

## 高フロアリング係数を用いた反復スペクトルサブトラクションによる ミュージカルノイズの低減の検討

### A study of musical tone reduction on iterative spectral subtraction based on higher flooring parameters

福森 隆寛<sup>†</sup> 森勢 将雅<sup>\*</sup> 西浦 敬信<sup>\*</sup> 南條 浩輝<sup>\*\*</sup>

Takahiro Fukumori<sup>†</sup> Masanori Morise<sup>\*</sup> Takanobu Nishiura<sup>\*</sup> Hiroaki Nanjo<sup>\*\*</sup>

#### 1. はじめに

マイクロホンを意識せずに音声対話が可能なハンズフリー音声対話システムは、遠隔通信システムの実現に重要なアプリケーションとして注目されている一方、背景雑音の混入が問題視されており、明瞭な会話は困難とされている。これまでに雑音を含む音声から雑音を抑圧する方法は検討されている。特にスペクトルサブトラクション(SS) [1]は、雑音混入音声から音声強調を行う方法であり、低演算量で雑音抑圧が可能のため広く利用されている。しかしSSによる雑音抑圧後の音声にはミュージカルノイズ[2]と呼ばれる残存雑音の発生が問題となる。ミュージカルノイズは受聴者にとって非常に不快な雑音であるため、ハンズフリー音声対話システムなどでSSを利用するには、雑音の抑圧に加えミュージカルノイズを抑圧する必要がある。そこで本稿ではミュージカルノイズが発生しにくいSS法を提案する。

#### 2. スペクトルサブトラクション法

スペクトルサブトラクション法(SS)は、目的信号と雑音が混合した観測信号のパワースペクトルから雑音のパワースペクトルを減算することで雑音を抑圧する方法である。目的信号と雑音が無相関である場合、式(1)を用いてスペクトル減算を行う。

$$|D(\omega, \tau)|^2 = |Y(\omega, \tau)|^2 - \alpha |\hat{N}(\omega)|^2, \quad (1)$$

$$|\hat{X}(\omega, \tau)|^2 = \begin{cases} |D(\omega, \tau)|^2 & (P) \\ \beta |Y(\omega, \tau)|^2 & (N), \end{cases}$$

$$(P) \quad |D(\omega, \tau)|^2 > \beta |Y(\omega, \tau)|^2$$

$$(N) \quad \text{otherwise}$$

ここで $\tau$ はフレーム番号を表す離散時間、 $Y(\omega, \tau)$ は観測信号(時間領域)のフレーム番号 $\tau$ におけるスペクトル、 $\hat{X}(\omega, \tau)$ は推定された雑音のパワースペクトルに基づいて算出された目的音声のスペクトル、 $\hat{N}(\omega)$ は観測信号の音声を含まない区間から推定された雑音のスペクトル、 $\alpha$ は減算係数を、 $\beta$ はフロアリング係数を表す。そして本稿では式(1)の(N)の演算をフロアリング処理とする。一般的に $\alpha$ は1よりも大きく、 $\beta$ は $0 < \beta \ll 1$ と定義されている。

<sup>†</sup> 立命館大学大学院理工学研究科

Graduate School of Science and Engineering, Ritsumeikan University

<sup>‡</sup> 立命館大学 情報理工学部

College of Information and Science, Ritsumeikan University

<sup>‡‡</sup> 龍谷大学 理工学部

Faculty of Science and Technology, Ryukoku

University

そして推定された目的音声のパワースペクトルと観測信号の位相を用いて逆フーリエ変換することで雑音抑圧された音声信号が得られる。しかし雑音抑圧後の信号にミュージカルノイズが発生するという問題がある。

#### 3. 提案手法

ミュージカルノイズは、スペクトログラム上に複数存在する局所的なピークである[3]とされているが、何故局所的なピークが生じるかはについて詳細に検討されていない。SS処理において1フレーム中のスペクトル減算で条件(P)と(N)が混在する場合に $\beta$ が0に近いとパワースペクトルのダイナミックレンジが大きくなりミュージカルノイズが発生すると考えられる。それに対して $\beta$ が1に近い場合では、条件(P)と(N)が混在する場合においてもパワースペクトルのダイナミックレンジは大きく変化しないが、雑音抑圧信号のパワーも十分に抑圧されないため、雑音抑圧量の低下が予想される。そこで1に近い $\beta$ で反復してSSを行うことにより、ミュージカルノイズの発生を抑えつつ、高い雑音抑圧性能の達成が期待できる。本稿では適切な $\beta$ と反復回数を決定するための客観・主観評価実験を実施する。

