

(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
2022年12月1日(01.12.2022)



(10) 国際公開番号

WO 2022/249302 A1

(51) 国際特許分類:
G10L 15/20 (2006.01)

(21) 国際出願番号: PCT/JP2021/019874

(22) 国際出願日: 2021年5月25日(25.05.2021)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 佐藤 宏 (SATO, Hiroshi); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 落合 翼 (OCHIAI,

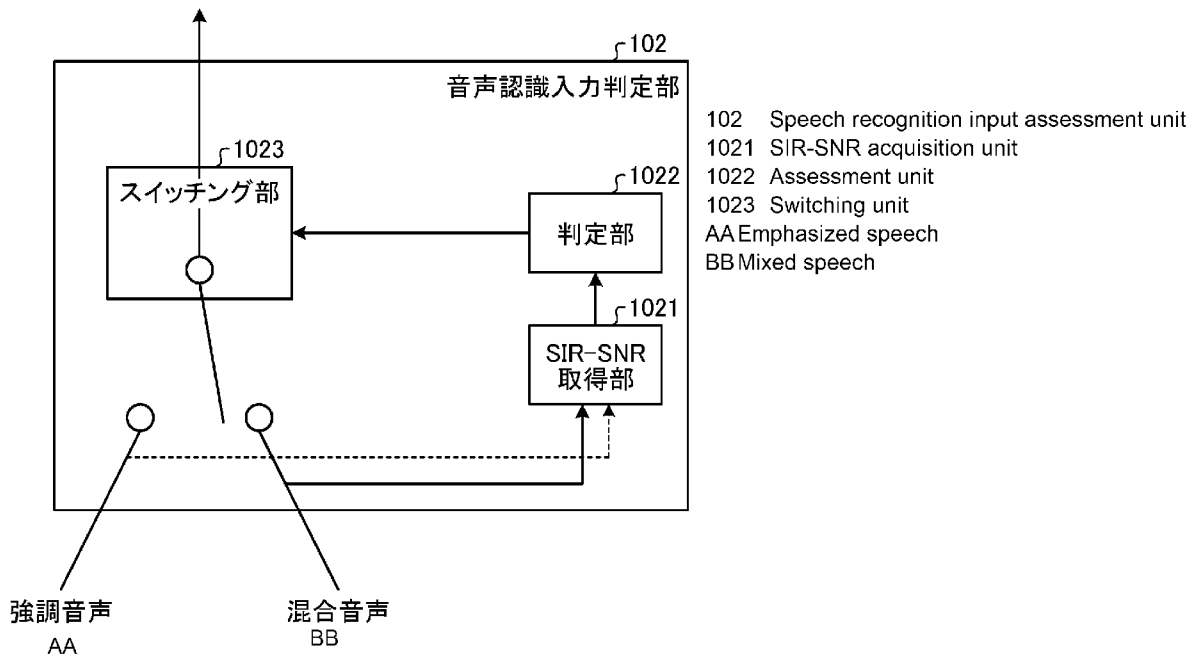
Tsubasa); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). デルクロア マーク (DELCROIX, Marc); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP). 木下 慶介 (KINOSHITA, Keisuke); 〒1808585 東京都武蔵野市緑町3丁目9-11 NTT 知的財産センタ内 Tokyo (JP).

(74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ,

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

(54) 発明の名称: 信号処理装置、信号処理方法及び信号処理プログラム



(57) Abstract: This speech recognition input assessment unit (102) has: an SIR-SNR acquisition unit (1021) that, from mixed speech in which the speech of another speaker overlaps the speech of a target speaker, acquires a signal-to-interference ratio (SIR), which is the ratio of the target speech and the interfering speaker's speech in the mixed speech, and a signal-to-noise ratio (SNR), which is the ratio of the target speech and noise, and/or acquires the mixed speech; and an assessment unit (1022) that, on the basis of the SIR and SNR, and/or on the basis of the mixed speech, assesses that

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, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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))

speech based on at least one of the mixed speech and emphasized speech, in which the mixed speech is emphasized, is the speech to be used in speech recognition.

(57) 要約: 音声認識入力判定部 (102) は、目的話者の音声に別の話者の音声重複している混合音声から、混合音声における、目的音声と干渉話者音声との比率である S I R (Signal to Interference Ratio) 及び混合音声の、目的音声と雑音との比率である S N R (Signal to Noise Ratio) と、混合音声とのうち少なくとも一つを取得する S I R - S N R 取得部 (1021) と、S I R 及び S N R と、前記混合音声とのうち少なくとも一つを基に、混合音声と、混合音声を強調した強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する判定部 (1022) と、を有する。

明 細 書

発明の名称：

信号処理装置、信号処理方法及び信号処理プログラム

技術分野

[0001] 本発明は、信号処理装置、信号処理方法及び信号処理プログラムに関する。

背景技術

[0002] 深層学習技術の発達により音声認識精度は向上し、音声認識の適用範囲は徐々に拡大している。しかしながら、現在でも複数人の声が重複して発せられた場合それを認識するのは困難である。

[0003] そこで、音声の重複に対処するための種々の技術が考案されている。ブラインド音源分離技術はそのうちの一つのアプローチであり、混合音に含まれる各音源に由来する音に分離を行うものである。ブラインド音源分離技術は、混合音のままでは音声認識が困難な音声を、各話者の音声に分離することで音声認識を可能にすることができる（例えば、非特許文献 1 参照）。

[0004] また、目的話者抽出技術が提案されている。目的話者抽出技術は、目的話者が事前登録した発話を補助的な情報として利用し、事前登録された話者の音声のみを、混合音から取得する技術である。この目的話者抽出技術は、混合音に含まれる話者数に関する事前情報を必要としない利点があり、実用上有用な技術である（例えば、非特許文献 2 参照）。目的話者抽出技術を用いて抽出した音声は目的話者の声だけを含むことから、音声認識が可能である。

先行技術文献

非特許文献

[0005] 非特許文献 1 : Dong Yu, et al, “PERMUTATION INVARIANT TRAINING OF DEEP MODELS FOR SPEAKER-INDEPENDENT MULTI-TALKER SPEECH SEPARATION”, 2017 IEEE International Conference on Acoustics, Sp

eech and Signal Processing (ICASSP), IEEE, 2017.

