の2つの層から出力されるベクトル形式のデータを連結（concatenate）する方法である。もう一つは、2つのベクトルの要素数が同じであれば、要素同士を乗じるという方法である。

[0044] 「再構成部の詳細について」

次に、図 1 1 を参照して、再構成部 1 2 Dの詳細について説明する。再構成部 1 2 Dは、抽出モデル部 1 2 Cの出力を音波形または音に類するデータ

に変換する。かかる処理を行うため、特徴量生成部 1 2 B の中の特徴量バッファ部 1 2 3 B から必要なデータを受け取る。

[0045] 再構成部 1 2 D は、複素スペクトログラム生成部 1 2 1 D と、逆短時間フーリエ変換部 1 2 2 D とを有している。複素スペクトログラム生成部 1 2 1 D は、抽出モデル部 1 2 C の出力及び特徴量生成部 1 2 B からのデータを統合して目的音の複素スペクトログラムを生成する。生成の仕方は、抽出モデル部の出力が振幅スペクトルか時間周波数マスクかによって変わる。振幅スペクトルの場合は、位相情報が欠落しているため、波形に変換するためには位相情報を追加する（復元する）必要がある。位相の復元のためには公知の技術が適用可能であり、例えば、特徴量バッファ部 1 2 3 B から同じタイミングのマイクロホン観測信号の複素スペクトルを取得し、そこから位相情報を取り出して振幅スペクトルと合成することで目的音の複素スペクトルを生成する。

[0046] 一方、時間周波数マスクの場合は、同じくマイクロホン観測信号の複素スペクトルを取得した後、複素スペクトルに時間周波数マスクを適用する（時間周波数ごとに乗算する）ことで目的音の複素スペクトルを生成する。時間周波数マスクの適用については、公知の方法（例えば、特開 2 0 1 5 - 5 5 8 4 3 号公報に記載の方法）を使用することができる。

[0047] 逆短時間フーリエ変換部 1 2 2 D は、複素スペクトルを波形に変換する。逆短時間フーリエ変換は、逆フーリエ変換およびオーバーラップ加算等からなる。これらの方法については公知の方法（例えば、特開 2 0 1 8 - 6 4 2 1 5 号公報に記載の方法）を適用することができる。

[0048] なお、後段のモジュールによっては、再構成部 1 2 D において波形以外のデータに変換したり、或いは、再構成部 1 2 D 自体を省略したりすることも可能である。例えば、後段のモジュールが発話区間検出および音声認識であり、そこで使用される特徴量が振幅スペクトルあるいはそこから生成可能なデータである場合、再構成部 1 2 D は、抽出モデル部 1 2 C の出力を振幅スペクトルに変換するだけで良い。さらに、抽出モデル部 1 2 C が振幅スペク

トルそのものを出力する場合は、再構成部12D自体を省略しても良い。

[0049] (抽出モデル部の学習系について)

次に、図12及び図13を参照して、抽出モデル部12Cの学習系について説明する。かかる学習系は抽出モデル部12Cに対して所定の学習を事前に行うために使用される。以下に説明する学習系は、抽出モデル部12Cを除き、信号処理装置10とは別システムであることを想定しているが、学習系にかかる構成が信号処理装置10に組み込まれていても良い。

[0050] 学習系の基本的な動作は、例えば、以下の(1)から(3)の通りであり、(1)から(3)までのプロセスを繰り返し行うことを学習と呼ぶ。(1) 目的音データセット21及び妨害音データセット22とから、入力特徴量及び教師データ(その入力特徴量に対する理想的な出力特徴量)が生成される。(2) 入力特徴量が抽出モデル部12Cに入力され、順方向伝播によって出力特徴量が生成される。(3) 出力特徴量と教師データとが比較され、誤差が小さくなるよう、換言すれば、損失関数における損失値を最小とするように抽出モデル内のパラメータが更新される。

[0051] 入力特徴量と教師データとのペアを、以下では学習データと適宜、称する。学習データは図13に示すような4通りを生成する。この図において、(a)は目的音と妨害音とが混合している場合に目的音を抽出することを学習するためのデータであり、(b)は静かな環境での発話を劣化なく出力させるためのデータであり、(c)はユーザが発話をしていない場合に無音を出力させるためのデータであり、(d)は静かな環境においてユーザが何も発話していない場合に無音を出力させるためのデータである。なお、図13の教示情報において「無し」とは、信号自体は存在するものの、その中には目的音に由来する成分が含まれていないことを意味する。

[0052] これら4通りの学習データは、場合分けによってそれぞれを所定の割合で生成する。あるいは後述のように、静かな環境で収録された無音に近い音を目的音および妨害音のデータセットに含めておくことで、場合分けなしに全組み合わせが生成されるようにしても良い。

[0053] 以下、学習系を構成するモジュールとその動作について説明する。目的音データセット21は、目的音波形とそれと同期した教示情報とのペアで構成される集合である。ただし、図13における(c)に対応する学習データや図13における(d)に対応する学習データを生成する目的で、静かな場所において人が発話していないときのマイクロホン観測信号およびそれに対応した補助センサの入力信号のペアも、このデータセットには含まれる。

[0054] 妨害音データセット22は、妨害音になり得る音で構成される集合である。音声も妨害音になり得るため、妨害音データセット22には音声も非音声も含まれる。さらに、図13における(b)に対応する学習データや図13における(d)に対応する学習データを生成するため、静かな場所で観測されたマイクロホン観測信号もこのデータセットには含まれる。学習時は、目的音波形と教示情報とからなるペアの一つが目的音データセット21からランダムに取り出される。その内の教示情報については、それが気導マイクロホンによって取得された場合は混合部24へ入力されるが、気導マイクロホン以外のセンサで取得された場合は特徴量生成部25へ直接入力される。目的音波形については、混合部23及び教師データ生成部26にそれぞれに入力される。他方、妨害音データセット22からはランダムに1個以上の音波形が取り出され、当該音波形が混合部23に入力される。補助センサが気導マイクロホン以外である場合は、妨害音データセット22から取り出された波形は混合部24にも入力される。

[0055] 混合部23は、目的音波形と1個以上の妨害音波形とを、所定の混合比(SN比(Signal Noise Ratio))で混合する。混合結果はマイクロホン観測信号に相当し、特徴量生成部25に送られる。混合部24は、補助センサ3が気導マイクロホンである場合に適用されるモジュールであり、音信号である教示情報に対し、妨害音を所定の混合比で混合する。混合部24において妨害音を混合する理由は、教示情報に妨害音がある程度まで混入しても良好な音源抽出が行えるようにするためである。

[0056] 特徴量生成部25に対する入力は2種類あり、一方はマイクロホン観測信

号であり、他方は教示情報あるいは混合部24の出力である。これら2種類のデータから入力特徴量を生成する。抽出モデル部12Cは、学習前および学習途中のニューラルネットワークであり、構成は図10と同一である。教師データ生成部26は、理想的な出力特徴量である教師データを生成する。教師データの形状は基本的に出力特徴量と同一であり、振幅スペクトルや時間周波数マスク等である。ただし後述の通り、抽出モデル部12Cの出力特徴量が時間周波数マスクである一方で、教師データが振幅スペクトルという組み合わせも可能である。

[0057] 図13に示したように、教師データは目的音及び妨害音の有無によって異なる。目的音が存在する場合は目的音に対応した出力特徴量、目的音が存在しない場合は無音に対応した出力特徴量である。比較部27は、抽出モデル部12Cの出力と教師データとを比較し、損失関数(loss function)における損失値が減少するように、抽出モデル部12Cに含まれるパラメータについての更新値を計算する。比較で使用する損失関数としては、平均二乗誤差等が使用可能である。比較の方法及びパラメータの更新方法については、ニューラルネットワークの学習アルゴリズムとして公知の方法を適用することができる。

[0058] [気導マイクロホン及び補助センサの具体例]

(具体例1)

次に、気導マイクロホン2及び補助センサ3の具体例について説明する。図14は、オーバーイヤ型ヘッドホン30における、気導マイクロホン2及び補助センサ3の具体例を示す図である。耳に被せる部品であるイヤークップ31の外側と内側とに、それぞれ、外側(耳介側とは反対側)マイクロホン32と内側(耳介側)マイクロホン33が設けられている。外側マイクロホン32及び内側マイクロホン33は、例えば、ノイズキャンセル用として設けられているマイクロホンを適用することができる。マイクロホンの種類としては外側も内側も気導マイクロホンであるが、使用目的が異なる。外側マイクロホン32が上述した気導マイクロホン2に相当し、目的音と妨

害音とが混合した音を取得するために用いられる。内側マイクロホン33が補助センサ3に相当する。

[0059] 人の発声器官は耳と繋がっているため、ヘッドホン装着者すなわちユーザの発話（目的音）は、大気を通じて外側マイクロホン32で観測される他に、内耳および外耳道を経由し、内側マイクロホン33でも観測される。妨害音は、外側マイクロホン32で観測される他に内側マイクロホン33でも観測されるが、イヤークップ31によってある程度減衰するため、内側マイクロホン33では妨害音よりも目的音が優勢な状態で音が観測される。しかしながら、内側マイクロホン33で観測された目的音は、内耳を経由しているために周波数分布が外側マイクロホン32由来のものとは異なり、また、体内で発生する発話以外の音（嚙下音など）が収録される場合もあるため、内側マイクロホン33で観測された音を他の人間が聴取したり、そのまま音声認識に入力したりすることは必ずしも適切ではない。

[0060] そこで本開示では、内側マイクロホン33で観測された音信号を音源抽出の教示情報として使用することで問題を解決する。具体的には、以下の（1）から（3）の理由によって問題が解決される。（1）抽出結果は、気導マイクロホン2である外側マイクロホン32の観測信号から生成され、更に、学習時に気導マイクロホン由来の教師データを用いるため、抽出結果における目的音の周波数分布は静かな環境で収録されたものに近い。（2）内側マイクロホン33で観測される音すなわち教示情報には、目的音だけでなく妨害音も混入する場合はあるが、学習時にはそのような教示情報および外側マイクロホン観測信号とから目的音を出力するようなデータを用いて対応付けを学習させるため、抽出結果は比較的クリーンな音声である。（3）嚙下音等が内側マイクロホン33で観測されても、その音は外側マイクロホン32では観測されないため、抽出結果には出現しない。

[0061] （具体例2）

図15は、片耳挿入型のイヤホン40における、気導マイクロホン2及び補助センサ3の具体例を示す図である。ハウジング41の外側に外側マイク

ロホン42が設けられている。外側マイクロホン42が、気導マイクロホン2に相当する。外側マイクロホン42により、空気中を伝わる目的音と妨害音とが混合した混合音が観測される。

[0062] イヤーピース43は、ユーザの外耳道に挿入される部分である。イヤーピース43の一部に内側マイクロホン44が設けられている。内側マイクロホン44が、補助センサ3に相当する。内側マイクロホン44では、内耳経由で伝わってきた目的音と、ハウジング部を通過して減衰した妨害音とが混合した音が観測される。音源抽出の方法については図14に示したヘッドホンと同様であるため、重複した説明を省略する。

[0063] (他の具体例)

なお、補助センサ3は、気導マイクロホンに限定されるものではなく、それ以外の種類のマイクロホンや、さらにはマイクロホン以外のセンサも使用可能である。

例えば、補助センサ3として、骨伝導マイクロホンや咽頭マイクロホンといった、体内を直接伝播する音波を取得可能なマイクロホンを使用しても良い。体内を伝播する音波は大気中を伝わる妨害音の影響を受けにくいため、これらのマイクロホンで取得された音信号はユーザのクリーンな発話音声に近いと考えられる。しかし実際には、図14のオーバーイヤー型ヘッドホン30における内側マイクロホン33を使用した場合と同様に、周波数分布の違いや嚙下音などの問題が発生する可能性がある。そこで、骨伝導マイクロホンや咽頭マイクロホンなどを補助センサ3として使用し、教示つき音源抽出を行うことで問題を解決する。

補助センサ3としては他に、音波以外の信号を検出するセンサ、例えば、光センサを適用することも可能である。音を発する物体の表面（例えば、筋肉）は振動しており、人体であれば発声器官の近くにある喉や頬の皮膚は自身が発する音声に応じて振動している。そのため、その振動を光センサによって非接触な方法で検出することで、発話自体の有無を検出したり、音声そのものを推定したりすることができる。

[0064] 例えば、振動を検出する光センサを用いて発話区間の検出を行う技術が提案されている。また、レーザを皮膚に当てることで生じる斑点の明るさを高フレームレートのカメラで観測し、その明るさの変化から音の推定を行う技術も提案されている。本例でも光センサを用いるが、光センサによる検出結果は、発話区間検出や音の推定のためではなく、教示つき音源抽出のために使用される。

[0065] 光センサを使用した具体例について説明する。レーザーポインターやLED等の光源から発する光を頬・喉・後頭部といった発声器官付近の皮膚に当てる。光を当てることで皮膚の上に光の斑点が生じる。斑点の明るさは光センサで観測される。この光センサが補助センサ3に相当し、ユーザの身体に装着されている。集光を容易にするため、光センサと光源とは一体化しても良い。

[0066] 携帯を容易にするため、気導マイクロホン2を光センサおよび光源と一体化しても良い。このモジュールに対して、気導マイクロホン2で取得された信号をマイクロホン観測信号として入力し、光センサで取得された信号を教示情報として入力する。

[0067] 上記の例は振動を検出する光センサを補助センサ3として使用するものであったが、ユーザの発話と同期した信号を取得するセンサであれば、他の種類のセンサも使用可能である。そのような例として、下顎や口唇付近の筋肉の筋電位を取得するための筋電センサや、下顎付近の動きを取得するための加速度センサ等が挙げられる。

[0068] [処理の流れ]

(全体の処理の流れ)

次に、実施の形態にかかる信号処理装置10で行われる処理の流れについて説明する。図16は、実施の形態にかかる信号処理装置10で行われる全体の処理の流れを示すフローチャートである。処理が開始されると、ステップST1では、気導マイクロホン2によりマイクロホン観測信号が取得される。そして、処理がステップST2に進む。

