

平成21年6月10日現在

研究種目：若手研究（A）
 研究期間：2006～2008
 課題番号：18680017
 研究課題名（和文） 変調伝達関数に基づいたブラインド残響音声回復法に関する研究
 研究課題名（英文） Study on the blind speech dereverberation method based on the Concept of Modulation Transfer Function
 研究代表者
 鶴木 祐史（UNOKI MASASHI）
 北陸先端科学技術大学院大学・情報科学研究科・准教授
 研究者番号：00343187

研究成果の概要：本研究では、音声明瞭度に関係づけられた変調伝達関数の概念に基づき、室内伝達特性を測定することなく、ブラインドで残響音を回復する方法を提案した。提案法は、適切な時間・周波数領域におけるパワーエンベロープ逆フィルタ処理とキャリア再生成処理で構成された。人工・実残響環境において大規模評価実験を行った結果、提案法が残響によって著しく低下した音声明瞭度ならびに音声認識率を効果的に改善できることを明らかにした。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	8,300,000	2,490,000	10,790,000
2007年度	5,400,000	1,620,000	7,020,000
2008年度	1,900,000	570,000	2,470,000
年度			
年度			
総計	15,600,000	4,680,000	20,280,000

研究分野：総合領域

科研費の分科・細目：情報学 知覚情報処理・知能ロボティクス

キーワード：変調伝達関数，残響音声，残響除去，音声明瞭度，逆フィルタ，パワーエンベロープ，キャリア，基本周波数推定

1. 研究開始当初の背景

残響は、音声信号の重要な情報を複雑に歪ませ、人間対機械、人間対人間の音声コミュニケーションを困難にさせる。そのため、残響の影響を受けた音声から元の音声へ物理的・知覚的に回復する研究は、音声認識や拡声会議通話システム、補聴システムといった様々な音声信号処理で非常に重要な課題である。特に、この要素技術を利用したバリアフリー・コミュニケーション技術は、福祉社

会への多大な貢献を期待できるものである。

これまでに提案されてきた数多くの残響除去法では、室内音響の伝達特性を事前に測定し、その逆特性を利用することを基本コンセプトとして残響音声の回復処理を実現してきた。このコンセプトを基とした方法では、伝達系の特性が最小位相特性を有していれば、逆フィルタを構成できるため、容易に残響除去を行うことができる（例えば、Neely & Allen (1972)の研究）。しかし、一般に伝達系は最小位相特性でない場合が多く、逆フィ

ルタを構成することができない。そのため、如何に適切な逆フィルタを導出するかということに焦点が当てられ、様々な方法が提案されてきた(代表例として、Miyoshi & Kaneda (1988) [MINT 法]や Wang & Itakura (1992))。しかし、実際の室内伝達特性は、時々刻々、環境の変化(例えば、室内の物の配置換えや人の移動、室温・湿度の変化など)とともに変動するため、残響除去の度に室内伝達特性を測定し、適応的に処理しなければならない。そのため、これらの方法は、応用面を考えた場合、現実的な処理体系であるとはいえない。

一方、これらとは異なるコンセプトとして、変調伝達関数の概念に基づくものがある。これは、Houtgast & Steeneken (1973) によって提案された音声明瞭度予測理論に直接関係するものである。このオリジナルの発想は、音声信号の時間的なエンベロープ(振幅の時間方向への変化を表す包絡線情報)が残響の影響により時間的に遅延しながら減衰する(変調伝達関数: MTF) ため、これを MTF の逆特性により回復を試みるというものである。これは、Langhaus & Strube (1982) の試みに続き、Mourjopoulos & Hammand (1983), Avendano & Hermansky (1996), 広林ら(1998, 2000) のモデルによって検討され、残響音声回復の可能性を示唆するに至った。

MTF は、伝達系の入出力間のパワーエンベロープの関係から得られ、事前に伝達系を測定する必要はないが、ブラインド処理として MTF の逆特性を利用するためには、MTF のパラメータをブラインド的に決定する必要がある。また、エンベロープ処理のみに着目しているため、回復されたエンベロープと位相情報を含む元のキャリアをそのまま利用して復元すると、波形レベルでの回復が十分に行われず、音声明瞭度そのものも顕著に回復されない。これらの問題点から、MTF に基づくアプローチでは、応用範囲が制限され(パワーエンベロープの回復のみ)、従来の方法には及ばなかった。