非特許文献2: Zmolikova Katerina, et al, "SpeakerBeam: Speaker Aware Neural Network for Target Speaker Extraction in Speech Mixtures", IEEE Journal of Selected Topics in Signal Processing 13.4 (2019): 800-814.

非特許文献3: Quan Wang, et al, "VoiceFilter-Lite: Streaming Targeted Voice Separation for On-Device Speech Recognition", arXiv preprint arXiv:2009.04323 (2020).

発明の概要

発明が解決しようとする課題

- [0006] 前述のとおり音声強調技術は、目的話者の音声が他の話者の音声と重複している場合や、背景雑音と重複している場合に、それら望ましくない音を除去し、目的話者の音声のみを取り出すことを可能とする技術である。しかしながら、望ましくない音を除去する際に、同時に所望の目的話者音声を歪ませてしまう場合がある。
- [0007] 音声強調技術を適用した音声に対して、音声認識を行う場合、この目的話者音声の歪みが、音声認識性能の劣化を引き起こすおそれがある。このような音声認識性能の劣化は、他話者の音声が必ずしも重複しない現実の音声に対して目的話者抽出を行う実運用上大きな課題である。
- [0008] 一方、音声認識は、ある程度の雑音であれば小さな性能劣化で動作することから、雑音の程度や種類によっては、むしろ目的話者抽出を行わない方が、音声認識性能が良い場合もある。すなわち、音声強調技術を適用することによる、雑音除去による効果と、それに伴うひずみの悪影響とを比較して、前者の方が大きい場合雑音除去を行うことが望ましく、後者の方が大きい場合音声強調を行わないことが望ましい。
- [0009] 例えば、非特許文献3では、雑音が他話者の音声である場合（音声雑音）に音声強調は効果的である一方、音声ではない場合（非音声雑音）に音声強調が音声認識を劣化させる可能性が高いという知見に基づき、雑音の種類が

他の話者の音声か、その他の雑音かを識別する二値識別機を導入し、雑音の種類が他の話者の音声であった場合のみ、音声強調を行う手法を提案している。すなわち、目的話者の音声に別の話者の音声が存在しない場合は音声強調をせず、重複した場合にのみ音声強調を行うことを提案している。

[0010] しかしながら、複数話者の音声重複して発せられた場合であっても、干渉話者音声の存在による性能劣化を、それを除去することによって生じる歪みによって生じる性能劣化が上回れば音声強調が劣化する場合がある。

[0011] 本発明は、上記に鑑みてなされたものであって、目的話者の音声に別の話者の音声重複している場合に、音声認識の精度を高めることができる入力音声として、混合音声と、混合音声を強調した強調音声との少なくとも一方を適切に判定することができる信号処理装置、信号処理方法及び信号処理プログラムを提供することを目的とする。

課題を解決するための手段

[0012] 上述した課題を解決し、目的を達成するために、本発明に係る信号処理装置は、目的話者の音声に別の話者の音声重複している混合音声から、混合音声における、目的音声と干渉話者音声との比率である SIR (Signal to Interference Ratio)、及び、混合音声の、目的音声と雑音との比率である SNR (Signal to Noise Ratio) と、混合音声とのうちの少なくとも一方を取得する取得部と、 SIR 及び SNR と、混合音声とのうち少なくとも一つを基に、混合音声と、混合音声を強調した強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する判定部と、を有することを特徴とする。

発明の効果

[0013] 本発明によれば、目的話者の音声に別の話者の音声重複している場合に、音声認識の精度を高めることができる入力音声として、混合音声と、混合音声を強調した強調音声との少なくとも一方を適切に判定することができる。

図面の簡単な説明

[0014] [図1]図1は、実施の形態に係る信号処理装置の構成の一例を模式的に示す図である。

[図2]図2は、図1に示す音声認識入力判定部の構成の一例を模式的に示す図である。

[図3]図3は、混合音声を強調した強調音声に対する音声認識性能を示す図である。

[図4]図4は、混合音声に対する音声認識性能を示す図である。

[図5]図5は、実施の形態に係る信号処理の処理手順を示すフローチャートである。

[図6]図6は、図5に示す音声認識入力判定処理の処理手順を示すフローチャートである。

[図7]図7は、実施の形態に係る信号処理装置を適用した場合の音声認識結果を示す図である。

[図8]図8は、強調音声を用いた場合の音声認識結果に対する、実施の形態に係る信号処理装置を用いた音声認識結果の性能向上率を示す図である。

[図9]図9は、プログラムが実行されることにより、信号処理装置が実現されるコンピュータの一例を示す図である。

発明を実施するための形態

[0015] 以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施形態により本発明が限定されるものではない。また、図面の記載において、同一部分には同一の符号を付して示している。

[0016] [実施の形態]

目的話者の音声に別の話者の音声重複した場合であっても、音声強調を行わない方がよい場合がありうる。例えば、干渉話者の音声が目話者の音声と比較して振幅が小さい場合、音声強調せずとも音声認識がある程度目的話者の音声を認識することが可能である場合がある。また、干渉話者の音声の他に、非音声雑音が存在する場合、音声強調による歪みの影響は大きくなる傾向が確認されており、このような場合については音声強調の悪影響が大

きく、もともとの混合音声認識したほうが音声認識にとって有利であることがある。

[0017] そこで、実施の形態に係る信号処理装置では、目的話者の音声に別の話者の音声重複している混合音声を対象とし、音声認識の精度を高めることができる入力音声として、混合音声と、混合音声を強調した強調音声との少なくとも一方に基づく音声を判定する。実施の形態に係る信号処理装置では、混合音声における、目的音声と雑音との比率である S N R (Signal to Noise Ratio)、及び、混合音声の、目的音声と干渉話者音声との比率である S I R (Signal to Interference Ratio) と、混合音声との少なくとも一つを基に、混合音声と強調音声との少なくとも一方に基づく音声を判定する。実施の形態に係る信号処理装置では、判定した混合音声と強調音声との少なくとも一方に基づく音声を用いて、音声認識処理を行うことで、音声認識精度の向上を図ることができる。

[0018] [信号処理装置]

次に、実施の形態に係る信号処理装置について説明する。図 1 は、実施の形態に係る信号処理装置の構成の一例を模式的に示す図である。

[0019] 信号処理装置 100 は、例えば、ROM (Read Only Memory)、RAM (Random Access Memory)、CPU (Central Processing Unit) 等を含むコンピュータ等に所定のプログラムが読み込まれて、CPU が所定のプログラムを実行することで実現される。また、信号処理装置 100 は、有線接続、或いは、ネットワーク等を介して接続された他の装置との間で、各種情報を送受信する通信インタフェースを有する。信号処理装置 100 は、音声強調部 101、音声認識入力判定部 102 及び音声認識部 103 を有する。