- [0069] ステップS T 2では、補助センサ3により1次元の時系列信号である教示情報が取得される。そして、処理がステップS T 3に進む。
- [0070] ステップS T 3では、音源抽出部12により、マイクロホン観測信号と教示情報とを用いて抽出結果、即ち、目的音信号が生成される。そして、処理がステップS T 4に進む。
- [0071] ステップS T 4では、一連の処理が終了したか否かが判断される。かかる判断処理は、例えば、信号処理装置10の制御部13により行われる。一連の処理が終了していない場合は、処理がステップS T 1に戻り、上述した処理が繰り返される。
- [0072] なお、図16では図示を省略しているが、ステップS T 3にかかる処理で目的音信号が生成された後、後処理部14による処理が行われる。上述したように、後処理部14による処理は、信号処理装置10が適用される機器に応じた処理（通話、録音、音声認識等）である。
- [0073] （音源抽出部による処理の流れ）
- 次に、図16のステップS T 3で行われる音源抽出部12による処理の流れを、図17のフローチャートを参照して説明する。
- [0074] 処理が開始されると、ステップS T 11では、AD変換部12AによるAD変換処理が行われる。具体的には、気導マイクロホン2により取得されたアナログ信号がデジタル信号であるマイクロホン観測信号に変換される。また、補助センサ3としてマイクロホンが適用されている場合には、補助センサ3で取得されたアナログ信号がデジタル信号である教示情報に変換される。そして、処理がステップS T 12に進む。
- [0075] ステップS T 12では、特徴量生成部12Bによる特徴量生成処理が行われる。具体的には、マイクロホン観測信号と教示情報とが、特徴量生成部12Bにより、それぞれ入力特徴量に変換される。そして処理がステップS T 13に進む。
- [0076] ステップS T 13では、抽出モデル部12Cによる出力特徴量生成処理が行われる。具体的には、ステップS T 12で生成された入力特徴量を抽出モ

デルであるニューラルネットワークに入力し、所定の順伝播処理（forward propagation）を行うことで出力特徴量を生成する。そして、処理がステップ S T 1 4 に進む。

[0077] ステップ S T 1 4 では、再構成部 1 2 D による再構成処理が行われる。具体的には、ステップ S T 1 3 で生成された出力特徴量に対して複素スペクトルの生成や逆短時間フーリエ変換等を適用することで、音波形またはそれに類するデータである目的音信号が生成される。そして、処理が終了する。

[0078] なお、音源抽出処理より後段の処理によっては、音波形以外のデータを生成したり再構成処理自体を省略したりしても良い。例えば、後段で音声認識を行う場合は、再構成処理において音声認識用の特徴量を生成しても良く、あるいは、再構成処理において振幅スペクトルを生成し、音声認識において振幅スペクトルから音声認識用の特徴量を生成しても良い。さらに、抽出モデルが振幅スペクトルを出力するように学習されている場合には、再構成処理自体をスキップしても良い。

[0079] なお、上述したフローチャートで示される処理の一部の処理順序が入れ替わっても良いし、複数の処理が並行して行われても良い。

[0080] [実施の形態により得られる効果]

本実施の形態によれば、例えば、以下の効果が得られる。

実施の形態の信号処理装置 1 0 は、目的音と妨害音とが混合した混合音（マイクロホン観測信号）を取得する気導マイクロホン 2 と、ユーザの発話と同期した 1 次元の時系列を取得する補助センサ 3 とを備えている。補助センサ 3 で取得された信号を教示情報とする教示つき音源抽出をマイクロホン観測信号に対して行うことで、妨害音が音声である場合はユーザの発話のみを選択的に抽出することができ、妨害音が非音声である場合は教示情報がない場合とくらべて入力データの情報量が増える分だけ高精度な抽出が可能となる。

教示つき音源抽出においては、マイクロホン観測信号および教示情報という入力データからクリーンな目的音への対応関係が予め学習されているモデ

ルを用いる。そのため、学習時に使用されたデータと同程度であれば教示情報には妨害音が含まれていても良い。更に、教示情報は音でも良いし音以外でも良い。即ち、教示情報が音であるという制限が不要となるため、発話に同期した任意の一次元の時系列信号を教示情報として使用することができる。

また、本実施の形態によれば、最小のセンサ数は気導マイクロホン2と補助センサ3との2個である。そのため、多数の気導マイクロホンを用いたビームフォーミング処理によって音源抽出を実現する場合と比べ、システム自体を小型化することができる。また、補助センサ3を携帯できるので、様々な場面に実施の形態を適用することができる。

例えば、教示情報として、1次元の時系列信号ではない信号、例えば、空間情報を含む画像情報を適用することも考えられる。しかしながら、発話しているユーザの顔画像（口元）を捉えるカメラをユーザ自身が装着し、移動しうるユーザの顔画像を常に取得することは困難である。これに対して実施の形態で使用される教示情報は、内耳経由で伝わってきたユーザの発話、発話者の皮膚の振動や発話者の口付近の筋肉の動き等であり、それらを観測するセンサをユーザが装着あるいは携帯することは容易である。そのため、ユーザが移動する状況であっても実施の形態を容易に適用することができる。

本実施の形態では、ユーザの発話と同期した信号を教示情報として使用するため、ユーザのクリーンな音声取得できない場合にも高精度の抽出を行うことができる。そのため、一つの信号処理装置10を複数人で共有したり、不特定多数が短時間ずつ使用したりするといったことも容易に実現することができる。

[0081] <2. 変形例>

以上、本開示の実施の形態について具体的に説明したが、本開示の内容は上述した実施の形態に限定されるものではなく、本開示の技術的思想に基づく各種の変形が可能である。以下、変形例について説明する。なお、変形例の説明において、上述した実施の形態にかかる構成と同一又は同質の構成に

については同一の参照符号を付し、重複した説明を適宜、省略する。

[0082] [変形例 1]

変形例 1 は、教示つき音源抽出と発話区間推定とを同時に推定する例である。上述した実施の形態では、音源抽出部 1 2 によって抽出結果を生成し、その抽出結果に基づいて発話区間推定部 1 4 C が発話区間情報を生成していたが、変形例 1 では、抽出結果が生成されると共に発話区間情報が生成される。

[0083] このような同時推定を行う理由は、妨害音も音声である場合における発話区間推定の精度を向上させるためである。この点について図 2 を参照して説明する。目的音だけでなく妨害音も音声である場合は、妨害音が非音声である場合よりも認識精度が大きく低下することがあり、その原因の一つは発話区間推定の失敗である。入力音が音声らしいか否かで発話区間を推定する方式では、目的音も妨害音も共に音声である場合は両者を区別できないため、妨害音のみが存在する区間も発話区間として検出され、それが認識誤りに繋がる。例えば、目的音の前後の時間に存在する妨害音を含んだ長い区間が発話区間として検出された結果、本来の目的音に由来する単語列の前後に妨害音由来の余計な単語列が連結したものが認識結果として得られたり、妨害音のみが鳴っているときにその部分が発話区間として検出された結果、余計な認識結果が生成されたりする場合などがある。

[0084] 音源抽出部 1 2 の抽出結果に対して発話区間推定を行う場合も、抽出結果に妨害音の消し残りが存在する限りは同じ問題が発生する可能性がある。すなわち、抽出結果は妨害音が完全に除去された理想的な信号（図 2 D 参照）とは限らず、目的音の前後に妨害音に由来する小さな音量の音声連結している場合がある。そのような信号に対して発話区間推定を行うと、本当の目的音よりも長い区間が発話区間として推定されたり、妨害音の消し残りが発話区間として検出されたりしてしまう可能性がある。

[0085] 発話区間推定部 1 4 C は、音源抽出部 1 2 の出力である抽出結果に加えて補助センサ 3 に由来する教示情報も使用することで区間推定精度の向上を意

図しているが、音声である妨害音が教示情報にも混入している場合（例えば図2Bにおいて妨害音4Bも音声である場合）には、本来の発話よりも長い区間が発話区間として推定される可能性が依然として残る。

[0086] そこで、ニューラルネットワークの学習時に、マイクロホン観測信号と教示情報という両入力からクリーンな目的音への対応関係を学習するだけでなく、両入力から発話区間の内外どちらであるかの判別結果への対応関係も合わせて学習されるようにする。そして、信号処理装置の使用時には、抽出結果の生成と発話区間の判別とが同時に行われるようにする（2種類の情報を出力する）ことで、上述した問題を解決する。即ち、抽出結果内に音声である妨害音の消し残りが存在していても、そのタイミングにおけるもう一方の出力が「発話区間の外側である」という判別結果であれば、妨害音が単独で存在する部分が発話区間として推定されてしまう問題を回避することができる。

[0087] 図18は、変形例1にかかる信号処理装置（信号処理装置10A）の構成例を示す図である。図18に示す信号処理装置10Aと、図6に具体的に示した信号処理装置10との相違は、信号処理装置10にかかる音源抽出部12と発話区間推定部14Cとが統合されて、音源抽出・発話区間推定部52というモジュールに置き換わっていることである。音源抽出・発話区間推定部52の出力は2系統ある。一方は音源抽出結果であり、この音源抽出結果が音声認識部14Dに送られる。他方は発話区間情報であり、この発話区間情報も音声認識部14Dに送られる。

[0088] 音源抽出・発話区間推定部52の詳細を図19に示す。音源抽出・発話区間推定部52と音源抽出部12との違いは、抽出モデル部12Cが抽出・検出モデル部12Fに置き換わったことと区間トラッキング部12Gが新たに設けられたことであり、それ以外のモジュールは音源抽出部12のモジュールと同一である。

[0089] 抽出・検出モデル部12Fの出力は2系統ある。一方の出力は再構成部12Dに出力され、音源抽出結果である目的音信号が生成される。他方の出力

は、区間トラッキング部 1 2 G に送られる。後者のデータは発話検出の判別結果であり、例えばフレームごとに二値化された判別結果である。即ち、そのフレームにおいてユーザの発話が存在する／しないを、「1」または「0」の値でそれぞれ表現する。発話の有無であって音声の有無ではないので、ユーザが発話していないタイミングにおいて音声である妨害音が発生した場合の理想値は「0」である。

[0090] 区間トラッキング部 1 2 G は、フレームごとの判別結果を時間方向にトラッキングすることで発話区間情報である発話開始時刻・終了時刻をそれぞれ求める。処理の例としては、1 という判定結果が所定の時間長以上に渡って継続したら発話の始端と見なし、同様に 0 という判定結果が所定の時間長以上継続したら発話の終端と見なす。あるいは、そのような規則に基づく方法の代わりに、ニューラルネットワークを使用した学習に基づく公知の方法によってトラッキングを行っても良い。

[0091] 上述した例では抽出・検出モデル部 1 2 F から出力される判別結果が二値であると説明したが、代わりに連続値を出力し、区間トラッキング部 1 2 G において所定の閾値によって二値化を行っても良い。こうして求めた音源抽出結果及び発話区間情報が音声認識部 1 4 D に送られる。

[0092] 次に、図 20 を参照して、抽出・検出モデル部 1 2 F の詳細について説明する。抽出・検出モデル部 1 2 F が、抽出モデル部 1 2 C と相違する点は、出力層が 2 種類（出力層 1 2 1 F 及び出力層 1 2 2 F）存在することである。出力層 1 2 1 F は、抽出モデル部 1 2 C の出力層 1 2 4 C と同様に動作することで、音源抽出結果に相当するデータを出力する。一方、出力層 1 2 2 F は、発話検出の判別結果を出力する。具体的には、フレームごとに二値化された判別結果などである。

[0093] 図 20 においては、出力側の分岐はその一つ前の層である中間層 n で発生しているが、それよりも入力層に近い側の中間層において分岐を発生させても良い。その場合、分岐が発生した中間層から各出力層までの層の個数はそれぞれ異なっても良く、一つの例として、出力データの一方が中間層か

ら出力されるようなネットワーク構造を用いても構わない。

- [0094] 次に、抽出・検出モデル部12Fの学習系について、図21を用いて説明する。抽出・検出モデル部12Fは抽出モデル部12Cと異なり2種類のデータを出力するため、抽出モデル部12Cとは異なる学習を行う必要がある。複数種類のデータを出力するニューラルネットワークを学習することはマルチタスク学習と呼ばれており、図21はマルチタスク学習機の一つである。マルチタスク学習については、公知の方法を適用できる。
- [0095] 目的音データセット61は、以下の(a)から(c)の3つの信号の組で構成される集合である。それは、(a)目的音波形(目的音である音声発話およびその前後に連結された所定の長さの無音からなる音波形)、(b)(a)と同期した教示情報、(c)(a)と同期した発話判別フラグである。
- [0096] 上記(c)の例として、(a)を所定の時間間隔(例えば図9の短時間フーリエ変換のシフト幅と同じ時間間隔)に分割した上で、各時間間隔内に発話が存在すれば「1」の値を、しなければ「0」の値を付与することで生成されるビット列などが考えられる。
- [0097] 学習時は、目的音データセット61からランダムに1個の組が取り出され、その内の教示情報が混合部64(教示情報が気導マイクロホンによって取得された場合)又は特徴量生成部65(それ以外の場合)へ、目的音波形が混合部63及び教師データ生成部66へ、発話判別フラグが教師データ生成部67へそれぞれ出力される。また、妨害音データセット62からランダムに1個以上の音波形が取り出され、取り出された音波形が、混合部63に送られる。教示情報が気導マイクロホンによって取得された場合には、妨害音の音波形は混合部64にも送られる。
- [0098] 抽出・検出モデル部12Fは2種類のデータを出力するため、それぞれについての教師データを用意する。教師データ生成部66は、音源抽出結果に対応した教師データを生成する。教師データ生成部67は、発話検出結果に対応した教師データを生成する。発話判別フラグが前述のようなビット列である場合は、それをそのまま教師データとして使用することができる。以降

では、教師データ生成部66により生成された教師データを教師データ1D、教師データ生成部67で生成された教師データを教師データ2Dとそれぞれ称する。

[0099] 抽出・検出モデル部12Fの出力が2種類あるため、比較部も2つ必要である。2種類の出力の内、音源抽出結果に相当する出力は比較部70に出力され、比較部70により教師データ1Dと比較される。比較部70の動作は、上述した図12における比較部27と同一である。一方、発話検出結果に相当する出力は比較部71に出力され、比較部71により教師データ2Dと比較される。比較部71も比較部70と同様に損失関数を使用するが、こちらは二値の分類器を学習するための損失関数である。