研究代表者は、MTF ベースの回復手法(特に、広林らの方法)を、波形レベルでの回復を可能とするモデルに拡張することができれば、モデルパラメータと音声認識等で利用される特徴との対応づけ(物理的回復の意味づけ)や、モデルパラメータと聴覚印象(音声明瞭度や残響感など)との対応づけ(知覚的回復の意味づけ)を踏まえて議論することが可能になり、その応用範囲を大幅に広げることができると考えた。更に、この手法が音声明瞭度に関して、ブラインド的に残響音声を回復できる可能性があることから、バリ

アフリー・コミュニケーションの要素技術として非常に重要で基礎的な位置にあると考えた。特に、音声明瞭度の改善効果に関しては、エンベロープとキャリアの両方の回復なしに、残響音声の正確な回復や著しい音声明瞭度の改善を達成できないと考えた。

MTF に基づくブラインド残響音声回復法を実現するための基礎検討として、まず彼らのモデルでの問題点を洗い出し、それらの解決法を考案した。次に、これをもとに検討を深め、ブラインド回復法のためのパラメータ決定法、ならびに音声を対象とした帯域分割型振幅変調モデルで表現される残響音声回復の実現方法を検討した。これらの基礎検討から、MTF に基づいたブラインド残響音声回復法の基本処理方式の実現可能性を見出し、計算シミュレーション上でその有効性を示すことができた。

2. 研究の目的

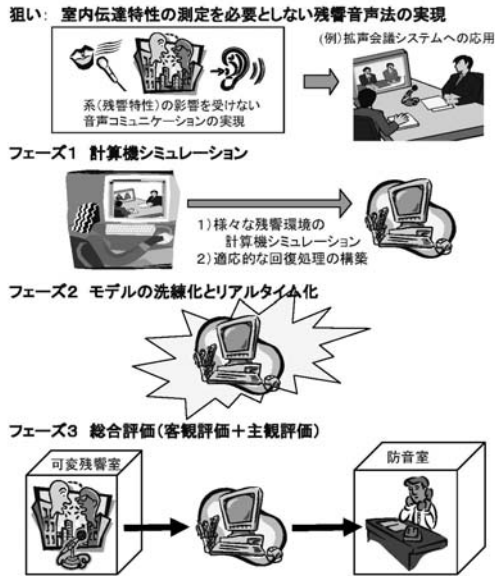
本研究の着想に至るまでに行った基礎検討すべてを踏まえると、ブラインド残響音声回復法を実現するためには、先の基本処理方式(案)において、(1) 適応的な残響回復処理の実現、(2) 残響回復と明瞭度改善の効果の対応づけ、(3) 実環境での評価と耐性、(4) 実時間処理の実現、という課題をクリアしなければならない。

本研究では、MTF の概念に基づき室内伝達特性を測定することなく、ブラインド的にかつ適応的に残響音声を回復する手法(基本処理方式)を提案する。また、現実問題への対策を検討するために、実時間処理への応用を狙ったプロトタイプモデルの構築を目指す。これにあたり、上記四つの課題(1)~(4)を解決し、本手法の有効性を示す。

3. 研究の方法

本研究の準備段階で、既に、変調伝達関数の概念に基づいたブラインド的な残響音声回復のための基本処理方式(案)を確立した。しかし、この方法はまだ試作段階にあり、完全なブラインド処理でかつ適応的に残響音声回復を可能とする方法を確立するためには、先に述べたように、(1) 適応的な残響回復処理の実現、(2) 残響回復と明瞭度改善の効果の対応づけ、(3) 実環境での評価と耐性ならびにモデルの洗練化、(4) 実時間処理の実現と総合評価という課題を実施する必要がある。そこで、これらを3段階のフェーズに分け、研究を遂行した。

- フェーズ1：計算機シミュレーション
課題(1)と(2)の実施
- フェーズ2：モデルの洗練化
課題(3)の実施
- フェーズ3：モデルの総合評価
課題(4)の実施