[0020] 音声強調部 101 は、公知の音声強調技術を用いて、入力される混合音声から、所望の音声を抽出する。音声強調部 101 の処理として、公知の音源分離技術や、公知の目的話者抽出技術を適用することができる。音源分離技術では、入力された混合音声を各話者に分離するため、各話者の音声に音声認識入力判定部 102 に与えられる。目的話者抽出技術では、入力と

して、目的話者に関する補助情報も併せて与えられる必要がある。目的話者に関する補助情報として、例えば目的話者が事前に登録した発話が用いられる。なお、目的話者に関する補助情報は、信号処理装置１００が保持してもよく、信号処理装置１００と通信可能である外部装置から入力されるものであってもよい。

[0021] 音声強調部１０１は、混合音声を強調した強調音声抽出する。例えば混合音声として、目的話者との干渉話者の音声の混合音声の混合音声強調部１０１に入力された場合、音声強調部１０１は、混合音声から、事前登録発話を発した目的話者の音声を抽出する。音声強調部１０１は、抽出した音声を強調音声として音声認識入力判定部１０２のＳＩＲ－ＳＮＲ取得部１０２１（後述）に入力する。入力される混合音声および強調音声としては、音声に対して特徴抽出などの何らかの変換を行ったものでもよいし、音声波形であってもよい。

[0022] 音声認識入力判定部１０２は、混合音声のＳＮＲ及び混合音声のＳＩＲと、混合音声とのうち少なくとも一つを基に、混合音声と、混合音声を強調した強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する。音声認識入力判定部１０２は、判定した混合音声と強調音声との少なくとも一方に基づく音声を音声認識部１０３に入力する。音声認識入力判定部１０２は、音声認識に入力する発話を、音声強調部１０１の出力である強調音声と、信号処理装置１００に入力された混合音声との少なくとも一方に基づく音声を、発話単位ないしはフレームなど一定区間単位で判定する。

[0023] 具体的に、音声認識入力判定部１０２は、音声の判定をハードに行う。例えば、音声認識入力判定部１０２は、「１」、「０」で、音声強調部１０１の出力である強調音声と、信号処理装置１００に入力された混合音声とのいずれか一方のみを音声認識部１０３に入力する。

[0024] また、音声認識入力判定部１０２は、音声の判定をソフトに行うこともできる。例えば、音声認識入力判定部１０２は、音声強調部１０１の出力であ

る強調音声と、信号処理装置 100 に入力された混合音声との一方に例えば「0.3」の重みづけを行い、他方に「0.7」の重みづけを行い、重みづけ後の混合音声と強調音声とを音声認識部 103 に入力する。

[0025] 音声認識部 103 は、音声認識入力判定部 102 が判定した、混合音声と強調音声との少なくとも一方に基づく音声を用いて、音声認識処理を行う。音声認識部 103 の音声認識処理には、公知の音声認識技術を適用することができる。

[0026] [SIR、SNR]

ここで、SIR の演算式を式 (1) に示し、SNR の演算式を式 (2) に示す。式 (1)、式 (2) において、S は、目的話者の音声であり、I は、干渉話者の音声であり、N は、雑音である。

[0027] [数1]

$$SIR = 10 \log_{10} \frac{\|S\|^2}{\|I\|^2} \quad \dots (1)$$

[0028] [数2]

$$SNR = 10 \log_{10} \frac{\|S\|^2}{\|N\|^2} \quad \dots (2)$$

[0029] 図 3 は、混合音声を強調した強調音声に対する音声認識性能を示す図である。図 4 は、混合音声に対する音声認識性能を示す図である。図 3 及び図 4 のいずれも、音声認識性能として、文字誤り率 (Character Error Rate) を評価した。

[0030] 図 3 及び図 4 に示すように、音声認識では、入力として、混合音声を強調した強調音声と、もともとの混合音声のいずれかが有利かは、SIR 及び SNR に依存することが実験的に判明している。具体的には、SIR から SNR を減じた値が大きいほど、強調音声よりも混合音声を入力とした方が、文字誤り率が低い。

[0031] このため、音声認識入力判定部 102 は、SIR 及び SNR を入力音声から推定することで、図 3 において音声認識の精度を高めることができるよう

なS N R及びS I Rの領域に対しては強調音声を選択し、そうでない領域については混合音声を選択して、音声認識部103に入力する。或いは、音声認識入力判定部102は、図3において音声認識の精度を高めることができるようなS N R及びS I Rの領域に対しては強調音声に、混合音声よりも大きな重みづけを行い、そうでない領域については混合音声に、強調音声よりも大きな重みづけを行い、重みづけ後の混合音声と強調音声とを音声認識部103に入力する。

[0032] [音声認識入力判定部]

次に、音声認識入力判定部102について説明する。図2は、図1に示す音声認識入力判定部102の構成の一例を模式的に示す図である。

[0033] 図2に示すように、音声認識入力判定部102は、S I R－S N R取得部1021（取得部）、判定部1022及びスイッチング部1023を有する。

[0034] S I R－S N R取得部1021は、システムに入力された混合音声から、混合音声におけるS I R、及び、混合音声におけるS N Rと、混合音声とのうち少なくとも一つを取得する。S I R－S N R取得部1021は、S I R及びS I Rを、フレームなど一定時間単位で取得してもよいし、発話単位で取得してもよい。

[0035] S I R－S N R取得部1021は、例えば、S I R及びS N Rを推定する推定モデルを用いて、S I R及びS N Rを推定する。推定モデルは、ニューラルネットワークなどによって構成される。推定モデルは、音声とS I R及びS N Rとのペアデータを予め学習させたものである。

[0036] また、音声強調部101によって、目的話者の音声と、それ以外の干渉話者の音声とが抽出された場合、S I R－S N R取得部1021は、抽出された目的話者の音声と、それ以外の干渉話者の音声を基に、式（1）及び式（2）を用いて、S I R及びS I Rを計算してもよい。

[0037] 判定部1022は、S I R－S N R取得部1021によって取得されたS I R及びS N Rと、混合音声とのうち少なくとも一つを基に、混合音声と、