[0100] パラメータ更新値計算部72は、2つの比較部70及び比較部71で計算されたそれぞれの損失値から、損失値が減少するように抽出・検出モデル部12Fのパラメータについての更新値を計算する。マルチタスク学習におけるパラメータ更新方法については公知の方法を利用できる。

[0101] [変形例2]

上述した変形例1においては、音源抽出結果と発話区間情報とを個別に音声認識部14D側に送り、音声認識部14D側で発話区間ごとへの分割と認識結果である単語列の生成とを行うことを想定していた。これに対して、変形例2は、音源抽出結果と発話区間情報とを統合したデータを一旦生成し、生成したデータを出力するようにしても良い。以下、変形例2について説明する。

[0102] 図22は、変形例2にかかる信号処理装置（信号処理装置10B）の構成例を示す図である。信号処理装置10Bが信号処理装置10Aと異なる点は、信号処理装置10Bでは、音源抽出・発話区間推定部52から出力される2種類のデータ（音源抽出結果及び発話区間情報）が区間外無音化部55に入力される点、及び、区間外無音化部55の出力が新たに設けられた発話分割部14H又は音声認識部14Dに入力される点である。その他の構成は、信号処理装置10Aの構成と同一である。

- [0103] 区間外無音化部55は、音信号である音源抽出結果に対して発話区間情報を適用することで新たな音信号を生成する。具体的には、区間外無音化部55は、発話区間外の時間に対応した音信号について、その部分を無音あるいは無音に近い音に置き換える処理を行う。無音に近い音とは、音源抽出結果に対して0に近い正の定数を乗じた信号などである。また、音の再生を行わない場合は、無音に置き換える代わりに、後段の発話分割部14Hや音声認識部14Dに悪影響を与えない種類のノイズに置き換えても良い。
- [0104] 区間外無音化部55の出力は切れ目のないストリームであり、それを音声認識部14Dに入力するために以下の(1)または(2)のどちらかの方法で対応する。(1)区間外無音化部55と音声認識部14Dとの間に、発話分割部14Hを追加する。(2)逐次音声認識と呼ばれる、ストリーム入力に対応した音声認識を用いる。(2)の場合、発話分割部14Hはなくても良い。発話分割部14Hとしては、公知の方法(例えば、特許第418244号に記載の方法)を適用することができる。
- [0105] 逐次音声認識は、公知の方法(例えば、特開2012-226068号公報に記載の方法)を適用することができる。区間外無音化部55が動作することにより、ユーザが発話している区間以外は無音(あるいは後段の動作に悪影響を与えない音)という音信号が入力されるため、それが入力される発話分割部14H又は音声認識部14Dは、音源抽出結果が直接入力される場合よりも正確な動作をすることが可能となる。また、音源・発話区間推定部52の後段に区間外無音化部55を設けることにより、逐次音声認識機を備えたシステムに対してだけでなく、発話分割部14Hと音声認識部14Dとが一体化されたシステムに対しても、本開示の教示つき音源抽出を適用することができる。
- [0106] 音源抽出結果に対して発話区間推定を行うと、妨害音も音声である場合に発話区間推定が妨害音の消し残りにも反応してしまい、誤認識に繋がったり、余計な認識結果が生成されたりする場合がある。変形例では、音源抽出と発話区間推定という2つの推定処理を同時に行うことで、音源抽出結果に妨

害音の消し残りが含まれていても、それとは独立に正確な発話区間推定が行われ、結果として音声認識精度を向上させることができる。

[0107] [その他の変形例]

その他の変形例について説明する。

上述した信号処理装置における全部又は一部の処理がクラウド上のサーバ等で行われても良い。また、目的音は人が発した音声以外（例えば、ロボットやペットの声）でも良い。また、補助センサは、人物以外のロボットやペットに装着されても良い。また、補助センサは、異なる種類の複数の補助センサでも良く、信号処理装置が使用される環境に応じて、使用される補助センサが切り替えられるようにしても良い。また、本開示は、オブジェクト毎の音源を生成する際にも適用することができる。

なお、図12の「混合部24」および図21の「混合部64」は、補助センサの種類によっては省略可能であるので、図12の「混合部24」および図21の「混合部64」のそれぞれには括弧を付している。

[0108] なお、本開示中に例示された効果により本開示の内容が限定して解釈されるものではない。

[0109] 本開示は、以下の構成も採ることができる。

(1)

目的音と前記目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、前記目的音と同期している1次元の時系列信号とが入力される入力部と、

前記1次元の時系列信号に基づいて、前記マイクロホン信号から前記目的音に対応する目的音信号を抽出する音源抽出部と

を有する信号処理装置。

(2)

前記音源抽出部は、前記1次元の時系列信号に基づいて生成される教示情報を使用して、前記目的音信号を抽出する

(1)に記載の信号処理装置。

(3)

前記補助センサは、前記目的音の発生源に装着されるセンサである

(1) 又は (2) に記載の信号処理装置。

(4)

前記マイクロホン信号は、第1のマイクロホンによって検出される信号であり、

前記補助センサは、前記第1のマイクロホンと異なる第2のマイクロホンである

(1) から (3) までの何れかに記載の信号処理装置。

(5)

前記第1のマイクロホンは、ヘッドホンのハウジングの外側に設けられるマイクロホンであり、前記第2のマイクロホンは、前記ハウジングの内部に設けられるマイクロホンである

(4) に記載の信号処理装置。

(6)

前記補助センサは、体内を伝播した音波を検出するセンサである

(1) から (4) に記載の信号処理装置。

(7)

前記補助センサは、音波以外の信号を検出するセンサである

(1) から (4) に記載の信号処理装置。

(8)

前記補助センサは、筋肉の動きを検出するセンサである

(7) に記載の信号処理装置。

(9)

前記音源抽出部により抽出された前記目的音信号を再生する再生部を有する

(1) から (8) までの何れかに記載の信号処理装置。

(10)

前記音源抽出部により抽出された前記目的音信号を外部機器に送信する通信部を有する

(1) から (8) までの何れかに記載の信号処理装置。

(11)

前記音源抽出部による抽出結果に基づいて、発話の有無を示す発話区間を推定し、その結果である発話区間情報を生成する発話区間推定部と、

前記発話区間における音声認識を行う音声認識部を有する

(1) から (8) までの何れかに記載の信号処理装置。

(12)

前記音源抽出部は、更に、発話の有無を示す発話区間を推定し、その結果である発話区間情報を生成する音源抽出・発話区間推定部として構成されており、

前記音源抽出・発話区間推定部は、前記目的音信号及び前記発話区間情報を出力する

(1) から (8) までの何れかに記載の信号処理装置。

(13)

前記音源抽出・発話区間推定部から出力される発話区間情報に基づいて、前記目的音信号における発話区間外の時間に対応する音信号を判定し、判定した前記音信号を無音化する区間外無音化部を有する

(12) に記載の信号処理装置。

(14)

前記音源抽出部は、前記マイクロホン信号に基づく第1の特徴量と、前記1次元の時系列信号に基づく第2の特徴量を入力とし、当該入力に対して順方向伝播処理を行うことにより、出力特徴量を出力する抽出モデル部を有する

(1) から (8) の何れか、(11) 又は (12) に記載の信号処理装置。

(15)

前記音源抽出部は、前記マイクロホン信号に基づく第1の特徴量と、前記1次元の時系列信号に基づく第2の特徴量を入力とし、当該入力に対して順方向伝播処理を行うことにより、複数の出力特徴量を出力する抽出・検出モデル部を有する

(1) から (8) の何れか、(12) 又は (13) に記載の信号処理装置。

(16)

前記出力特徴量に基づいて少なくとも前記目的音信号を生成する再構成部を有する

(14) 又は (15) に記載の信号処理装置。

(17)

入力特徴量から前記出力特徴量への対応関係が予め学習されている

(14) 又は (15) に記載の信号処理装置。

(18)

目的音と前記目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、前記目的音と同期している1次元の時系列信号とが入力部に入力され、

前記1次元の時系列信号に基づいて、前記マイクロホン信号から前記目的音に対応する目的音信号が音源抽出部により抽出される

信号処理方法。

(19)

目的音と前記目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、前記目的音と同期している1次元の時系列信号とが入力部に入力され、

前記1次元の時系列信号に基づいて、前記マイクロホン信号から前記目的音に対応する目的音信号が音源抽出部により抽出される

信号処理方法をコンピュータに実行させるプログラム。

符号の説明

- [0110] 2 . . . 気導マイクロホン
- 3 . . . 補助センサ
- 1 0, 1 0 A, 1 0 B . . . 信号処理装置
- 1 1 . . . 入力部
- 1 2 . . . 音源抽出部
- 1 2 C . . . 抽出モデル部
- 1 2 D . . . 再構成部
- 1 4 A . . . 音再生部
- 1 4 B . . . 通信部
- 3 2, 3 3, 4 2, 4 4 . . . マイクロホン
- 5 2 . . . 音源抽出・発話区間推定部
- 5 5 . . . 区間外無音化部

請求の範囲

- [請求項1] 目的音と前記目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、前記目的音と同期する1次元の時系列信号とが入力される入力部と、
前記1次元の時系列信号に基づいて、前記マイクロホン信号から前記目的音に対応する目的音信号を抽出する音源抽出部と
を有する信号処理装置。
- [請求項2] 前記音源抽出部は、前記1次元の時系列信号に基づいて生成される教示情報を使用して、前記目的音信号を抽出する
請求項1に記載の信号処理装置。
- [請求項3] 前記補助センサは、前記目的音の発生源に装着されるセンサである
請求項1に記載の信号処理装置。
- [請求項4] 前記マイクロホン信号は、第1のマイクロホンによって検出される信号であり、
前記補助センサは、前記第1のマイクロホンと異なる第2のマイクロホンである
請求項1に記載の信号処理装置。
- [請求項5] 前記第1のマイクロホンは、ヘッドホンのハウジングの外側に設けられるマイクロホンであり、前記第2のマイクロホンは、前記ハウジングの内部に設けられるマイクロホンである
請求項4に記載の信号処理装置。
- [請求項6] 前記補助センサは、体内を伝播した音波を検出するセンサである
請求項1に記載の信号処理装置。
- [請求項7] 前記補助センサは、音波以外の信号を検出するセンサである
請求項1に記載の信号処理装置。
- [請求項8] 前記補助センサは、筋肉の動きを検出するセンサである
請求項7に記載の信号処理装置。
- [請求項9] 前記音源抽出部により抽出された前記目的音信号を再生する再生部

を有する

請求項 1 に記載の信号処理装置。

[請求項10] 前記音源抽出部により抽出された前記目的音信号を外部機器に送信する通信部を有する

請求項 1 に記載の信号処理装置。

[請求項11] 前記音源抽出部による抽出結果に基づいて、発話の有無を示す発話区間を推定し、その結果である発話区間情報を生成する発話区間推定部と、

前記発話区間における音声認識を行う音声認識部を有する

請求項 1 に記載の信号処理装置。

[請求項12] 前記音源抽出部は、更に、発話の有無を示す発話区間を推定し、その結果である発話区間情報を生成する音源抽出・発話区間推定部として構成されており、

前記音源抽出・発話区間推定部は、前記目的音信号及び前記発話区間情報を出力する

請求項 1 に記載の信号処理装置。

[請求項13] 前記音源抽出・発話区間推定部から出力される発話区間情報に基づいて、前記目的音信号における発話区間外の時間に対応する音信号を判定し、判定した前記音信号を無音化する区間外無音化部を有する

請求項 1 2 に記載の信号処理装置。

[請求項14] 前記音源抽出部は、前記マイクロホン信号に基づく第 1 の特徴量と、前記 1 次元の時系列信号に基づく第 2 の特徴量を入力とし、当該入力に対して順方向伝播処理を行うことにより、出力特徴量を出力する抽出モデル部を有する

請求項 1 に記載の信号処理装置。

[請求項15] 前記音源抽出部は、前記マイクロホン信号に基づく第 1 の特徴量と、前記 1 次元の時系列信号に基づく第 2 の特徴量を入力とし、当該入力に対して順方向伝播処理を行うことにより、複数の出力特徴量を出

力する抽出・検出モデル部を有する

請求項 1 に記載の信号処理装置。

[請求項16] 前記出力特徴量に基づいて少なくとも前記目的音信号を生成する再構成部を有する

請求項 1 4 に記載の信号処理装置。

[請求項17] 入力特徴量から前記出力特徴量への対応関係が予め学習されている請求項 1 4 に記載の信号処理装置。

[請求項18] 目的音と前記目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、前記目的音と同期している 1 次元の時系列信号とが入力部に入力され、

