

## 科学研究費助成事業 研究成果報告書

令和 2 年 5 月 26 日現在

機関番号：22604

研究種目：研究活動スタート支援

研究期間：2018～2019

課題番号：18H06482・19K21546

研究課題名(和文)音環境理解に向けた音声信号の位相・振幅同時復元の理論構築の研究

研究課題名(英文)Study of speech signal amplitude and phase reconstructions for understanding sound environment

研究代表者

若林 佑幸(Wakabayashi, Yukoh)

首都大学東京・システムデザイン研究科・特任助教

研究者番号：80826462

交付決定額(研究期間全体)：(直接経費) 2,300,000円

研究成果の概要(和文)：本研究では、雑音によって劣化した音声信号から雑音の影響を除去し、真の音声信号を復元する理論の構築とその応用を実施した。従来研究では、音声信号の振幅・位相という二つの特徴に基いた復元アルゴリズムを別々に構築するのが一般的であった。これに対し、本研究ではこれら二つの特徴に深い関連があることに着目し、新しい復元理論の構築を行った。この関連性を考慮することで復元性能を改善した。また、他の応用として発話区間検出という基礎的な信号処理への適用を実施し、関連性を考慮した新たな手法の提案を行った。

研究成果の学術的意義や社会的意義

本研究において得られた、雑音抑圧と音声区間検出に対する結果が示すことは以下の通りである。学術的意義としては、これまで別々に研究されてきた振幅特徴と位相特徴の関連を考慮し、統合的に取り扱うことが信号処理の性能を改善する上で重要であり、より高性能な信号処理アルゴリズムの構築に繋がることが確認できたことである。社会的意義としては、雑音抑圧や音声区間検出の高性能化に伴い、今後の情報社会において必須となる音声認識性能の向上や遠隔会話システムにおける、より円滑な会話の実現が達成できることが挙げられる。

研究成果の概要(英文)：This project tackled the challenges of clean speech reconstruction from noisy observation by using relationships between speech amplitude and phase features.

In general, previous studies have separately constructed amplitude- and phase-based noise reduction algorithms. In contrast, the principal investigator proposed new algorithm that integrates the two features and confirmed its superiority over separately-handled method. In addition, he applied the relationship to another audio signal processing such as voice activity detection and showed that new integration method achieves higher performance than only amplitude-based method.

研究分野：音響信号処理

キーワード：位相信号処理 音声強調 雑音抑圧 音声区間検出

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

近年の情報社会を支える技術として、音声に代表されるような音響信号の処理技術は格段に需要が高まっている。音響信号処理の分野では、信号の振幅・位相情報に基づき、それぞれ別々に処理アルゴリズムを構築するのが一般的である。従来、振幅情報が重視されてきたが、位相情報を取り扱う信号処理が近年注目を集めている。振幅と位相の間には関連があるとされているが、これらそれぞれに基づく信号処理手法の間には関連がなく、独立に研究が進められていた。音響信号処理の高性能化のためには音響信号の振幅・位相情報の関連を考慮した理論や手法が必要とされている。

### 2. 研究の目的

本研究では、大きく次の二つを目的とする。

(1) 環境雑音によって劣化した音声信号から雑音の影響を除去し、所望の音声信号を復元することを目標とする。復元するうえで信号の振幅・位相の関連性を考慮に入れた復元アルゴリズムの提案を行う。

(2) 振幅・位相の関連性は上記雑音除去への応用だけでなく、その他の音響信号処理に応用できるものと期待される。そのため、基礎的な信号処理の一つである音声区間検出へ応用する。音声区間検出もこれまで振幅の情報のみを用いた手法が検討されているが、位相情報は全く考慮されていない。

### 3. 研究の方法

本研究では、音響信号の振幅・位相の関連性としてこれらが時間・周波数方向に同期するという特徴を利用した。本研究で取り組む二つの目的において異なる同期の考え方を導入する。