混合音声を強調した強調音声との少なくとも一方に基づく音声の音声認識に使用する音声として判定する。判定部1022は、SIR-SNR取得部1021によって推定或いは計算されたSIR及びSNRと、混合音声とのうち少なくとも一つを基に、混合音声と、混合音声を強調した強調音声とのどちらが音声認識にとって有利かを判定する。

[0038] 例えば、判定部1022は、SIR及びSNRを用いた所定のルールを用いて、混合音声と強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する。判定部1022は、音声の判定をハードに行う場合、例えば、式(3)の場合に強調音声を選択するというルールを用いる。

[0039] [数3]

$$SIR - SNR \leq 10 \quad \dots (3)$$

[0040] また、判定部1022は、音声認識の入力として、混合音声と強調音声とのいずれかが有利かであるかがSIR及びSNRに依存すること(図3, 4参照)を基に、SIRとSNRとの組み合わせに応じて、混合音声に対する重み値と強調音声に対する重み値を予め設定したルールを用いて判定を行ってもよい。

[0041] 判定部1022は、音声の判定をソフトに行う場合、音声認識部103の入力として、混合音声と強調音声との少なくとも一方に基づく音声を識別する識別モデルを用いて、混合音声と強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定してもよい。

[0042] 識別モデルは、ニューラルネットワークなどから構成される。識別モデルは、混合音声と強調音声とのいずれが音声認識に使用する音声として有利であるかを示す教師ラベルが付与された、混合音声から取得された特徴量と、SNR及びSIRと、の少なくとも一方を入力として予め学習させたものである。

[0043] 識別モデルがSNR及びSIRを入力とする場合は、混合音声と強調音声とのいずれが音声認識に使用する音声として有利であるかを示す教師ラベル

と、推定あるいは計算されたS N R及びS I Rのペアデータとをもとに、識別モデルの学習を行う。

[0044] 識別モデルは、混合音声から取得された特徴量を入力として、混合音声と強調音声とのいずれが音声認識にとって有利であるかを直接推定することも可能である。その場合は混合音声と強調音声とのいずれが音声認識に使用する音声として有利であることを示す教師ラベルと、混合音声から抽出された特徴量のペアデータとをもとに、識別モデルの学習を行う。

[0045] 識別モデルがS N R及びS I Rと混合音声の両方を入力する場合は、それらと混合音声と強調音声とのいずれが音声認識に使用する音声として有利であることを示す教師ラベルのペアデータを基に識別モデルの学習を行う。

[0046] 識別モデルは、混合音声及び強調音声に対する、音声認識に使用する音声としての有利さを示す重みを出力する。また、識別モデルは、音声認識に使用する音声として有利である音声として、混合音声または強調音声の一方を選択した判定結果を出力するものであってもよい。

[0047] スイッチング部1023は、判定部1022の判定結果をもとに、混合音声と強調音声との少なくとも一方に基づく音声を音声認識部103へ入力する。図4のように音声の判定をハードに行う場合、スイッチング部1023は、判定部1022の選択結果をもとに、強調音声または混合音声を音声認識部103へ入力する。音声の判定をソフトに行う場合（不図示）、スイッチング部1023は、判定部1022が出力した重みに応じて混合音声と強調音声とに重みづけを行い、重みづけ後の混合音声と強調音声とを音声認識部103に出力する。

[0048] [信号処理の処理手順]

次に、信号処理装置100が実行する信号処理について説明する。図5は、実施の形態に係る信号処理の処理手順を示すフローチャートである。

[0049] 混合音声の入力を受け付けると（ステップS1）、音声強調部101は、入力された混合音声から、目的話者の音声を抽出し、抽出した音声を、混合音声を強調した強調音声として音声認識入力判定部102に入力する音声強

調処理を行う（ステップS 2）。

[0050] 音声認識入力判定部102は、混合音声のSNR及び混合音声のSIRと、混合音声とのうち少なくとも一つを基に、混合音声と、混合音声を強調した強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する音声認識入力判定処理を行う（ステップS 3）。

[0051] 音声認識部103は、音声認識入力判定部102が判定した、混合音声と強調音声との少なくとも一方に基づく音声を用いて、音声認識処理を行う音声認識処理を行い（ステップS 4）、音声認識結果を出力する（ステップS 5）。

[0052] [音声認識入力判定処理の処理手順]

次に、図5の音声認識入力判定処理（ステップS 3）について説明する。図6は、図5に示す音声認識入力判定処理の処理手順を示すフローチャートである。

[0053] 図6に示すように、SIR-SNR取得部1021は、システムに入力された混合音声から、混合音声におけるSIR及び混合音声におけるSNRと、混合音声とのうち少なくとも一つを取得する取得処理を行う（ステップS 11）。

[0054] 判定部1022は、SIR-SNR取得部1021によって取得されたSIR及びSNRと、混合音声とのうち少なくとも一つを基に、混合音声と、混合音声を強調した強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する判定処理を行う（ステップS 12）。スイッチング部1023は、判定部1022の判定結果をもとに、混合音声と強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として音声認識部103へ入力する（ステップS 13）。

[0055] [実施の形態の効果]

このように、信号処理装置100では、音声認識の入力として、混合音声と強調音声とのいずれかが有利かであるかがSIR及びSNRに依存することに基づいて、SIR及びSNRと、混合音声とのうち少なくとも一つを基

に、混合音声と強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する。

[0056] 図7は、実施の形態に係る信号処理装置100を適用した場合の音声認識結果の一例を示す図である。図8は、強調音声を用いた場合の音声認識結果に対する、実施の形態に係る信号処理装置100を用いた音声認識結果の性能向上率の一例を示す図である。図7及び図8は、強調音声及び混合音声のうち、式(3)の場合に強調音声を選択するというルールを用いて、音声認識部103へ入力する音声を選択した場合の音声認識結果である。なお、図7及び図8のいずれも、音声認識性能として、文字誤り率(Character Error Rate)を評価した。

[0057] 図7に示す枠W1、図8に示す枠W11の領域は、推定したSIR、SNRに基づいて、入力信号が混合音声に選択されていると思われる領域である。図7に示す枠W1、図8に示す枠W11の音声認識結果より、式(3)の場合に強調音声を選択するというルールを用いて音声認識部103への信号を選択する信号処理装置100の方が、強調音声のみを用いた場合と比して、音声認識精度が高いことが分かった。