前記 1 次元の時系列信号に基づいて、前記マイクロホン信号から前記目的音に対応する目的音信号が音源抽出部により抽出される

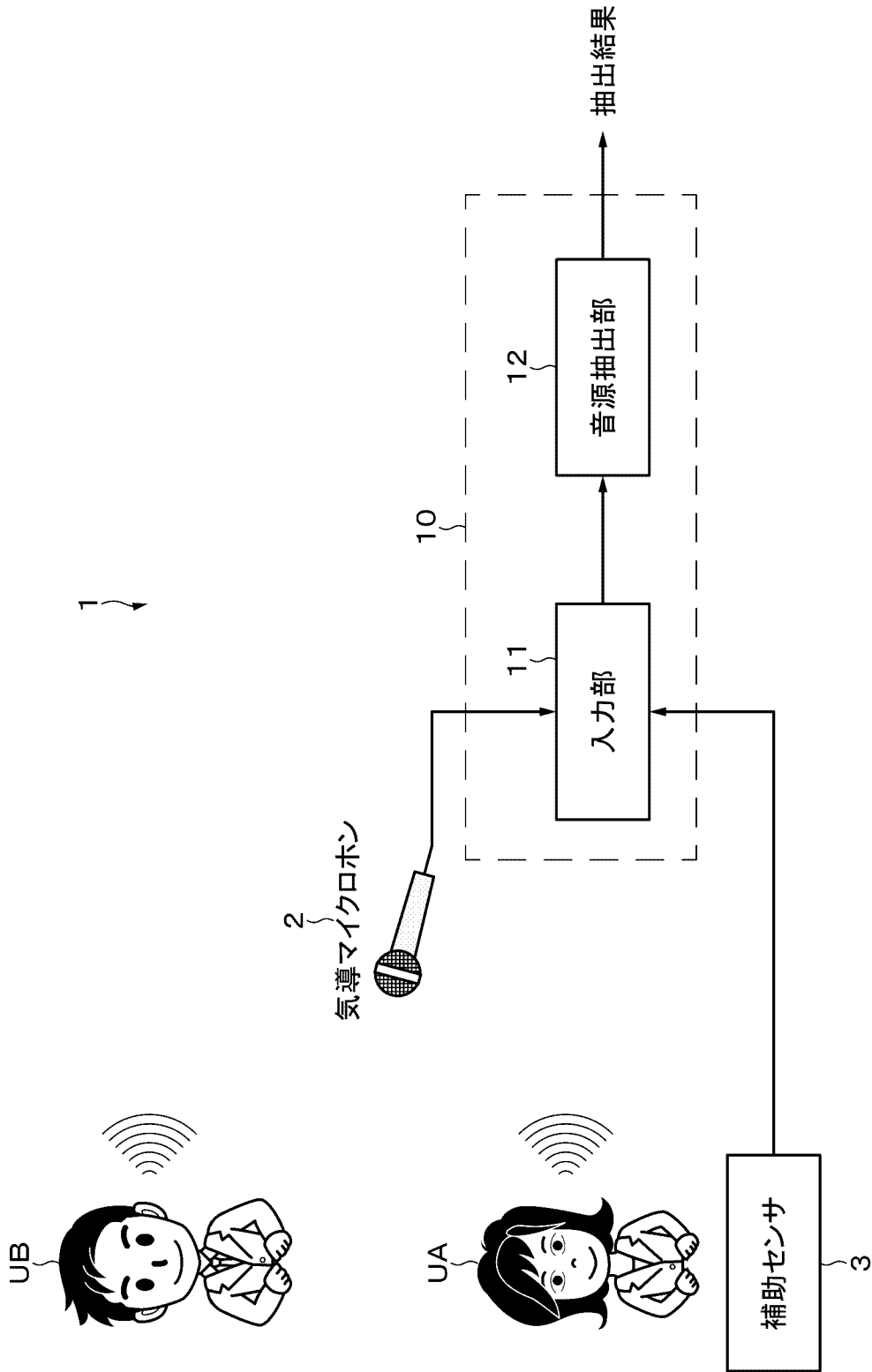
信号処理方法。

[請求項19] 目的音と前記目的音以外の音とが混合された混合音を含むマイクロホン信号と、補助センサによって取得された、前記目的音と同期している 1 次元の時系列信号とが入力部に入力され、

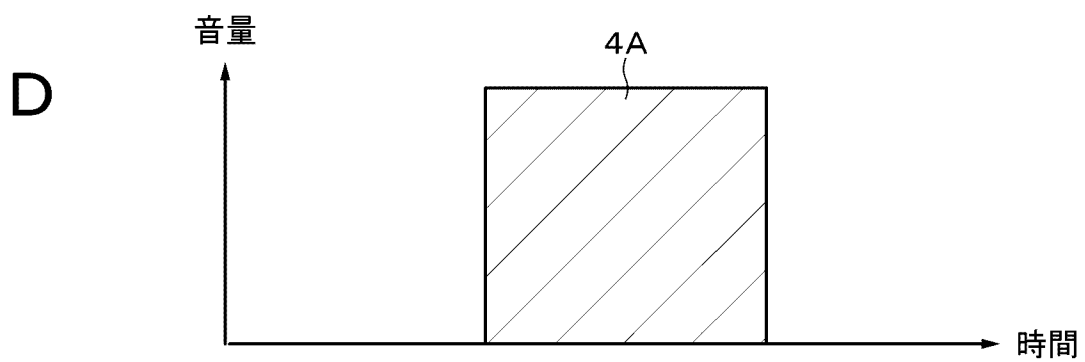
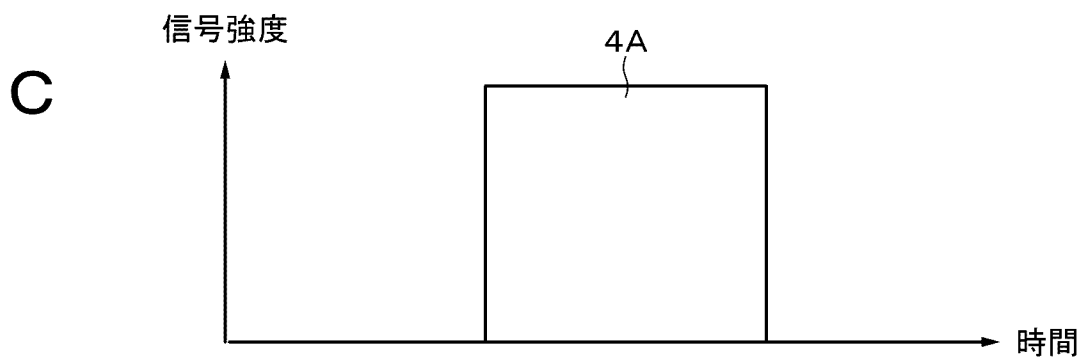
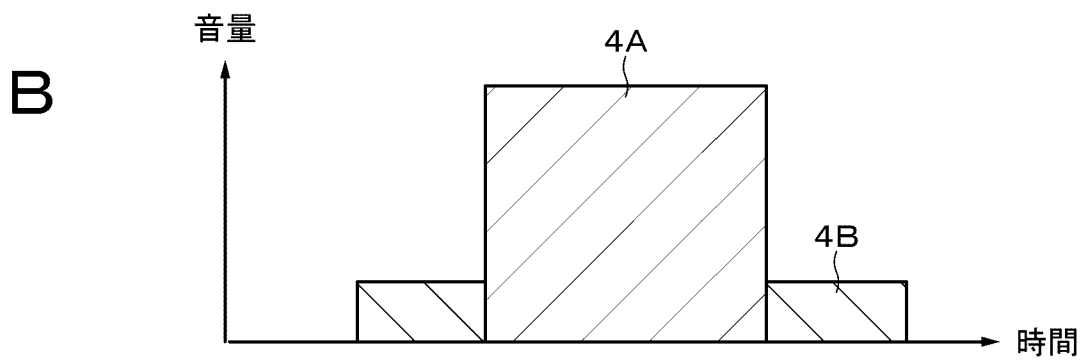
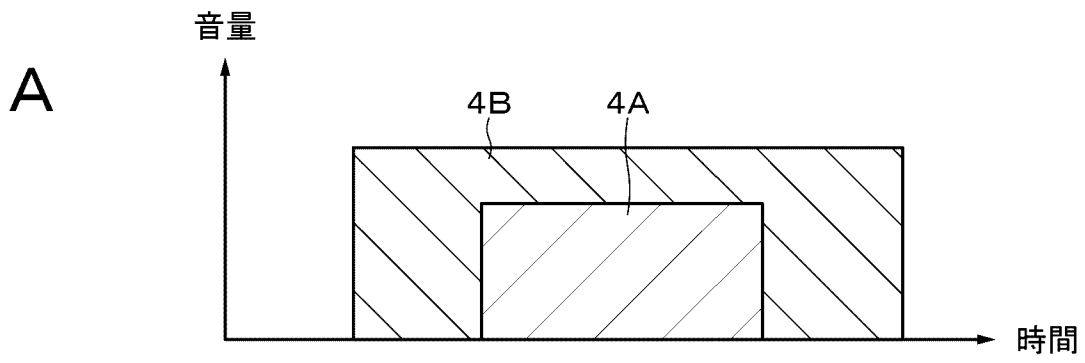
前記 1 次元の時系列信号に基づいて、前記マイクロホン信号から前記目的音に対応する目的音信号が音源抽出部により抽出される

信号処理方法をコンピュータに実行させるプログラム。

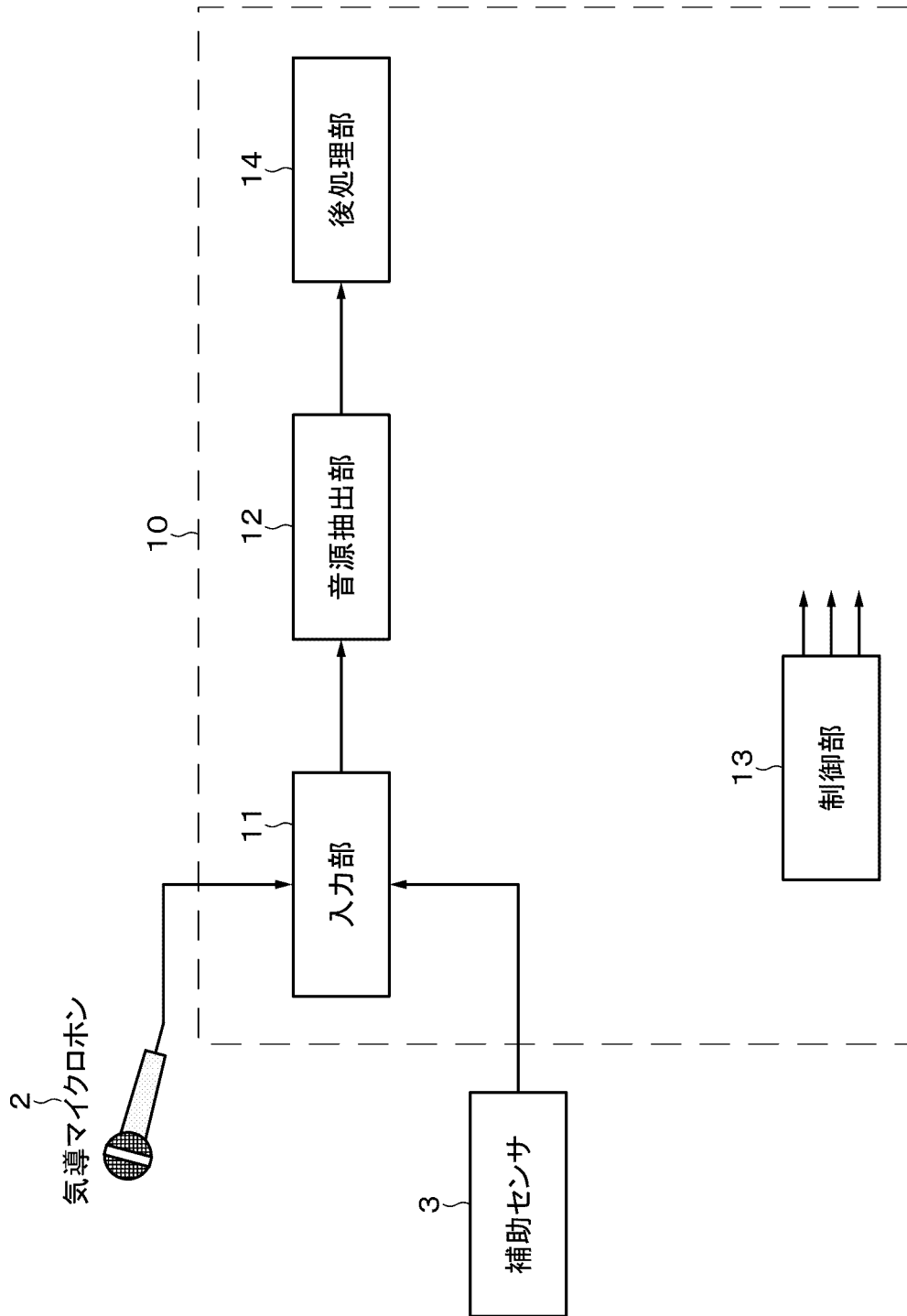
[図1]



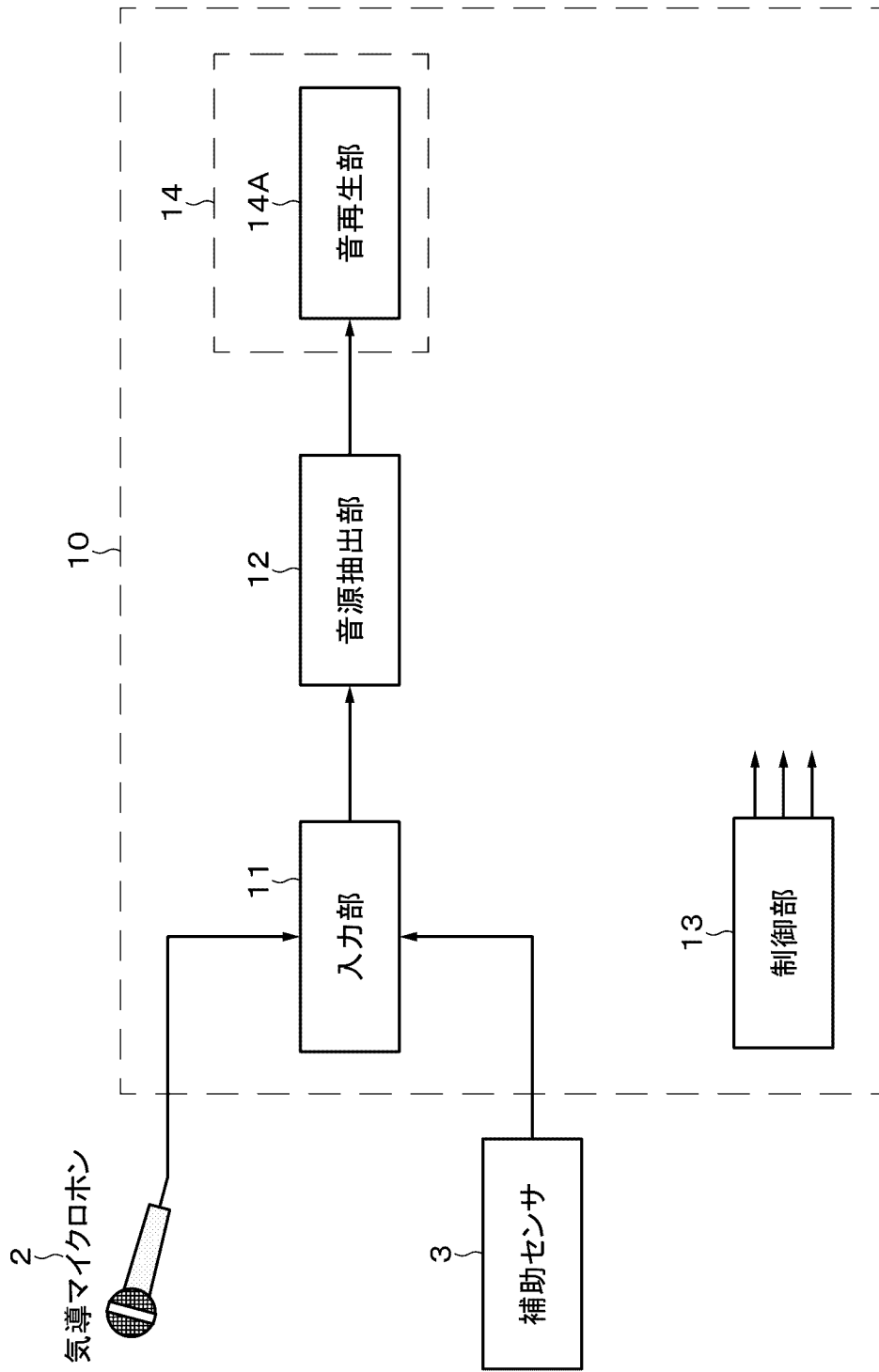
[図2]