4. 研究成果

(1) フェーズ1

現段階の基本処理方式(案)では、事前に定められた周波数分割幅で、残響音声を時間・周波数表現し、各サブバンド信号上でエンベロープとキャリアに分解した。その後、エンベロープについてはパワーエンベロープ逆フィルタ処理(系のMTFの逆特性を補償する処理)を行うことで残響成分を取り除き、クリーンな音声のエンベロープに回復させた。キャリアについては、原音声の基本周波数とパワー成分から有声・無声(UV)判定を行い、チャンネルボコーダ型の音源信号を駆動することで、キャリアを再生成した。これらを再び組み合わせることで、サブバンド信号を復元し、残響回復音声を再合成した。この基本処理方式(案)については、次のような検討を行い、それらの処理を実現した。

①MTF ベースの信号モデル化

基本処理方式(案)は、振幅変調モデルに基づいて信号を表現している。MTFの概念に基づくと、これには、エンベロープ逆フィルタ法とパワーエンベロープ逆フィルタ法の二つが存在する。計算機シミュレーションにより、二つの手法の有効性を検討したところ、人工信号・音声信号ともにパワーエンベロープ逆フィルタ処理が有効であることがわか

った。同様に、キャリア表現にも検討したところ、雑音キャリアと複合音キャリアとも、いずれも相違ない結果となることがわかった。このことから、信号表現は振幅変調モデルで定義されるが、回復処理はパワーエンベロープ上で行うこととした。

②理想 MTF と合理的な逆 MTF フィルタ

残響環境の場合、MTFは一種の低域通過特性となる。そのため、パワーエンベロープ逆フィルタ(逆MTFフィルタ)は高域強調処理となり、エンベロープ上の高周波成分のゴミを過剰に強調する恐れがある。本研究では、高域成分を過剰に強調し過ぎない合理的な逆MTFフィルタを設計した。

③キャリア再生成処理

基本処理方式(案)では、原音声の基本周波数を既知としてキャリア再生成処理を行っていた。これを完全にブラインド処理化するために、既存の基本周波数推定法とパワー推定、UV判定法を組み合わせ、音源駆動モデルを作成し、キャリア再生成処理を実現した。

④適応的な時間・周波数分解について

大規模シミュレーションを利用して帯域分割幅について検討した結果、工学的によく利用される定Q分割ではなく、対象信号の共変調特性(音声特有の特徴を保持するため)を満たし、可能な限り広い帯域幅(サブバンド信号のキャリア間の無相関性を満たすため)で、適応的に周波数分割する必要があることを明らかにした。また、その一次近似として、平均値を定帯域幅分割として活用できることもわかった。同様に、(サブバンド)音声信号の無音区間に対応して、適応的に時間分割することも有効であることがわかった。そこで、これらの特性を可能な限り満たすために、対象信号に特化して、適応的に時間分割・周波数分割を可能とするフィルタバンクを実現した。

(2) フェーズ2

以上の検討を踏まえ、基本処理方式を確立し、様々な残響特性を考慮した大規模シミュレーションを行うことで、提案モデルの洗練化を行った。その結果、次の点について再検討し、モデルの改良を行った。

①室内残響インパルス応答の仮定

本研究では、統計的な検討から、指数減衰型のパワーエンベロープをもつ残響インパルス応答(Schroeder, 1981)を仮定して、逆MTFを算出した。特殊なケースを除いて、ほぼこの仮定で拡散音場における残響インパルス応答を模擬できるが、初期反射をもつ

指数減衰型のパワーエンベロープには対応できないことがわかった。そこで、新たに後者のパワーエンベロープに対して、MTF 特性を算出し、逆フィルタを設計した。

②残響時間推定の改良

パワーエンベロープ逆フィルタ処理をブラインドで行うためには、残響時間に関係するパラメータ値を知る必要である。既存の推定法では、残響時間が長い場合に過小推定する傾向にあるため、この推定法を改良した。これまでのパラメータ推定では、実測の残響時間と完全に一致しなかったが、改良法におけるパラメータ推定では、ほぼ実測の値と一致することから、残響時間のブラインド推定法も同時に確立することができた。

③残響環境下での F0 推定法とキャリア再生成法の改良

キャリア再生成処理で駆動される音源信号は、特に補聴を考えたときに、回復音声の音質に影響を与えることがわかった。同時に残響時間が長くなるにつれ、駆動音源の特性が劣化する傾向にあった。これは UV 判定だけでなく、基本周波数 F0 自体の推定が失敗していることに起因していた。そのため、残響音声から正確に F0 推定する方法を開発した。これに基づき、残響音声からの F0 推定だけでなく、パワーエンベロープ逆フィルタから推定された回復音声のパワー推定も同時に利用して、正確な UV 判定をすることでキャリア再生成法を改良した。