[0058] SIR、SNRの組み合わせ全体において、平均すると、信号処理装置100の方が、強調音声のみを用いた場合と比して、音声認識精度が高いことが分かった。さらに、SIRの値が比較的大きい領域(例えば、枠W11、W12参照)において、信号処理装置100の方が、強調音声のみを用いた場合と比して、音声認識精度が特に高いことが明らかになった。

[0059] これより、信号処理装置100によれば、目的話者の音声に別の話者の音声重複している場合であっても、音声認識の精度を高めることができる入力音声として、混合音声と、混合音声を強調した強調音声との少なくとも一方に基づく音声を適切に判定する。そして、信号処理装置100は、音声認識の精度を高めることができる入力音声を適切に判定することで、音声強調の歪みによる性能劣化を防ぎ、音声認識精度を高めることができる。

[0060] なお、本実施の形態では、音声強調部101は、音声認識の精度を高める

ことができる入力音声として混合音声を選択された場合、音声強調処理を行わなくともよい。このように、音声強調部 101 は、音声強調処理が必要な区間、すなわち、音声認識入力判定部 102 によって、音声認識の精度を高めることができる入力音声として強調音声を選択された区間についてのみ音声強調処理を行う。

[0061] [実施の形態のシステム構成について]

信号処理装置 100 の各構成要素は機能概念的なものであり、必ずしも物理的に図示のように構成されていることを要しない。すなわち、信号処理装置 100 の機能の分散及び統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散または統合して構成することができる。

[0062] また、信号処理装置 100 においておこなわれる各処理は、全部または任意の一部が、CPU、GPU (Graphics Processing Unit)、及び、CPU、GPU により解析実行されるプログラムにて実現されてもよい。また、信号処理装置 100 においておこなわれる各処理は、ワイヤードロジックによるハードウェアとして実現されてもよい。

[0063] また、実施の形態において説明した各処理のうち、自動的におこなわれるものとして説明した処理の全部または一部を手動的に行うこともできる。もしくは、手動的におこなわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。この他、上述及び図示の処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて適宜変更することができる。

[0064] [プログラム]

図 9 は、プログラムが実行されることにより、信号処理装置 100 が実現されるコンピュータの一例を示す図である。コンピュータ 1000 は、例えば、メモリ 1010、CPU 1020 を有する。また、コンピュータ 1000 は、ハードディスクドライブインタフェース 1030、ディスクドライブインタフェース 1040、シリアルポートインタフェース 1050、ビデオ

アダプタ１０６０、ネットワークインタフェース１０７０を有する。これらの各部は、バス１０８０によって接続される。

[0065] メモリ１０１０は、ROM１０１１及びRAM１０１２を含む。ROM１０１１は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライブインタフェース１０３０は、ハードディスクドライブ１０９０に接続される。ディスクドライブインタフェース１０４０は、ディスクドライブ１１００に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ１１００に挿入される。シリアルポートインタフェース１０５０は、例えばマウス１１１０、キーボード１１２０に接続される。ビデオアダプタ１０６０は、例えばディスプレイ１１３０に接続される。

[0066] ハードディスクドライブ１０９０は、例えば、OS (Operating System) １０９１、アプリケーションプログラム１０９２、プログラムモジュール１０９３、プログラムデータ１０９４を記憶する。すなわち、信号処理装置１００の各処理を規定するプログラムは、コンピュータ１０００により実行可能なコードが記述されたプログラムモジュール１０９３として実装される。プログラムモジュール１０９３は、例えばハードディスクドライブ１０９０に記憶される。例えば、信号処理装置１００における機能構成と同様の処理を実行するためのプログラムモジュール１０９３が、ハードディスクドライブ１０９０に記憶される。なお、ハードディスクドライブ１０９０は、SSD (Solid State Drive) により代替されてもよい。

[0067] また、上述した実施の形態の処理で用いられる設定データは、プログラムデータ１０９４として、例えばメモリ１０１０やハードディスクドライブ１０９０に記憶される。そして、CPU１０２０が、メモリ１０１０やハードディスクドライブ１０９０に記憶されたプログラムモジュール１０９３やプログラムデータ１０９４を必要に応じてRAM１０１２に読み出して実行する。

[0068] なお、プログラムモジュール１０９３やプログラムデータ１０９４は、ハ

ードディスクドライブ１０９０に記憶される場合に限らず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ１１００等を介してＣＰＵ１０２０によって読み出されてもよい。あるいは、プログラムモジュール１０９３及びプログラムデータ１０９４は、ネットワーク（ＬＡＮ（Local Area Network）、ＷＡＮ（Wide Area Network）等）を介して接続された他のコンピュータに記憶されてもよい。そして、プログラムモジュール１０９３及びプログラムデータ１０９４は、他のコンピュータから、ネットワークインタフェース１０７０を介してＣＰＵ１０２０によって読み出されてもよい。

[0069] 以上、本発明者によってなされた発明を適用した実施の形態について説明したが、本実施の形態による本発明の開示の一部をなす記述及び図面により本発明は限定されることはない。すなわち、本実施の形態に基づいて当業者等によりなされる他の実施の形態、実施例及び運用技術等は全て本発明の範疇に含まれる。

符号の説明

- [0070] １００ 信号処理装置
 １０１ 音声強調部
 １０２ 音声認識入力判定部
 １０３ 音声認識部
 １０２１ ＳＩＲ－ＳＮＲ取得部
 １０２２ 判定部
 １０２３ スイッチング部

請求の範囲

- [請求項1] 目的話者の音声に別の話者の音声が重複している混合音声から、前記混合音声における、目的音声と干渉話者音声との比率である SIR (Signal to Interference Ratio)、及び、前記混合音声の、目的音声と雑音との比率である SNR (Signal to Noise Ratio) と、前記混合音声とのうち少なくとも一つを取得する取得部と、
- 前記 SIR 及び前記 SNR と、前記混合音声とのうち少なくとも一つを基に、前記混合音声と、前記混合音声を強調した強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する判定部と、
- を有することを特徴とする信号処理装置。
- [請求項2] 前記判定部は、前記 SIR 及び前記 SNR を用いた所定のルールを用いて、前記混合音声と前記強調音声との少なくとも一方に基づく音声を、前記音声認識に使用する音声として判定することを特徴とする請求項1に記載の信号処理装置。
- [請求項3] 前記判定部は、前記混合音声と前記強調音声とのいずれが音声認識に使用する音声として有利であるかを示す教師ラベルが付与された、前記混合音声から取得された特徴量と、前記 SNR 及び前記 SIR と、の少なくとも一方を入力として学習された識別モデルを用いて、前記混合音声と前記強調音声との少なくとも一方に基づく音声を、前記音声認識に使用する音声として判定することを特徴とする請求項1に記載の信号処理装置。
- [請求項4] 前記判定部が判定した前記混合音声と前記強調音声との少なくとも一方に基づく音声を用いて、音声認識処理を行う音声認識部をさらに有することを特徴とする請求項1～3のいずれか一つに記載の信号処理装置。
- [請求項5] 前記混合音声から、前記目的話者の音声を抽出し、抽出した音声を前記強調音声として前記取得部に入力する音声強調部をさらに有する