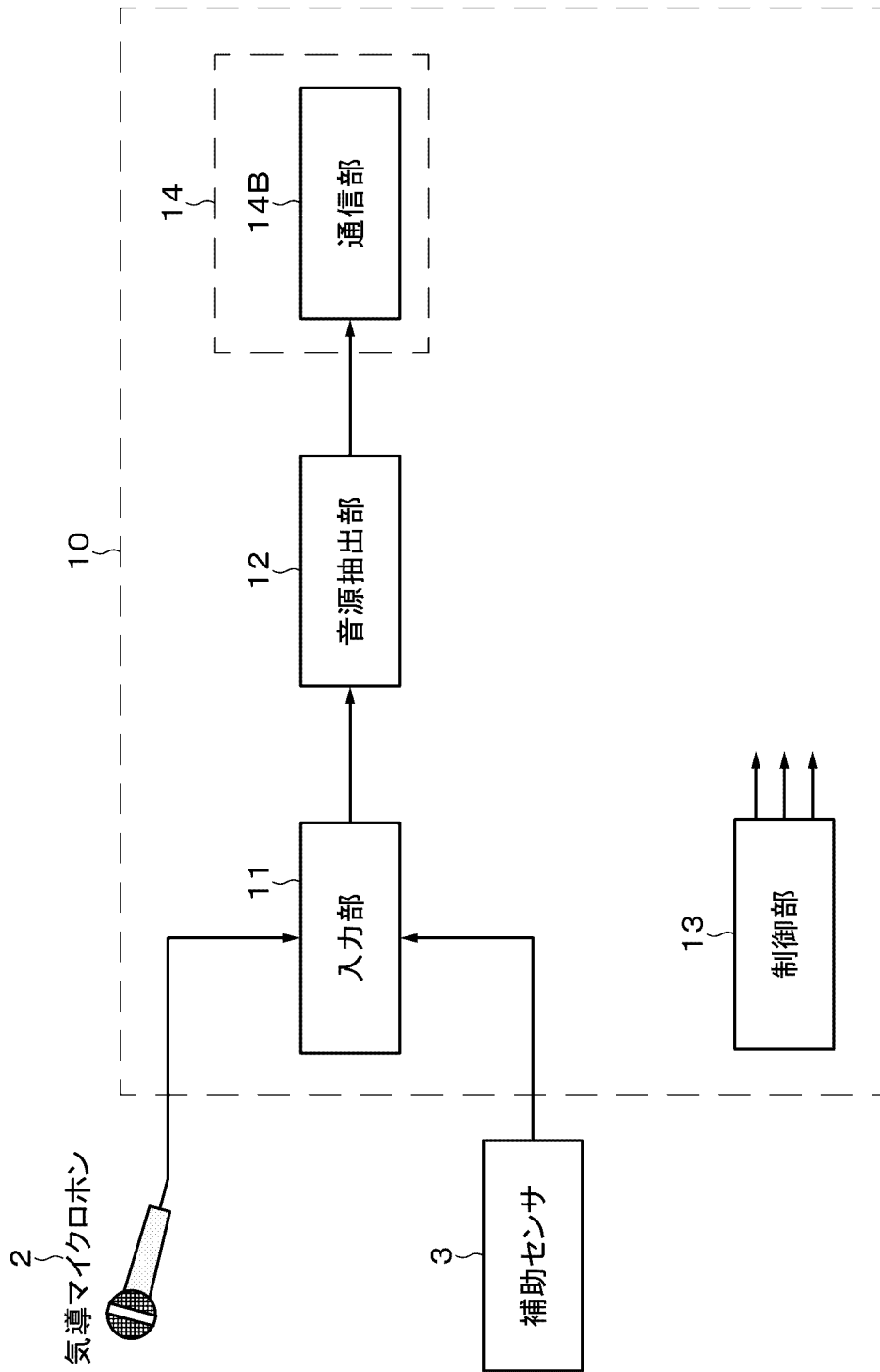
[図3]



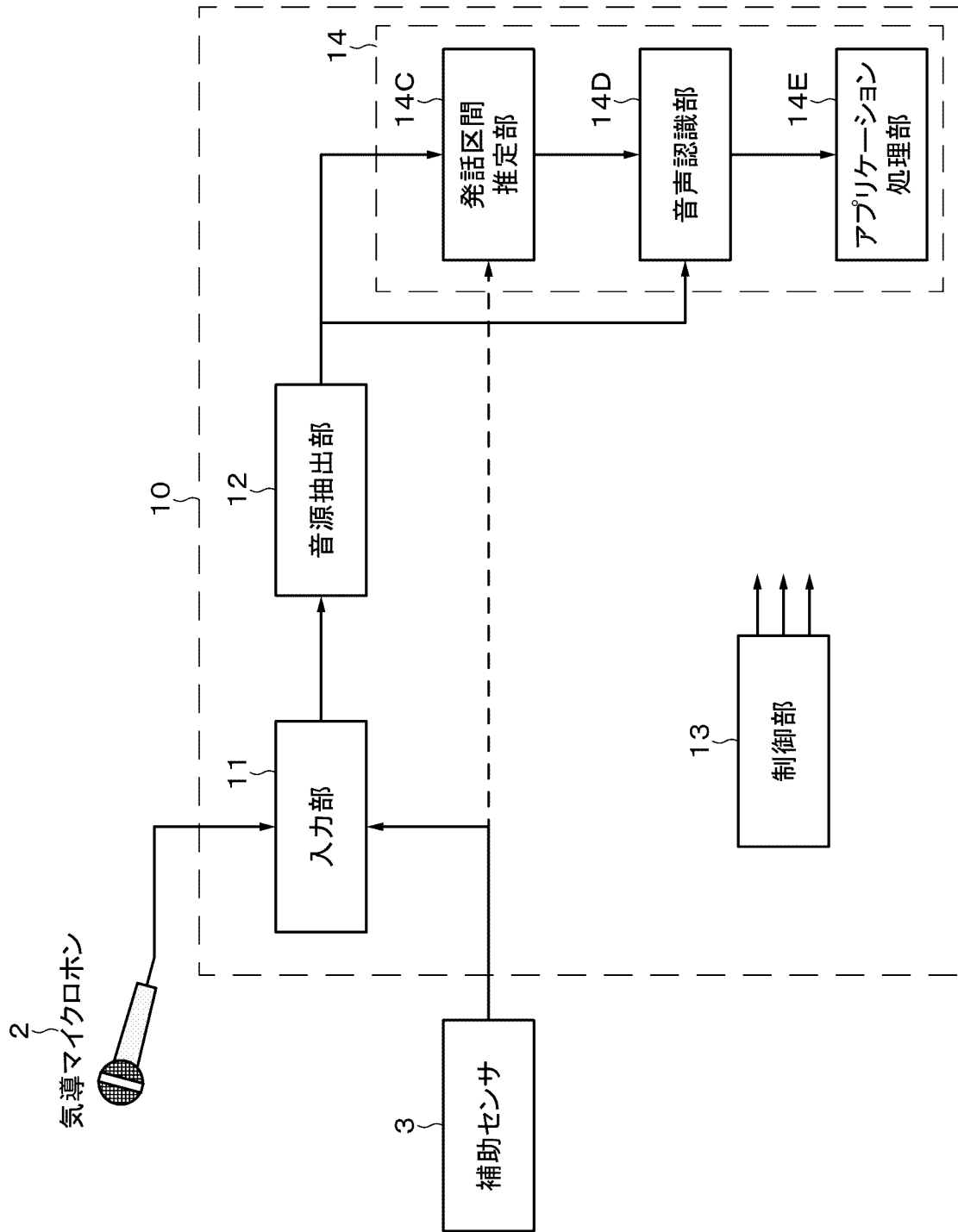
[図4]



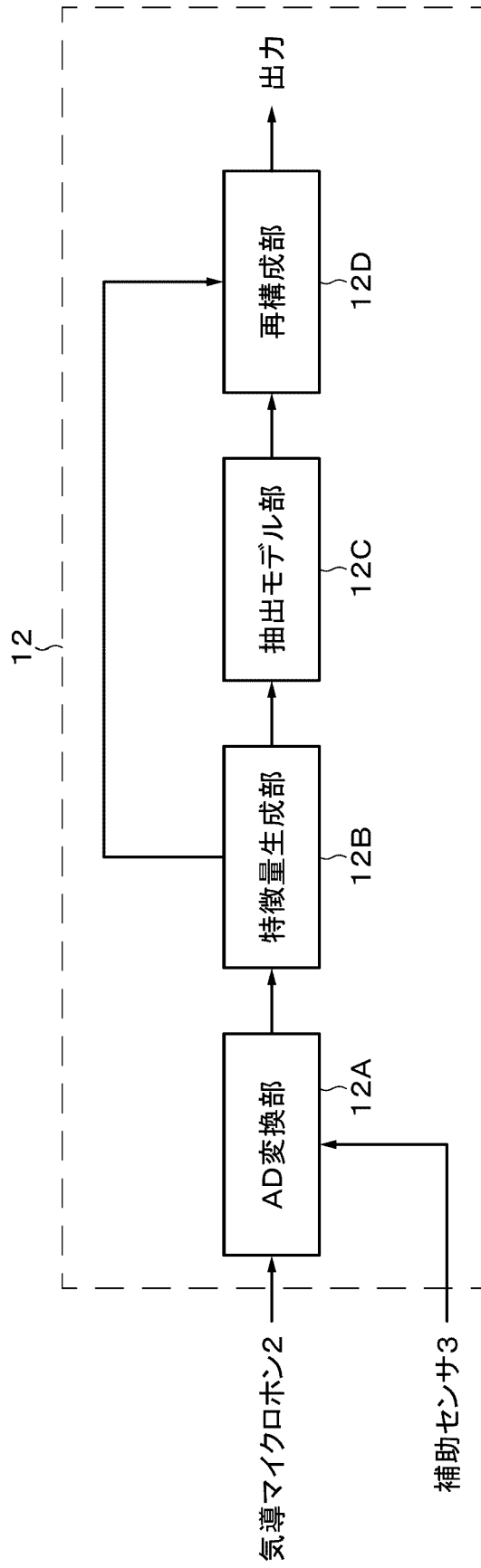
[図5]



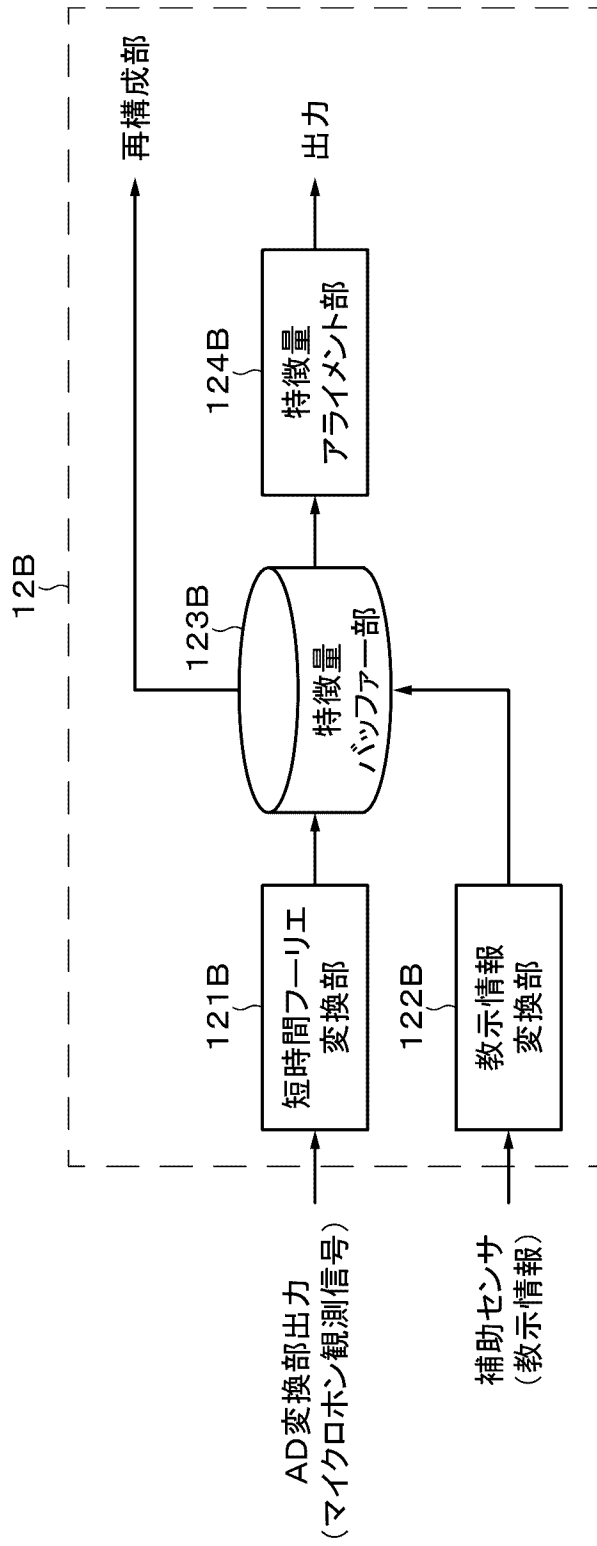
[図6]