#### (1) 雑音抑圧

本目的において二つのアプローチをとる。第一は、既存の位相復元手法による結果を基にし、確率的に既存の振幅復元手法を統合させ同期するアプローチである。研究代表者がこれまでに取り組んできた位相復元手法の枠組みを既存の振幅復元手法の中に組み込んだものである。第二は、既存の振幅復元手法による結果に最も同期した位相を更新アルゴリズムにより復元し、統合するアプローチである。更新アルゴリズムには Griffin らによる手法【文献1】を拡張した手法を提案し、利用した。

#### (2) 音声区間検出

本目的において、位相情報の時間・周波数方向の変動が特に振幅情報と同様の挙動をするという事実を用いる。図1に示す予備的な解析の結果、位相の二階微分値(d)が一階微分(c)や位相情報そのもの(b)と比較し、振幅(a)と似た情報を持つことを示した。この位相の二階微分情報を用い、音声区間検出のアルゴリズム構築を行った。アルゴリズムの構築には二段階存在する。二階微分情報の利用手段と振幅との統合手段である。この二点について理論的・実験的な側面から取り組んだ。

### 4. 研究成果

(1) 研究代表者がこれまでに取り組んだ位相復元手法で雑音により劣化した位相情報から真の位相情報を推定することができる。この情報を、振幅推定のための既存の確率モデルへ組み込んだ。これは容易に解を求めることができ、その解を新たな振幅復元値とした。音質に関するシミュレーション実験を行った。図2は異なる2環境における信号対雑音(SN)比の結果であり、値が大きいほど高性能であることを示す。図2により、提案手法(proposed)は振幅・位相の独立処理(baseline 1)と比較し性能の改善が見られたが、他の独立処理(baseline 2)と比較して格段に優れた性能を示さなかった。また、理論的に考える上限値(oracle)には遠い。この原因として考えるのは、振幅と位相の復元のアプローチの違いである。本実験で適用した位相復元手法は音声の周波数方向の特徴を基に推定するのに対し、振幅復元は確率モデルを基に推定するため、推定基準があていない。統一した枠組みへの拡張が必要である。これらの結果は【文献2】にてまとめられた。

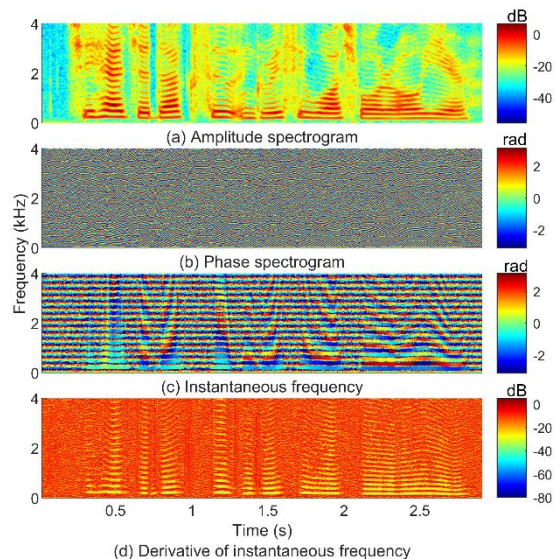


図1：時間周波数領域における音声の振幅、位相の変動

アルゴリズムの構築には二段階存在する。二階微分情報の利用手段と振幅との統合手段である。この二点について理論的・実験的な側面から取り組んだ。

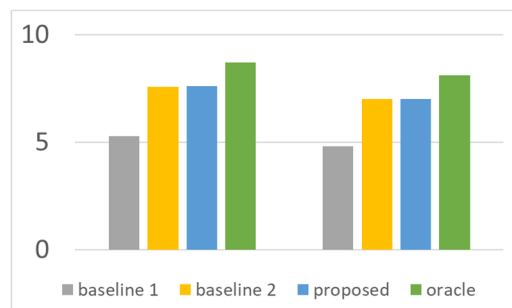


図2：取組(1)における音質評価結果

提案手法(proposed)は振幅・位相の独立処理(baseline 1)と比較し性能の改善が見られたが、他の独立処理(baseline 2)と比較して格段に優れた性能を示さなかった。また、理論的に考える上限値(oracle)には遠い。この原因として考えるのは、振幅と位相の復元のアプローチの違いである。本実験で適用した位相復元手法は音声の周波数方向の特徴を基に推定するのに対し、振幅復元は確率モデルを基に推定するため、推定基準があていない。統一した枠組みへの拡張が必要である。これらの結果は【文献2】にてまとめられた。