ことを特徴とする請求項 1 ～ 4 のいずれか一つに記載の信号処理装置。

[請求項6]

信号処理装置が実行する信号処理方法であって、

目的話者の音声に別の話者の音声重複している混合音声から、前記混合音声における、目的音声と干渉話者音声との比率である S I R (Signal to Interference Ratio)、及び、前記混合音声の、目的音声と雑音との比率である S N R (Signal to Noise Ratio) と、前記混合音声とのうち少なくとも一つを取得する工程と、

前記 S I R 及び前記 S N R と、前記混合音声とのうち少なくとも一つを基に、前記混合音声と、前記混合音声を強調した強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定する工程と、

を含んだことを特徴とする信号処理方法。

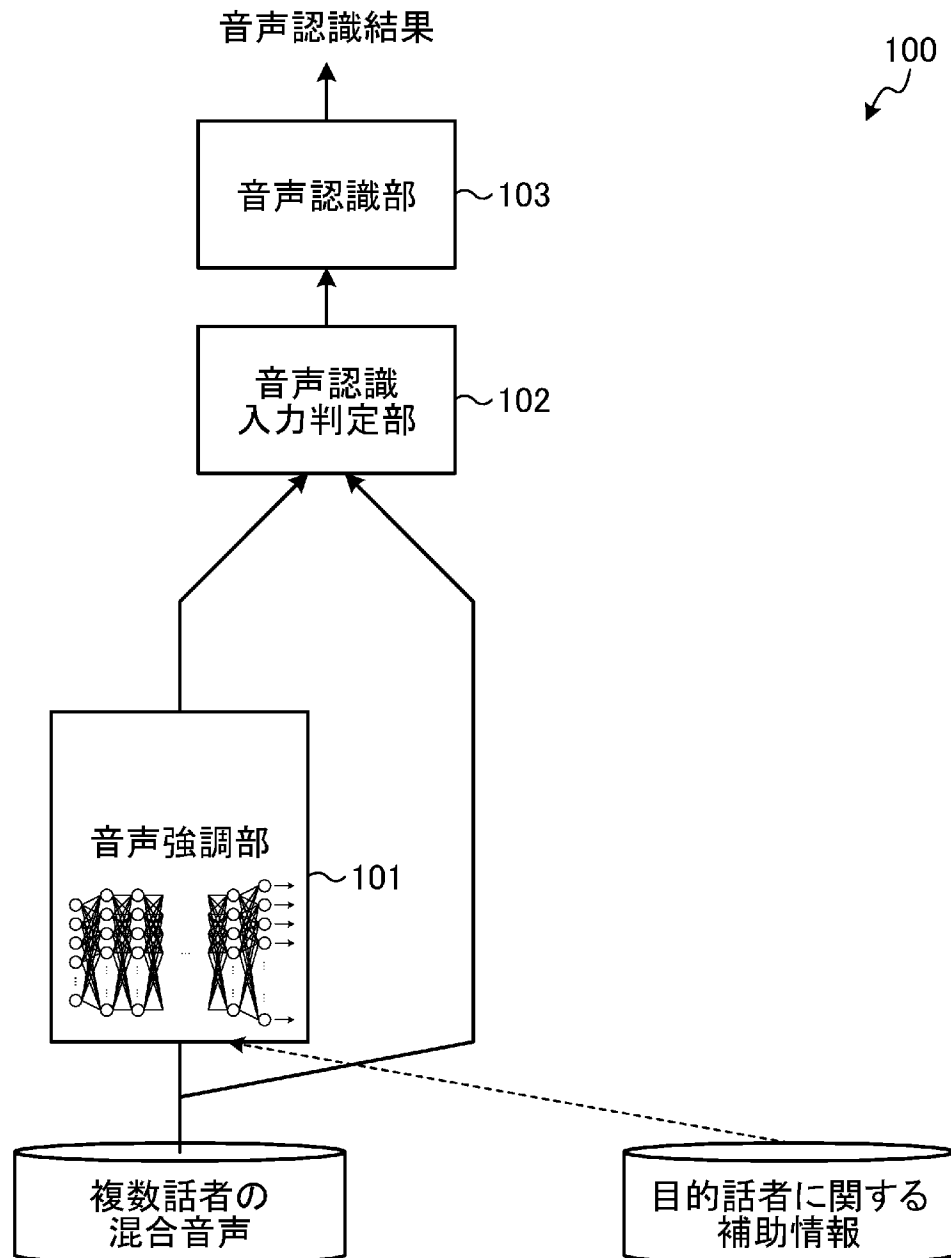
[請求項7]

目的話者の音声に別の話者の音声重複している混合音声から、前記混合音声における、目的音声と干渉話者音声との比率である S I R (Signal to Interference Ratio)、及び、前記混合音声の、目的音声と雑音との比率である S N R (Signal to Noise Ratio) と、前記混合音声とのうち少なくとも一つを取得するステップと、