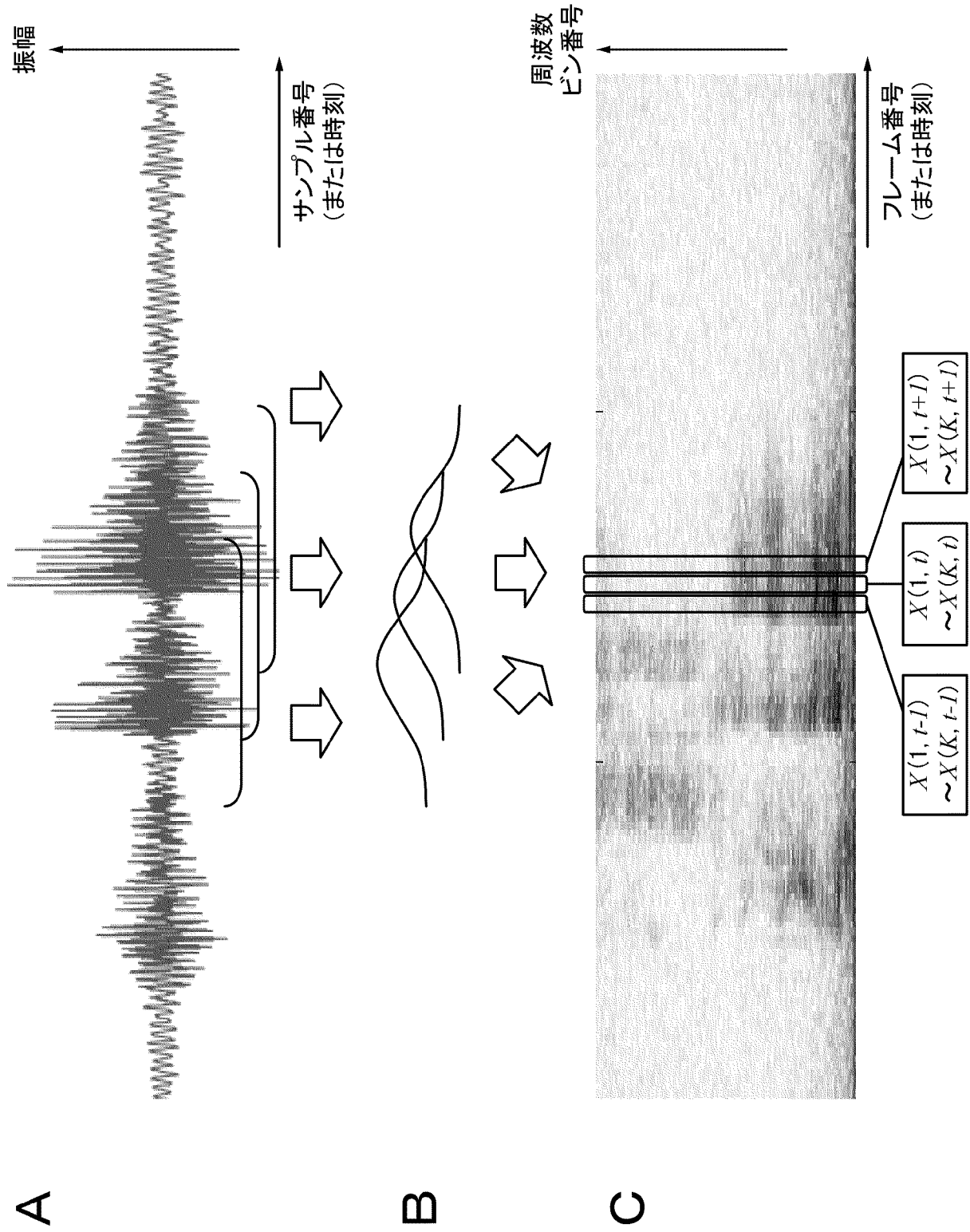
[図7]



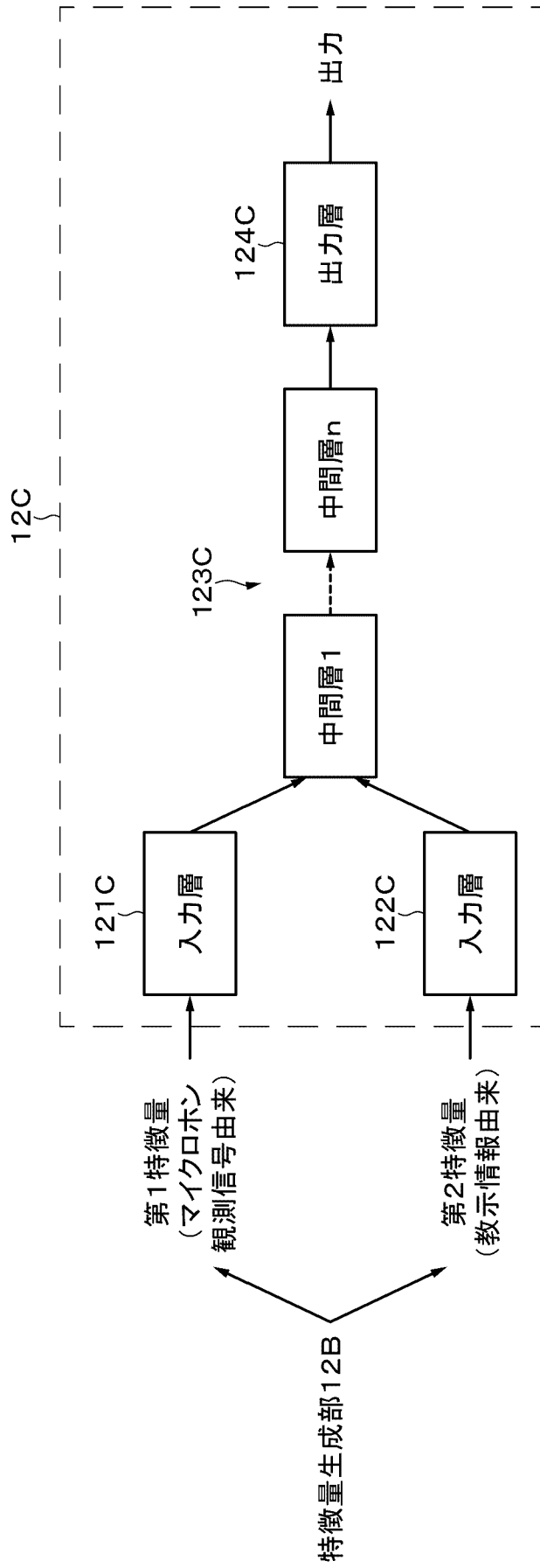
[図8]



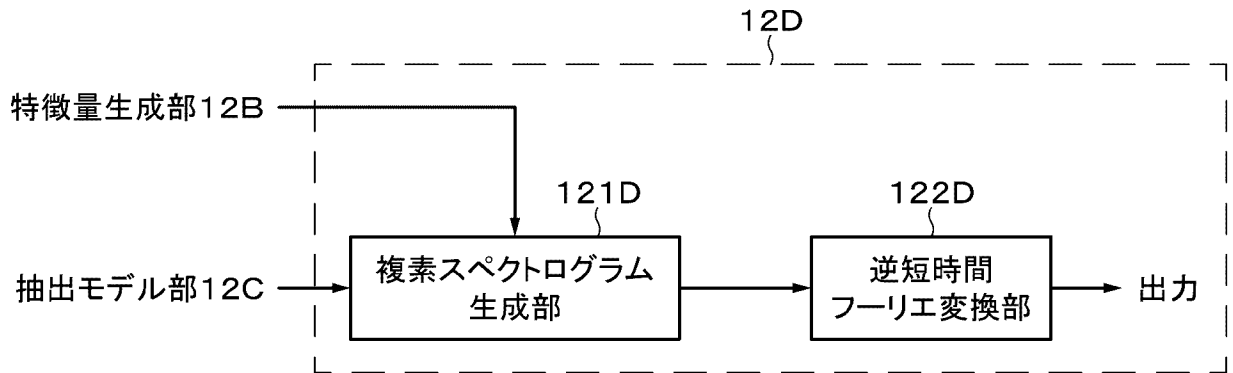
[図9]



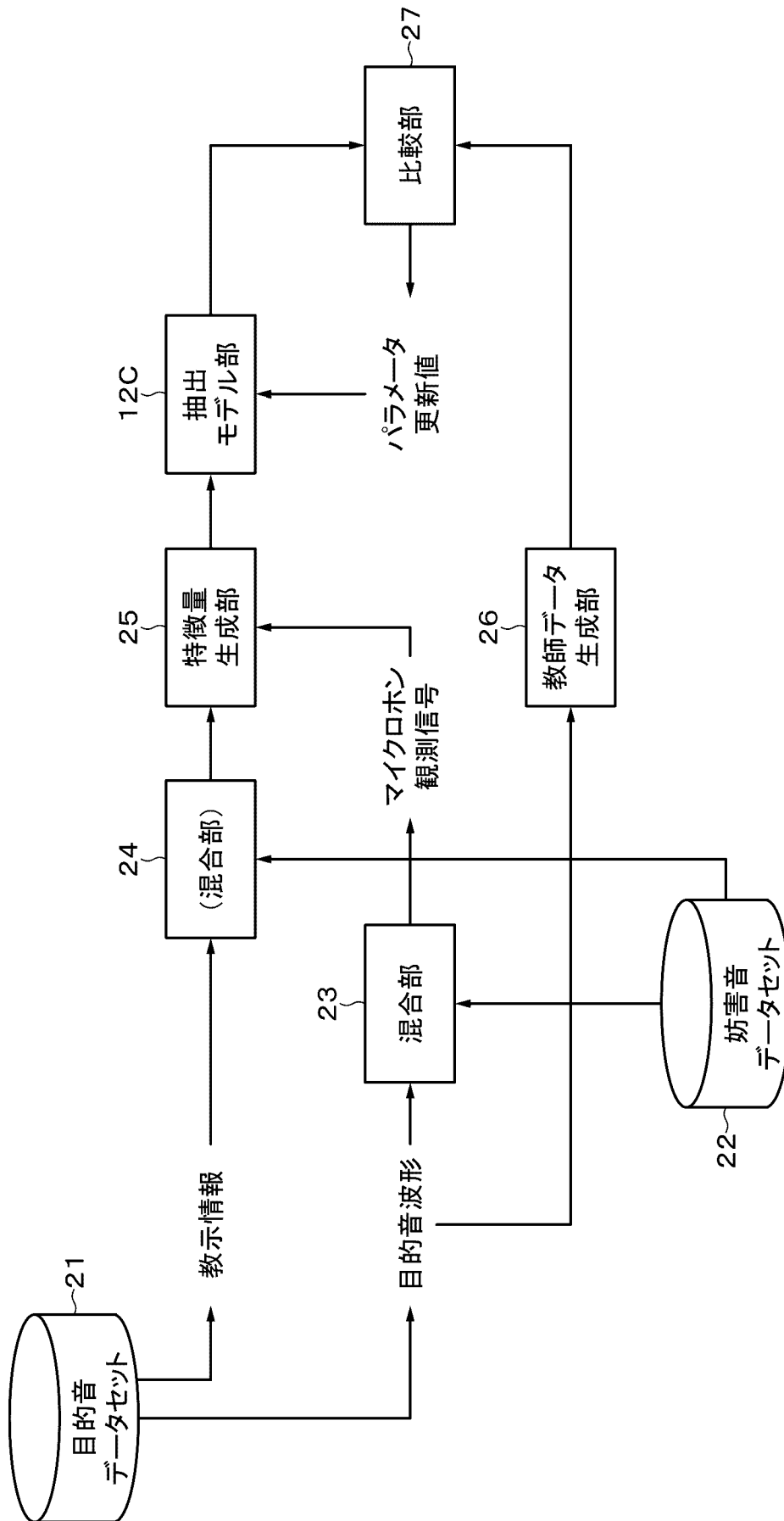
[図10]



[図11]



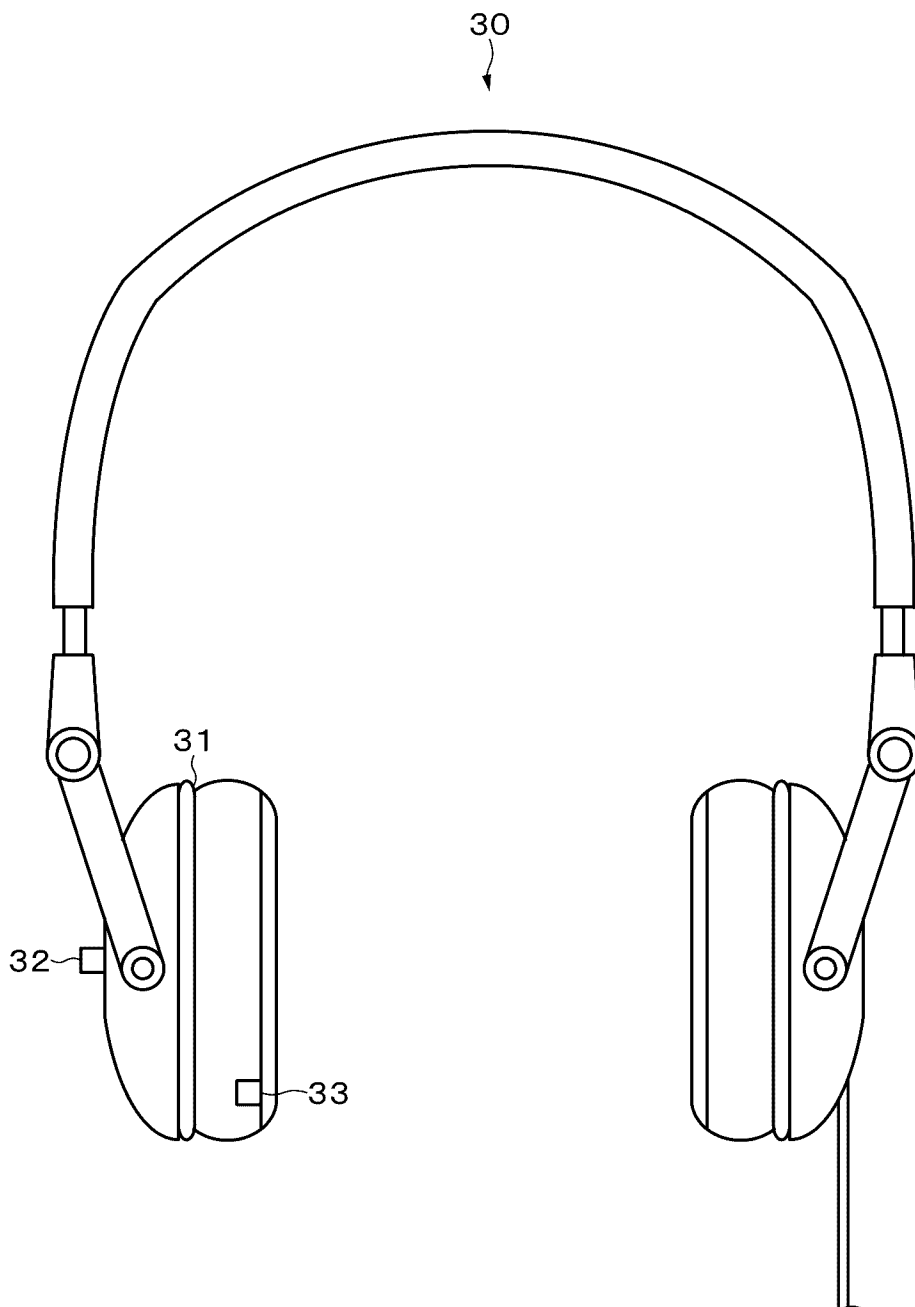
[図12]