(3) フェーズ 3

以上の改良を踏まえて、MTF に基づいたブラインド残響音声回復法を確立した。ここでは、提案法の処理方式により残響音声の回復度合い（明瞭度の回復、残響感の低減）を知るために、聴取実験ならびに大規模な計算機シミュレーションによる評価実験（PESQ, LSD, WSS, LPC 距離, LLR, 明瞭度重み付 LSD など）を利用した評価を行った。この結果、提案手法により、残響によって低下した音声の明瞭度を改善できることを確認した。特に、単語理解度試験を行った結果、親密度の低い単語ほど残響の影響を強く受け、その認識率も著しく低下していた。これに対し、提案法は約 30%認識率を改善することができた。

次に提案法を音声認識の前処理として利用したときの効果を調べた。AURORA2J で公開された音声データベースを利用して、人工的な残響環境ならびに実残響環境において認識精度を調べたところ、残響時間が 0.2 s 内では従来法でも認識率 90%程度を保持できたが、残響時間が 1s を超えると、60%、2s で 40%まで低下した。これに対し、提案法では

全体的に認識率を改善し、残響時間が 1s 以上で約 10%の改善を生み出していた。

以上により、提案手法の総合評価から、提案法が補聴処理としてだけでなく音声認識等にも有効に機能していることを確認できた。本提案法で利用されるパワーエンベロープ逆フィルタ処理は、一次の IIR フィルタで実現されているため、既にリアルタイム処理化が可能であるが、時間・周波数分解で利用するフィルタバンクが分析再合成のため FIR 型で実現されており、リアルタイム処理には若干の問題が残っている。また F0 推定を含むキャリア再生成処理は、現在のところまだオフライン処理でしか実現できていないため、当面の課題として残った。

しかしながら、本研究により、残響環境の室内インパルス応答を測定せずに、残響音声をブラインドで回復できる処理体系を確立できたという成果は大きい。特に、本研究の特色である、伝達系の影響（時変な要因）を受けない処理体系であることと、変調伝達関数に基づいた処理であるため、音声明瞭度回復に直結した信号処理を実現できたことは、今後の聴覚的信号処理としての応用・発展に期待できる。

本研究は、残響環境に特化した音声明瞭度を回復する方法であったが、実環境では雑音も混在した複雑な環境である。変調伝達関数の概念は、雑音に対しても同様に適用できることから、今後は、雑音・残響環境におけるブラインド音声回復法の実現を目指したい。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 22 件）

1. 鶴木祐史, “変調伝達関数に基づく音声信号処理(3)―残響環境下の基本周波数推定と残響時間ブラインド推定―,”信号処理学会論文誌, Vol. 13, No. 2, pp. 91-101, March 2009, 解説論文(査読有).
2. Yutaka Yamasaki and Masashi Unoki, “A study on the noise suppression method based on the MTF concept,” Proc. 2009 RISP International workshop on Nonlinear Circuits and Signal Processing (NCSP09), pp. 157-160, March 2009, 査読有.
3. 鶴木祐史, “変調伝達関数に基づく音声信号処理(2)―ブラインド残響音声回復法―,”信号処理学会論文誌, Vol. 13, No. 1, pp. 3-12, Jan. 2009, 解説論文(査読有).
4. Xugang Lu, Masashi Unoki, and Masato Akagi, “A comparative evaluation of MTF-based blind restoration of the sub-band power envelopes of speech as a