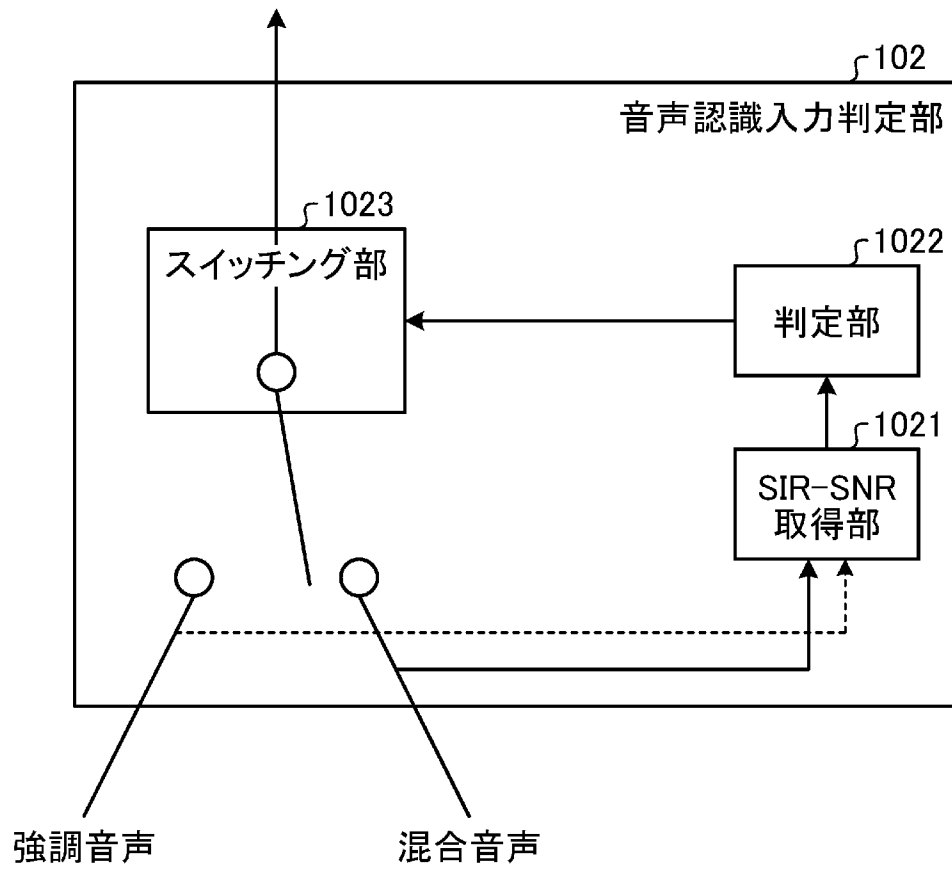
前記 S I R 及び前記 S N R と、前記混合音声とのうち少なくとも一つを基に、前記混合音声と、前記混合音声を強調した強調音声との少なくとも一方に基づく音声を、音声認識に使用する音声として判定するステップと、

をコンピュータに実行させるための信号処理プログラム。

[図1]



[図2]



[図3]

SNR SIR		SNR		
		20	10	0
	0	8.4	10.5	30.5
	5	6.5	8.7	26.9
	10	5.6	7.7	23.8
	15	5.2	7.5	22.7
	20	5.9	8.4	22.4
	25	9.2	9.5	22.4

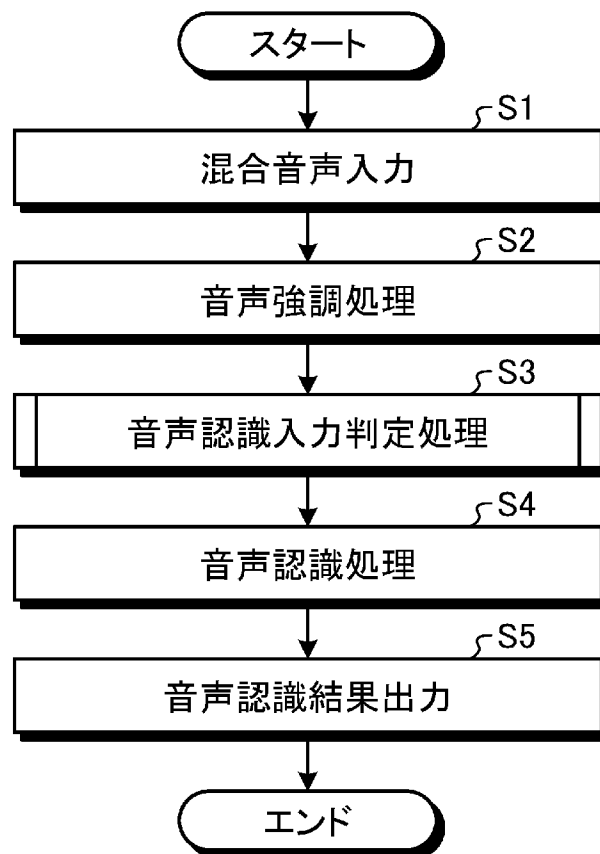
[%]

[図4]

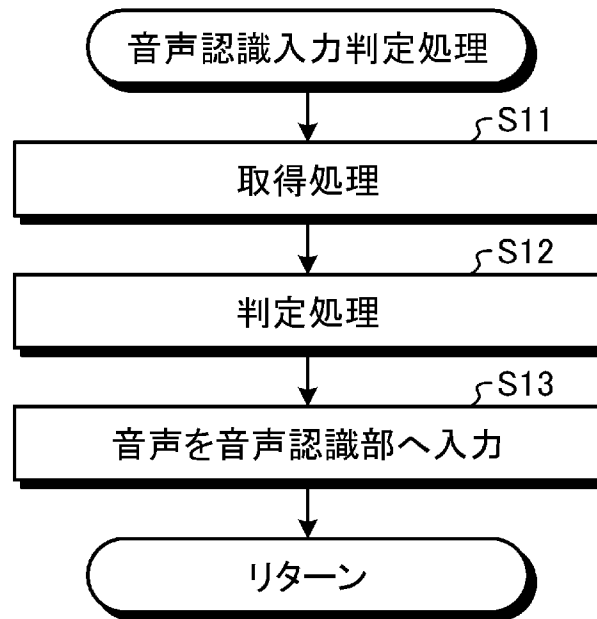
SNR SIR		SNR		
		20	10	0
	0	80.5	80.1	82.6
	5	40.5	36.5	39.2
	10	16.1	13.2	19.3
	15	7.6	6.9	14.7
	20	5.0	5.6	13.5
	25	4.4	5.4	13.2

[%]

[図5]



[図6]



[図7]

SNR				
SIR				
		20	10	0
	0	8.4	10.5	30.5
	5	6.6	8.9	27.8
	10	6.5	8.4	19.9
	15	5.9	6.9	16.0
	20	5.4	6.8	15.5
	25	7.6	8.2	15.6

[%]