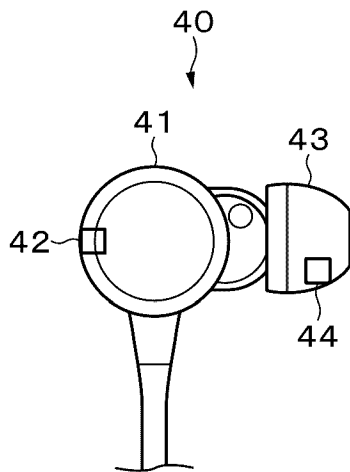
[図13]

	目的音	教示情報	妨害音	入力特徴量	教師データ
a)	有り	有り	有り	目的音+妨害音	目的音
b)	有り	有り	無し	目的音のみ	目的音
c)	無し	無し	有り	妨害音のみ	無音
d)	無し	無し	無し	無音	無音

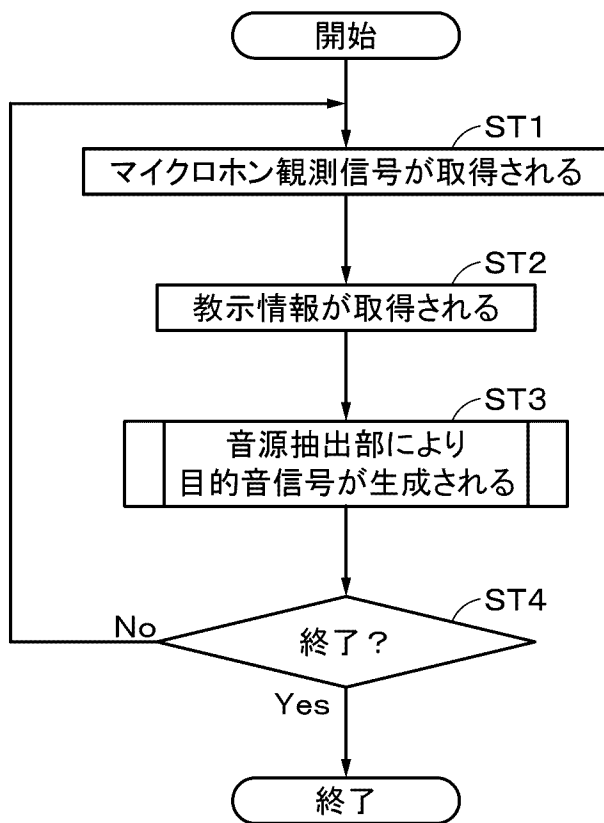
[図14]



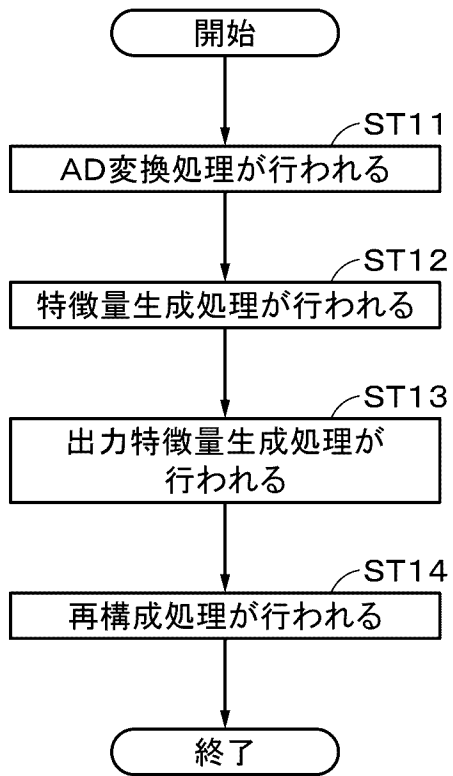
[図15]



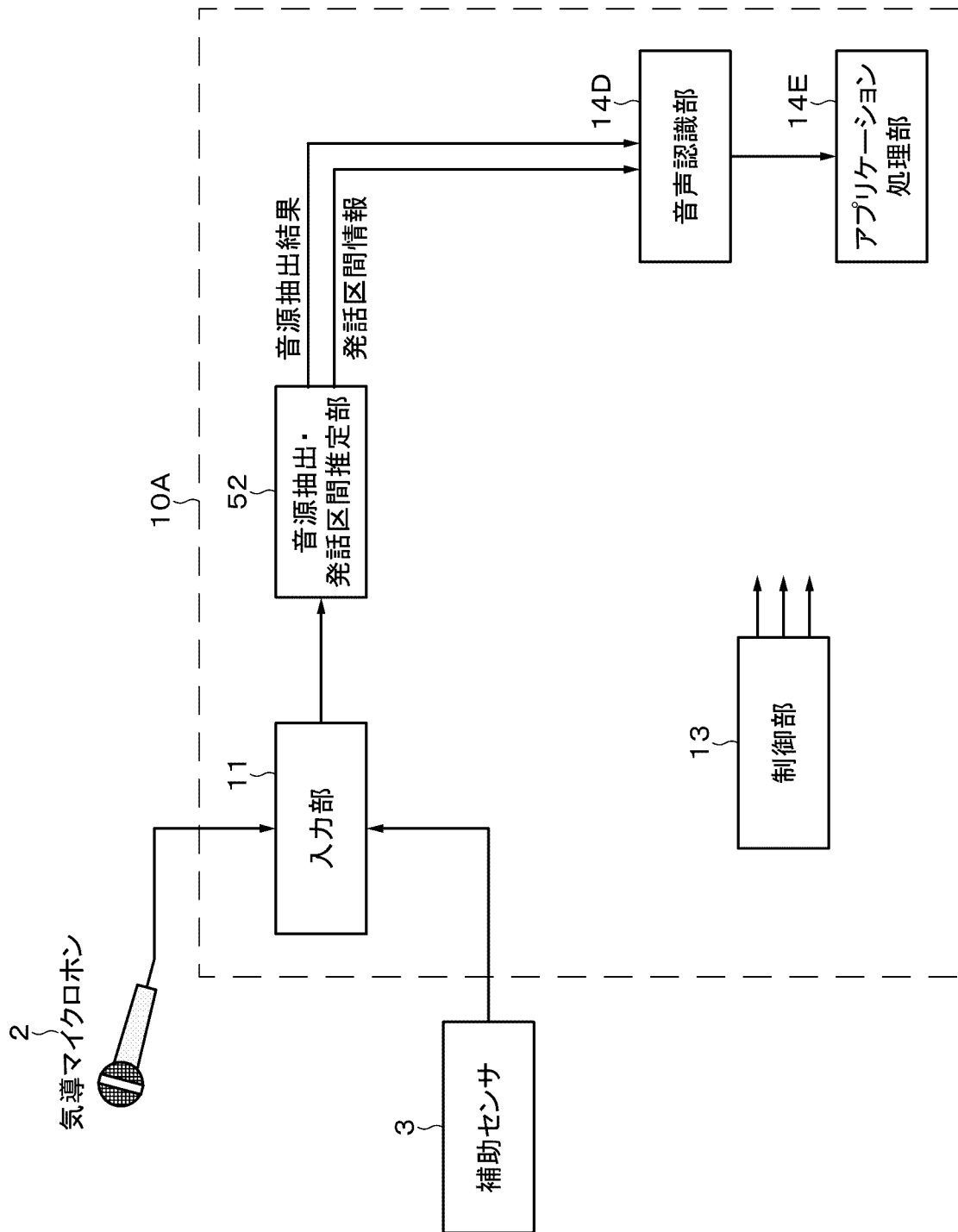
[図16]



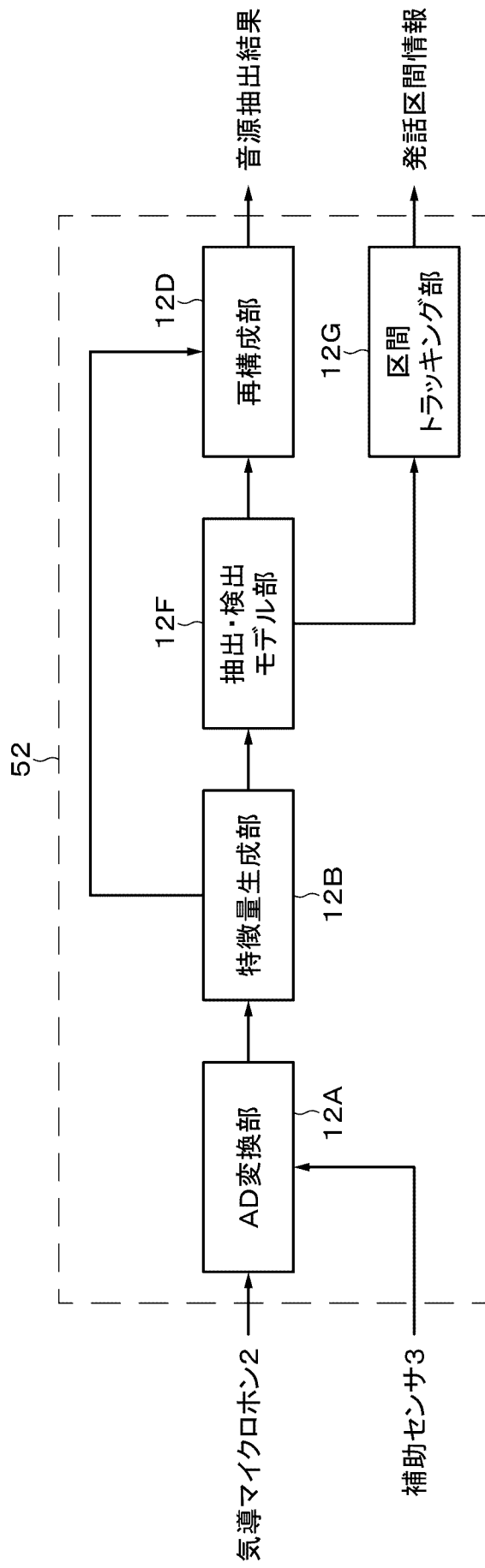
[図17]



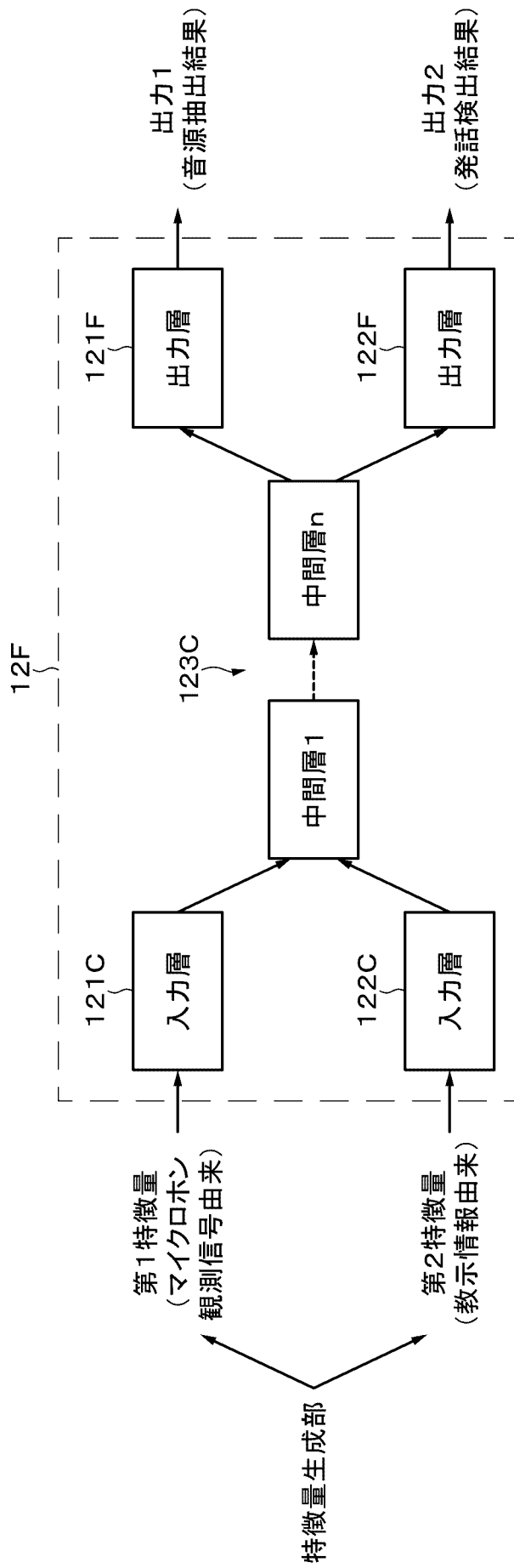
[図18]