(2) 取り組み(1)は位相の変動を振幅復元に利用した方法である。それに対し、振幅の変動を位相復元に利用する方法も考えられる。このアプローチは多くの先行研究があり、中でも有名な手法として Griffin らによる更新アルゴリズム【文献1】がある。この手法は、ある振幅が与えられたときに最もその振幅に合った位相を繰り返し演算により求めるため、時間周波数解析による振幅情報が必要である。本取り組みでは、Griffin らによるアルゴリズムの位相復元メカニズムと時間周波数解析との依存関係に着目し、時間周波数解析に冗長度パラメータ  $\alpha$  を導入した。 $\alpha$  が増えるとアルゴリズムがより良い値に収束しやすいことを実験的に示すことができた。このアプローチを雑音除去に応用することができる。既存の振幅復元手法により振幅の雑音除去を行ったうえで、その値を、 $\alpha$  導入位相復元に適用した。音声信号に対する箱ひげ図による実験結果を図3に示す。縦軸は音質を評価する客観指標 PESQ であり、値が大きいほど高音質であることを示す。図を見ると、パラメータを大きくすることで、音質も改善する傾向であることを確認できる。PESQ が 0.1 ポイント改善することは、聴覚的に小さいながらも違いを感じることができる程度である。この結果は【文献3】にまとめられた。

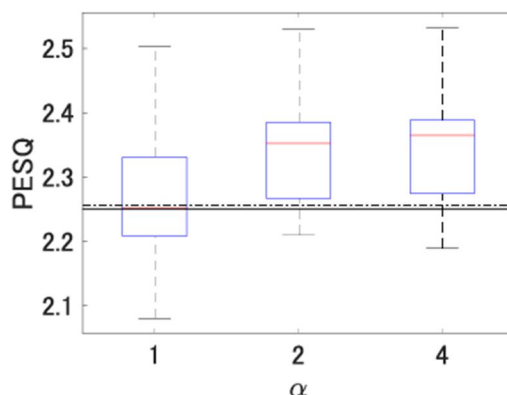


図3：取組(2)における音質評価結果

(3) 音声信号の振幅・位相の変動の同期を利用して、発話区間であるか非発話区間であるかを検出する問題、音声区間検出に取り組んだ。位相の二階微分を Derivative of instantaneous frequency (DIF)と定義し、この挙動を調査した。DIF のヒストグラムは図4に示すように音声区間(上段)と非音声区間(下段)で形状が異なる。これは背後にある各状態の確率モデルが異なることを示している。この差を音声区間検出に利用した。対象の信号の DIF ヒストグラムが事前に設定した非音声区間のモデル形状に近い場合、非音声区間、そうでなければ音声区間と判定した。位相情報に基づく検出手法と振幅情報に基づく手法を統合するため、次の二つのアプローチを提案した。一つ目は振幅、位相情報によるそれぞれの検出結果を論理和で統合するもの(統合手法A)、二つ目は確率分布上で統合するもの(統合手法B)である。既存の振幅情報のみによる検出【文献4】では、雑音環境下での正解率が 62.3%であったのに対し、統合手法Aでは 64.4%、統合手法Bでは 70.5%を示した。振幅・位相の関連を考慮することで性能の改善が確認できた。統合手法A、Bによる結果はそれぞれ【文献5】、【文献6】にまとめられた。