- front-end processor for automatic speech recognition systems,” *Acoustical Science & Technology*, Vol. 29, No. 6, pp. 351–361, Nov. 2008, 査読有.
5. 鶴木祐史, “変調伝達関数に基づく音声信号処理(1)ーパワーエンベロップ逆フィルタ処理の原理とその応用についてー,” *信号処理学会論文誌*, Vol. 12, No. 5, pp. 339–348, Sept. 2008, 解説論文(査読有).
 6. Rico Petrick, Xugang Lu, Masashi Unoki, Masato Akagi, and Ruediger Hoffmann, “Robust Front End Processing for Speech Recognition in Reverberant Environments: Utilization of Speech Characteristics,” *Proc. Interspeech2008*, pp. 658–661, Sept. 2008, 査読有.
 7. Rico Petrick, Masashi Unoki, Anish Mittal, Calros Segura, and Ruediger Hoffmann, “A Comprehensive Study on the Effects of Room Reverberation on Fundamental Frequency Estimation,” *Proc. Interspeech2008*, pp. 131–134, Sept. 2008, 査読有.
 8. Masashi Unoki and Sota Hiramatsu, “MTF-based method of blind estimation of reverberation time in room acoustics,” *Proc. EUSIPCO2008*, August 2008, CD-ROM, 査読有.
 9. Sota Hiramatsu and Masashi Unoki, “A Study on the Blind Estimation of Reverberation Time in Room Acoustics,” *Journal of Signal Processing*, Vol. 12, No. 4, pp. 323–326, July 2008, 査読有.
 10. Xugang Lu, Masashi Unoki, and Masato Akagi, “An MTF-based blind restoration of temporal power envelopes as a front-end processor for automatic speech recognition systems in reverberant environments,” *Proc. Acoustics 08 Paris*, pp. 1419–1424, July 2008, 査読無.
 11. Masashi Unoki and Sota Hiramatsu, “Blind estimation method of reverberation time based on concept of modulation transfer function,” *Proc. Acoustics 08 Paris*, pp. 4489–4494, July 2008, 査読無.
 12. Masashi Unoki, Toshihiro Hosorogiya, and Yuichi Ishimoto, “Comparative evaluations of robust and accurate F0 estimates in reverberant environments,” *Proc. 2008 International Conference on Acoustics, Speech, and Signal Processing (ICASSP2008)*, pp. 4569–4572, March 2008, 査読有.
 13. Masashi Unoki and Toshihiro Hosorogiya, “Estimation of fundamental frequency of reverberant speech by utilizing complex cepstrum analysis,” *Journal of Signal Processing*, vol. 12, no. 1, pp. 31–44, Jan. 2008, 査読有.
 14. Xugang Lu, Masashi Unoki, and Masato Akagi, “Comparative evaluation of MTF-based feature extraction for speech recognition in reverberant environments,” *Proc. XII International Conference Speech and Computer (SPECOM2007)*, pp. 124–133, Oct. 2007, 査読無.
 15. Masashi Unoki and Toshihiro Hosorogiya, “A study on robust and accurate F0 estimation in reverberant environments by utilizing complex cepstrum analysis,” *Proc. XII International Conference Speech and Computer (SPECOM2007)*, pp. 222–232, Oct. 2007, 査読無.
 16. Xugang Lu and Masashi Unoki, “A study on restoration methods for temporal envelope of reverberant speech based on the modulation transfer function concept,” *Proc. the Japan-China Joint Conference on Acoustics*, P-2-11, June 2007, 査読無.
 17. Masashi Unoki and Toshihiro Hosorogiya, “Estimation of fundamental frequency of reverberant speech by utilizing complex cepstrum analysis,” *JAIST Research Report, IS-RR-2007-008*, June 2007, 査読無.
 18. Xugang Lu and Masashi Unoki, “Dereverberation models for temporal envelope restoration based on the modulation transfer function concept,” *Proc. 2007 RISP International workshop on Nonlinear Circuits and Signal Processing (NCSP07)*, pp. 394–400, March 2007, 査読有.
 19. Toshihiro Hosorogiya and Masashi Unoki, “A study on the estimation method of the fundamental frequency from a reverberant speech,” *Proc. 2007 RISP International workshop on Nonlinear Circuits and Signal Processing (NCSP07)*, pp. 575–578, March 2007, 査読有.
 20. Yohei Shibano and Masashi Unoki, “An improved carrier regeneration process in the MTF-based speech dereverberation method using instantaneous amplitude and phase,” *Proc. 2007 RISP International workshop on Nonlinear Circuits and Signal Processing (NCSP07)*, pp. 221–224, March

- 2007, 査読有.
21. Xugang Lu, Masashi Unoki, and Masato Akagi, "A robust feature extraction based on the MTF concept for speech recognition in reverberant environment," Proc. Interspeech 2006, pp. 2546-2549, Sept. 2006, 査読有.
 22. Masashi Unoki, Masato Toi, and Masato Akagi, "Refinement of an MTF-based speech dereverberation method using optimal inverse-MTF filter," Proc. XI International Conference Speech and Computer (SPECOM2006), pp. 323-326, June 2006, 査読無.