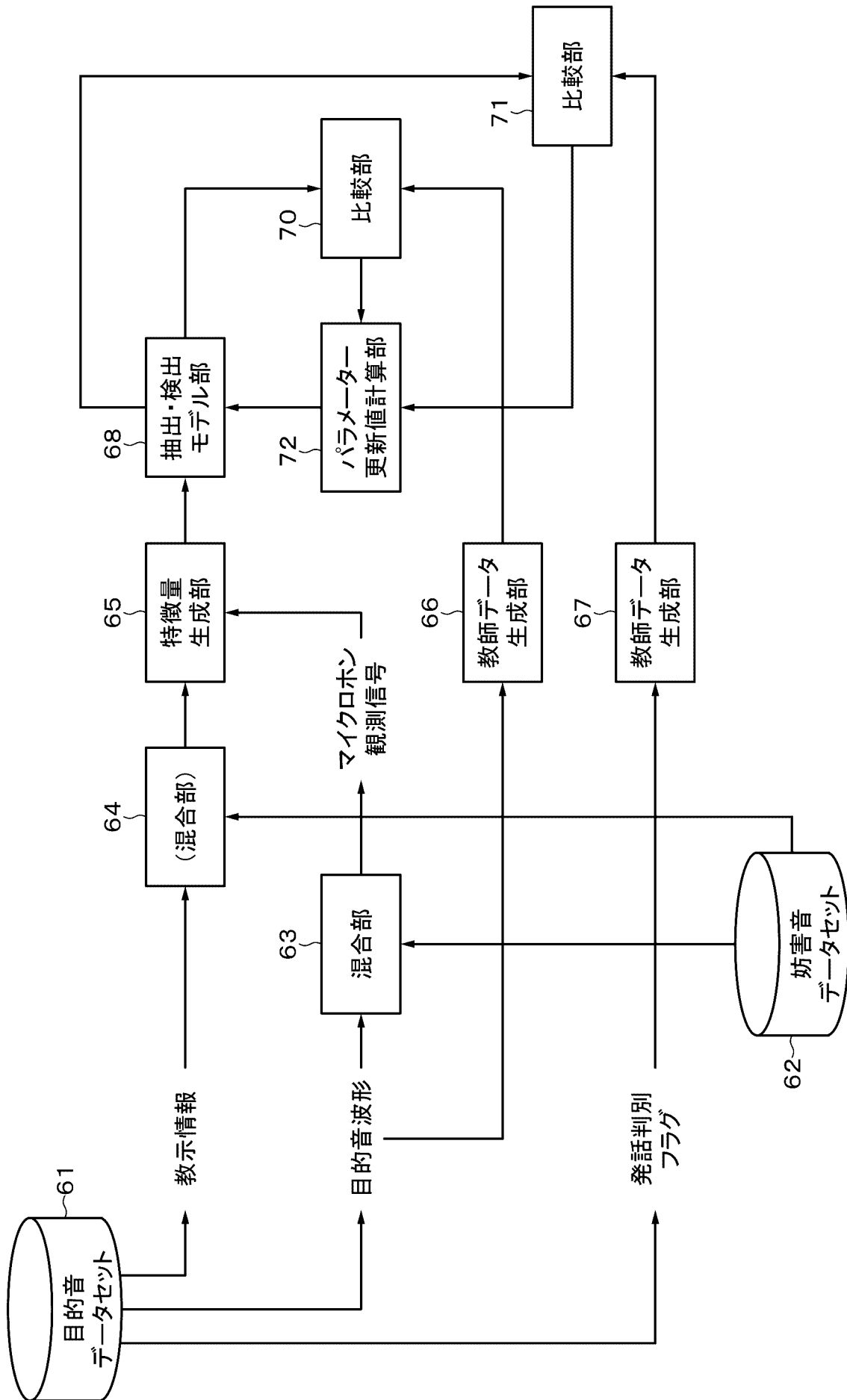
[図19]



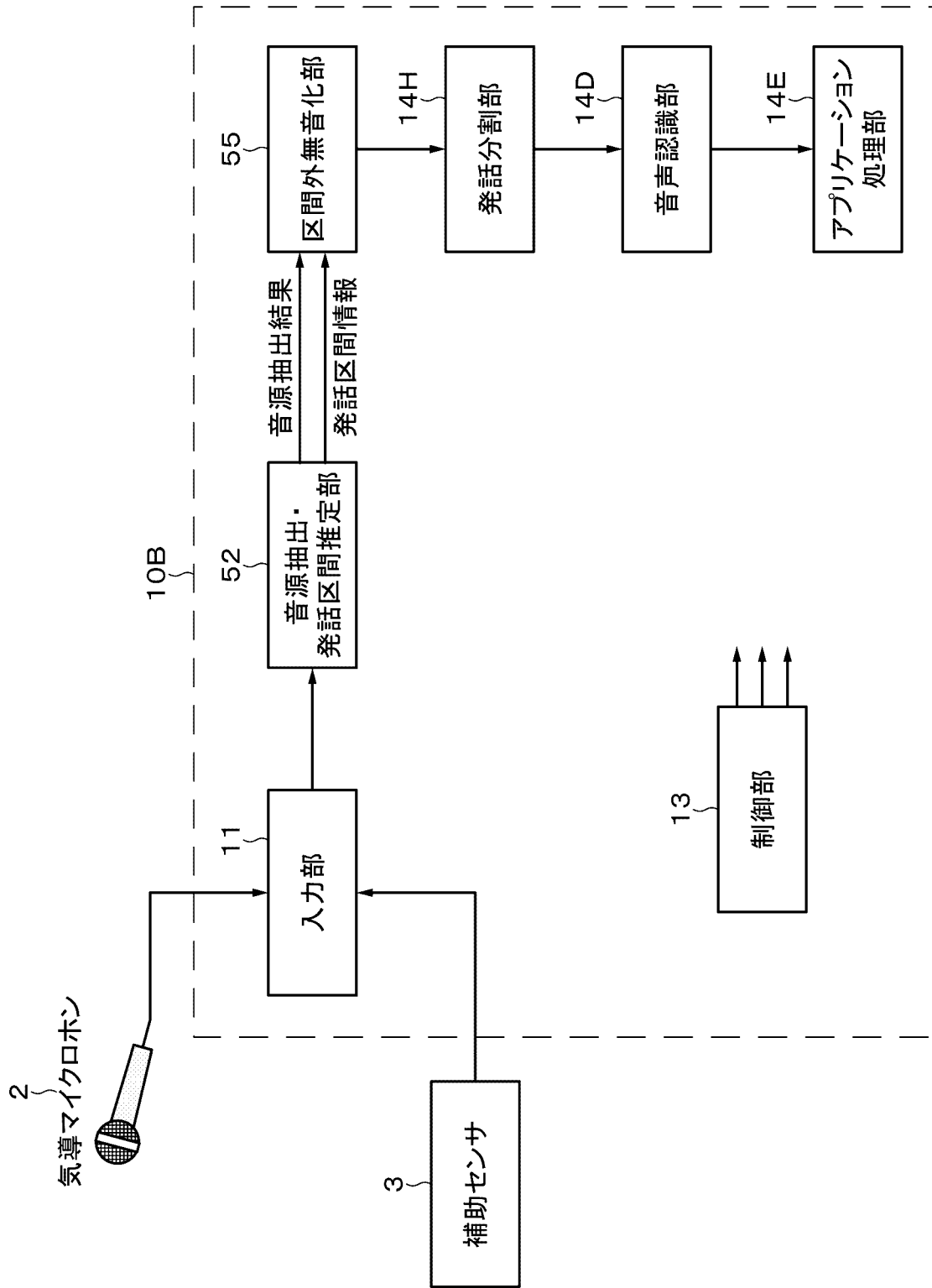
[図20]



[図21]



[図22]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/005061

A. CLASSIFICATION OF SUBJECT MATTER Int.Cl. H04R1/10(2006.01)i, H04R3/00(2006.01)i, G10L21/0272(2013.01)i, G10L15/20(2006.01)i, G10L25/84(2013.01)i FI: G10L21/0272100A, G10L25/84, G10L15/20370E, H04R1/10101A, H04R3/00320 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. H04R1/10, H04R3/00, G10L21/0272, G10L15/20, G10L25/84 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Published examined utility model applications of Japan 1922-1996 Published unexamined utility model applications of Japan 1971-2020 Registered utility model specifications of Japan 1996-2020 Published registered utility model applications of Japan 1994-2020 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.		
X	JP 2007-251354 A (NATIONAL UNIVERSITY CORPORATION SAITAMA UNIVERSITY) 27.09.2007 (2007-09-27), paragraphs [0001], [0002], [0007], [0016]-[0028], [0047], [0086], fig. 1-4	1-4, 6, 9-10, 18-19		
Y	paragraphs [0002], [0021], [0028], [0086]	5, 7, 8, 11-17		
Y	WO 2016/002358 A1 (SONY CORPORATION) 07.01.2016 (2016-01-07), paragraph [0016], fig. 1	5		
Y	JP 6-12483 A (CANON INC.) 21.01.1994 (1994-01-21), paragraphs [0027]-[0032], fig. 7	7-8		
Y	JP 4-276799A (RICOH CO., LTD.) 01.10.1992 (1992-10-01), paragraphs [0003], [0005]-[0008], fig. 1-4	11-13		
Y	JP 11-224098 A (MEIDENSHA CORPORATION) 17.08.1999 (1999-08-17), paragraphs [0011], [0017], fig. 1	13		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.				
<table style="width:100%; border:none;"> <tr> <td style="width:50%; border:none;"> * 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 </td> <td style="width:50%; border:none;"> "I" 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 </td> </tr> </table>			* 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	"I" 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
* 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	"I" 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 15.04.2020	Date of mailing of the international search report 12.05.2020			
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.			

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/005061

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2019/017403 A1 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 24.01.2019 (2019-01-24), paragraphs [0005], [0057]-[0068], fig. 3	14, 16-17
Y	JP 2016-180839 A (JAPAN BROADCASTING CORPORATION) 13.10.2016 (2016-10-13), paragraphs [0024]-[0027], fig. 2	15

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2020/005061

JP 2007-251354 A	27.09.2007	(Family: none)
WO 2016/002358 A1	07.01.2016	US 2017/0200442 A1 paragraph [0056], fig. 1 EP 3163902 A CN 106664473 A
JP 6-12483 A	21.01.1994	(Family: none)
JP 4-276799 A	01.10.1992	(Family: none)
JP 11-224098 A	17.08.1999	(Family: none)
WO 2019/017403 A1	24.01.2019	(Family: none)
JP 2016-180839 A	13.10.2016	(Family: none)

<p>A. 発明の属する分野の分類（国際特許分類（IPC）） H04R 1/10(2006.01)i; H04R 3/00(2006.01)i; G10L 21/0272(2013.01)i; G10L 15/20(2006.01)i; G10L 25/84(2013.01)i FI: G10L21/0272 100A; G10L25/84; G10L15/20 370E; H04R1/10 101A; H04R3/00 320</p>										
<p>B. 調査を行った分野</p>										
<p>調査を行った最小限資料（国際特許分類（IPC）） H04R1/10; H04R3/00; G10L21/0272; G10L15/20; G10L25/84</p>										
<p>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</p> <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2020年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2020年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2020年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2020年	日本国実用新案登録公報	1996 - 2020年	日本国登録実用新案公報	1994 - 2020年
日本国実用新案公報	1922 - 1996年									
日本国公開実用新案公報	1971 - 2020年									
日本国実用新案登録公報	1996 - 2020年									
日本国登録実用新案公報	1994 - 2020年									
<p>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）</p>										
<p>C. 関連すると認められる文献</p>										
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								
X	JP 2007-251354 A (国立大学法人埼玉大学) 27.09.2007 (2007 - 09 - 27) [0001] - [0002]、[0007]、[0016] - [0028]、 [0047]、[0086]、第1-4図	1-4, 6, 9-10, 18-19								
Y	[0002]、[0021]、[0028]、[0086]	5, 7, 8, 11-17								
Y	WO 2016/002358 A1 (ソニー株式会社) 07.01.2016 (2016 - 01 - 07) [0016]、第1図	5								
Y	JP 6-12483 A (キヤノン株式会社) 21.01.1994 (1994 - 01 - 21) [0027] - [0032]、第7図	7 - 8								
Y	JP 4-276799 A (株式会社リコー) 01.10.1992 (1992 - 10 - 01) [0003]、[0005] - [0008]、第1-4図	11 - 13								
Y	JP 11-224098 A (株式会社明電舎) 17.08.1999 (1999 - 08 - 17) [0011]、[0017]、第1図	13								
Y	WO 2019/017403 A1 (日本電信電話株式会社) 24.01.2019 (2019 - 01 - 24) [0005]、[0057] - [0068]、第3図	14, 16 - 17								
<p><input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> パテントファミリーに関する別紙を参照。</p>										
* 引用文献のカテゴリー	<p>“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抵触するものではなく、発明の原理又は理論の理解のために引用するもの</p>									
“A” 特に関連のある文献ではなく、一般的な技術水準を示すもの	<p>“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</p>									
“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	<p>“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</p>									
“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）	<p>“&” 同一パテントファミリー文献</p>									
“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										
国際調査を完了した日	国際調査報告の発送日									
15.04.2020	12.05.2020									
名称及びあて先	権限のある職員（特許庁審査官）									
日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	大野 弘 5Z 9175 電話番号 03-3581-1101 内線 3591									

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
Y	JP 2016-180839 A (日本放送協会) 13.10.2016 (2016 - 10 - 13) [0024] - [0027]、第2図	15

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2020/005061

引用文献	公表日	パテントファミリー文献	公表日
JP 2007-251354 A	27.09.2007	(ファミリーなし)	
WO 2016/002358 A1	07.01.2016	US 2017/0200442 A1 [0056]、第1図 EP 3163902 A CN 106664473 A	
JP 6-12483 A	21.01.1994	(ファミリーなし)	
JP 4-276799 A	01.10.1992	(ファミリーなし)	
JP 11-224098 A	17.08.1999	(ファミリーなし)	
WO 2019/017403 A1	24.01.2019	(ファミリーなし)	
JP 2016-180839 A	13.10.2016	(ファミリーなし)	