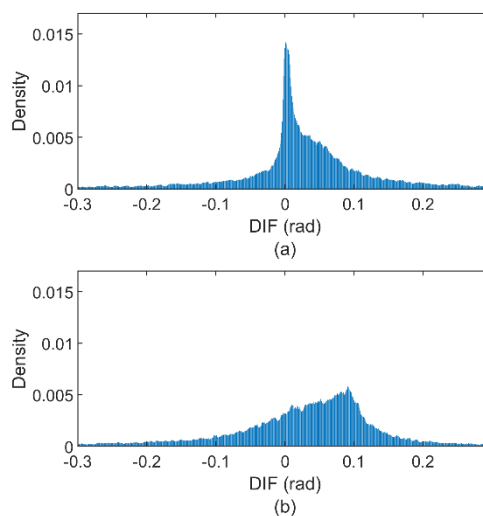


図4：音声(上)/非音声(下)区間における DIF ヒストグラムの例

< 引用文献 >

D. W. Griffin and J. Lim, "Signal estimation from modified short-time Fourier transform," IEEE Trans. on Acoust., Speech and Signal Process., vol. 32, no. 2, pp. 236–243, 1984.

**Yukoh Wakabayashi** and Nobutaka Ono, "Maximum a posteriori estimation of spectral gain with harmonic-structure-based phase reconstruction for phase-aware speech enhancement," Proc. APSIPA ASC 2018, Hawaii, USA, Nov. 2018.

**Yukoh Wakabayashi** and Nobutaka Ono, "Griffin-Lim phase reconstruction using short-time Fourier transform with zero-padded frame analysis," Proc. APSIPA ASC 2019, pp.1863-1867, Lanzhou, China, Nov. 2019.

J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," IEEE signal processing letters, vol. 6, no. 1, pp. 1–3, 1999.

Binh Thien Nguyen, **Yukoh Wakabayashi**, Takahiro Fukumori, and Takanobu Nishiura, "Derivative of Instantaneous Frequency for Voice Activity Detection Using Phase-based Approach," Proc. APSIPA ASC 2019, Lanzhou, China, Nov. 2019.

Binh Thien Nguyen, **Yukoh Wakabayashi**, Takahiro Fukumori, and Takanobu Nishiura, "A phase-based voice activity detection using statistical likelihood ratio of the derivative of instantaneous frequency," 日本音響学会 2020 年春季研究発表会, Saitama, Mar. 2020.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 0件/うち国際共著 0件/うちオープンアクセス 2件）

|  |                       |
|--|-----------------------|
| 1. 著者名<br>Yukoh Wakabayashi  | 4. 巻<br>40            |
| 2. 論文標題<br>Speech enhancement using harmonic-structure-based phase reconstruction                              | 5. 発行年<br>2019年       |
| 3. 雑誌名<br>Acoustic Science & Technology  | 6. 最初と最後の頁<br>162-169 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br><a href="https://doi.org/10.1250/ast.40.162">https://doi.org/10.1250/ast.40.162</a> | 査読の有無<br>無            |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である）  | 国際共著<br>-             |

|  |                       |
|--|-----------------------|
| 1. 著者名<br>若林佑幸   | 4. 巻<br>75            |
| 2. 論文標題<br>調波構造に基づく位相復元を利用した音声強調   | 5. 発行年<br>2019年       |
| 3. 雑誌名<br>日本音響学会誌  | 6. 最初と最後の頁<br>139-146 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br><a href="https://doi.org/10.20697/jasj.75.3_139">https://doi.org/10.20697/jasj.75.3_139</a> | 査読の有無<br>無            |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である）  | 国際共著<br>-             |