[学会発表] (計 10 件)

1. Rico Petrick, Xugang Lu, Masashi Unoki, Masato Akagi, and Ruediger Hoffmann, "Robust Front End Processing for Speech Recognition in Reverberant Environments: Utilization of Speech Properties," IEICE Tech. Rep. SP2008-44, pp. 7-12, 岩手県立大アイーナキャンパス, July 18, 2008.
2. Masashi Unoki, Rico Petrick, Anish Mittal, and Ruediger Hoffmann, "Effects of Room Reverberation on Robust and Accurate F0 Estimates," IEICE Tech. Rep. SP2008-43, pp. 1-6, 岩手県立大学アイーナキャンパス, July 18, 2008.
3. 平松壮太, 鶴木祐史, "変調伝達関数に基づいた残響時間のブラインド推定," 日本音響学会春季研究発表会, 2-6-10, pp. 683-684, 千葉工業大学, March 18, 2008.
4. Lu Xugang, Masashi Unoki, and Masato Akagi, "Comparative evaluations of modulation transfer function based dereverberation for robust speech recognition," 日本音響学会春季研究発表会講演論文, 1-10-12, pp. 39-40, 千葉工業大学, March 17, 2008.
5. 鶴木祐史, 細呂木谷敏弘, 石本祐一, "残響環境下におけるロバストで正確な基本周波数推定法の比較評価," 電子情報通信学会技術報告, SP2007-168, pp. 7-12, (TL2007-73, WIT2007-73) 島根大学, Jan. 25, 2008.
6. 平松壮太, 鶴木祐史, "変調伝達関数に基づいた残響時間のブラインド推定法の検討," 日本音響学会聴覚研究会資料, vol. 37, no. 11, H-2007-149, 京都大学, Jan. 23, 2008.
7. Masashi Unoki and Toshihiro Hosorogiya, "Robust and accurate F0 estimation for

- reverberant speech by utilizing complex cepstrum analysis," J. Acoust. Soc. Am., Vol. 122, No. 5, Pt. 2.2, pp. 3021, New Orleans, Louisiana, USA, Nov. 29, 2007.
8. 鶴木祐史, 細呂木谷敏弘, "複素ケプストラム分析を利用した残響音声の基本周波数推定法," 電子情報通信学会技術報告, Vol. 107, No. 186, pp. 31-36, EA2007-44, 東北大学, Aug. 9, 2007.
 9. 細呂木谷敏弘, 鶴木祐史, "残響環境における音声の基本周波数推定法に関する検討," 日本音響学会春季研究発表会, 1-Q-1, pp. 265-266, 千葉工業大学, March 17, 2007.
 10. Masashi Unoki, Masato Toi, Yohei Shibano, and Masato Akagi, "Suppression of speech intelligibility loss through an MTF-based speech dereverberation method," J. Acoust. Soc. Am., vol. 120, no. 5, Pt. 2, 5pAA12, pp. 3360, Honolulu, Hawaii, USA, Dec. 2, 2006.

[産業財産権]

○出願状況 (計 1 件)

名称: 残響時間推定装置及び残響時間推定方法

発明者: 鶴木祐史, 平松壮太

権利者: 北陸先端科学技術大学院大学

種類: 特願

番号: 2008-95540,

出願年月日: 平成 20 年 3 月 4 日

国内外の別: 国内

[その他]

平成 20 年度「良いシーズをつなぐ知の連携システム (つなぐしくみ)」支援課題

<http://www.jst.go.jp/pr/info/info624/besshil.html>

J-STORE:

http://jstore.jst.go.jp/cgi-bin/prompt/detail.cgi?prompt_id=6800

<http://jstshingi.jp/abst/jst/090619/program.html#6>

6. 研究組織

(1) 研究代表者

鶴木 祐史 (UNOKI MASASHI)

北陸先端科学技術大学院大学・情報科学研究科・准教授

研究者番号: 00343187

(2) 研究分担者

(3) 連携研究者