W1

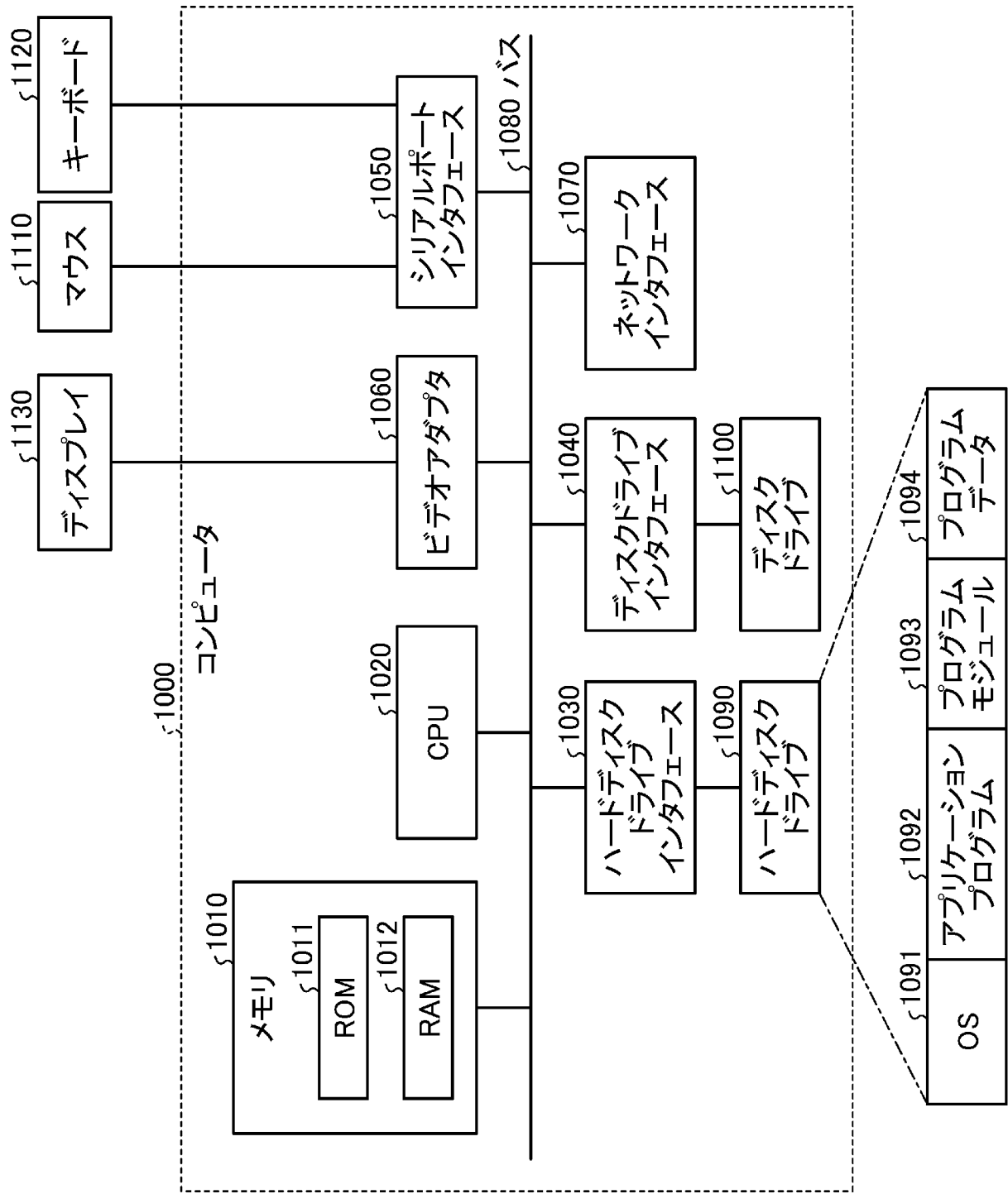
[図8]

SNR				
SIR				
		20	10	0
	0	0%	0%	0%
	5	-2%	-2%	-3%
	10	-16%	-9%	16%
	15	-13%	8%	30%
	20	8%	19%	31%
	25	17%	14%	30%

[%]

W12 W11

[図9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/019874

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/20(2006.01)i

FI: G10L15/20 353

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)

G10L15/00-15/34

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 1971-2021

Registered utility model specifications of Japan 1996-2021

Published registered utility model applications of Japan 1994-2021

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

IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2018-205512 A (ARTSIDE KK) 27 December 2018 (2018-12-27) paragraphs [0011]-[0126]	1-7
A	JP 2002-236497 A (ALPINE ELECTRONICS INC) 23 August 2002 (2002-08-23) paragraphs [0011]-[0018]	1-7
A	JP 2007-264327 A (MATSUSHITA ELECTRIC WORKS LTD) 11 October 2007 (2007-10-11) paragraphs [0032]-[0056]	1-7

☐ 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
20 July 2021 (20.07.2021)

Date of mailing of the international search report
10 August 2021 (10.08.2021)

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

Information on patent family members

International application no.

PCT/JP2021/019874

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2018-205512 A	27 Dec. 2018	(Family: none)	
JP 2002-236497 A	23 Aug. 2002	(Family: none)	
JP 2007-264327 A	11 Oct. 2007	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G10L 15/20(2006.01)i FI: G10L15/20 353		
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）） G10L15/00-15/34		
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年 日本国公開実用新案公報 1971-2021年 日本国実用新案登録公報 1996-2021年 日本国登録実用新案公報 1994-2021年		
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語） IEEE Xplore		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2018-205512 A（富士通株式会社）27.12.2018（2018-12-27） [0011]-[0126]	1-7
A	JP 2002-236497 A（アルパイン株式会社）23.08.2002（2002-08-23） [0011]-[0018]	1-7
A	JP 2007-264327 A（松下電工株式会社）11.10.2007（2007-10-11） [0032]-[0056]	1-7
☐ 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、一般的な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 “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 20.07.2021		国際調査報告の発送日 10.08.2021
名称及びあて先 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		権限のある職員（特許庁審査官） 菊池 智紀 5Z 3352 電話番号 03-3581-1101 内線 3591

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/JP2021/019874

引用文献			公表日	パテントファミリー文献	公表日
JP	2018-205512	A	27.12.2018	(ファミリーなし)	
JP	2002-236497	A	23.08.2002	(ファミリーなし)	
JP	2007-264327	A	11.10.2007	(ファミリーなし)	