〔学会発表〕 計11件（うち招待講演 1件/うち国際学会 3件）

|  |
|--|
| 1. 発表者名<br>若林佑幸, 小野順貴                    |
| 2. 発表標題<br>周波数を冗長化したSTFTによる位相復元の音声強調への利用 |
| 3. 学会等名<br>日本音響学会2019秋季研究発表会             |
| 4. 発表年<br>2019年                          |

|  |
|--|
| 1. 発表者名<br>若林佑幸, 小野順貴                    |
| 2. 発表標題<br>周波数を冗長化したSTFTによる位相復元の音声強調への利用 |
| 3. 学会等名<br>日本音響学会2019秋季研究発表会             |
| 4. 発表年<br>2019年                          |

|   |
|---|
| 1. 発表者名<br>Binh Thien Nguyen, Yukoh Wakabayashi, Takahiro Fukumori, and Takanobu Nishiura |
| 2. 発表標題<br>Speech analysis using the second derivative of phase spectrum                  |
| 3. 学会等名<br>日本音響学会2019秋季研究発表会  |
| 4. 発表年<br>2019年   |

|  |
|--|
| 1. 発表者名<br>Yukoh Wakabayashi and Nobutaka Ono  |
| 2. 発表標題<br>Griffin-Lim phase reconstruction using short-time Fourier transform with zero-padded frame analysis |
| 3. 学会等名<br>APSIPA ASC 2019 (国際学会)  |
| 4. 発表年<br>2019年  |

|  |
|--|
| 1. 発表者名<br>Binh Thien Nguyen, Yukoh Wakabayashi, Takahiro Fukumori, and Takanobu Nishiura                |
| 2. 発表標題<br>Derivative of Instantaneous Frequency for Voice Activity Detection Using Phase-based Approach |
| 3. 学会等名<br>APSIPA ASC 2019 (国際学会)  |
| 4. 発表年<br>2019年  |

|  |
|--|
| 1. 発表者名<br>若林佑幸, 小野順貴                    |
| 2. 発表標題<br>周波数方向冗長化STFTに基づく位相復元の音楽信号への適用 |
| 3. 学会等名<br>日本音響学会2020春季研究発表会             |
| 4. 発表年<br>2020年                          |

|   |
|---|
| 1. 発表者名<br>Binh Thien Nguyen, Yukoh Wakabayashi, Takahiro Fukumori, and Takanobu Nishiura   |
| 2. 発表標題<br>A phase-based voice activity detection using statistical likelihood ratio of the derivative of instantaneous frequency |
| 3. 学会等名<br>日本音響学会2020春季研究発表会  |
| 4. 発表年<br>2020年   |

|  |
|--|
| 1. 発表者名<br>若林 佑幸, 小野 順貴                  |
| 2. 発表標題<br>短時間フーリエ変換の周波数方向冗長化とその位相復元への応用 |
| 3. 学会等名<br>日本音響学会2019年春季研究発表会            |
| 4. 発表年<br>2019年                          |

|   |
|---|
| 1. 発表者名<br>Yukoh Wakabayashi and Nobutaka Ono   |
| 2. 発表標題<br>Maximum a posteriori estimation of spectral gain with harmonic-structure-based phase reconstruction for phase-aware speech enhancement |
| 3. 学会等名<br>APSIPA ASC 2018 (国際学会)   |
| 4. 発表年<br>2018年   |

|                                    |
|------------------------------------|
| 1. 発表者名<br>若林佑幸                    |
| 2. 発表標題<br>音声強調における位相復元とアレイ処理への効果  |
| 3. 学会等名<br>電子情報通信学会 応用音響研究会 (招待講演) |
| 4. 発表年<br>2018年                    |

|  |
|--|
| 1. 発表者名<br>若林 佑幸, 小野 順貴                      |
| 2. 発表標題<br>音声強調における位相スペクトルと振幅スペクトル同時推定手法の一検討 |
| 3. 学会等名<br>日本音響学会2018年秋季研究発表会                |
| 4. 発表年<br>2018年                              |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

|  | 氏名<br>(ローマ字氏名)<br>(研究者番号) | 所属研究機関・部局・職<br>(機関番号) | 備考 |
|--|---------------------------|-----------------------|----|
|  |                           |                       |    |