#### 4. 評価実験

提案法の有効性を示すために、客観・主観評価実験を行った。客観評価は提案SS法が従来SS法に比べ雑音抑圧量が優れていることを示すために行う。そして主観評価は提案手法がミュージカルノイズの影響を従来法よりも抑圧可能であることを示すために行う。

実験には「IPA-98-testset」[4]を音声データベースとして用いた。なお収録音声の標本化周波数は16 kHz、量子化ビット数は16 bitである。音声データベースに付加する雑音はホワイトノイズとした。そしてSS処理におけるフレーム長を64 ms、シフト長を32 ms、高速フーリエ点数を1024点、減算係数 $\alpha$ を2.0とし、推定する雑音パワースペクトルは先頭7フレームの平均パワースペクトルとした。そしてフロアリング係数 $\beta$ は従来法の0.01に加えて、0.1から0.1間隔で0.9までとし、SSの反復回数は0~50回までとした。

##### 4.1 NRRに基づく客観評価実験

はじめに提案法が従来法と同雑音抑圧性能を達成可能か検証するための客観評価を行う。雑音抑圧性能を評価するためにNRR(Noise Reduction Rate)を用いて評価する。NRRは雑音信号に対する雑音抑圧信号のエネルギー比を示し、以下の定義式を用いて算出される。

$$NRR = 10 \log_{10} \left( \frac{\sum_{m=0}^{M-1} n^2(m)}{\sum_{m=0}^{M-1} \hat{x}^2(m)} \right) \quad (2)$$

ここで  $n(n)$  は雑音信号,  $\hat{x}(n)$  を雑音抑圧信号, そして  $M$  は両信号長を表す. 本実験における  $\hat{x}^2(n)$  は観測信号における目的信号  $x(m)$  を 0 として計算される. 雑音抑圧性能が高いほど NRR は高い数値を示す.

図1にNRRの評価結果を示す.  $\beta$  が小さく反復を行わない従来法は, 反復回数0回の提案法よりも高いNRRを示した. 一方, 提案法のNRRは反復回数の増加に伴い, 従来のSSよりも高い数値を示すことがわかる. 本実験より, 高いフロアリング係数  $\beta$  とSSの反復回数を組み合わせた提案手法は従来のSS法 ( $\beta : 0.01$ , 反復なし) よりも高い雑音抑圧性能を達成できたといえる.

## 4.2 主観評価実験

提案法によってミュージカルノイズの発生を抑圧可能であることを示すため, 2種類の主観評価実験を行った. 評価実験は10名の被験者(男女各5名)を対象として行った. なお事前にホワイトノイズと従来SS法によってホワイトノイズを抑圧した信号を学習させている. 実験中に提示する信号は音圧レベルを70 dBA, 提示時間を3秒間, 提示順序をランダムとした. 被験者は提示された信号がホワイトノイズとミュージカルノイズのどちらに近いのかを5段階評価(1:ミュージカルノイズ, 5:雑音各種)で評価した.

### 4.2.1 $\beta$ に基づくミュージカルノイズ発生を抑圧

はじめに高いフロアリング係数を用いた提案手法によるミュージカルノイズ発生を抑圧効果を検証した. SSの反復回数はフロアリング係数ごとに変更し, 従来SS法 ( $\beta : 0.01$ , 反復回数: 0回)と同NRRを示した反復回数とした. また本実験は同一音源を2回を含む, 合計20種類の信号(提示雑音: 10種類, 実験回数: 2回)を提示した.

図2に実験結果を示す. なお図中の  $x$  は全回答の平均を縦線は標準偏差を表す. 図より高いフロアリング係数  $\beta$  を用いた提案法では, 従来SS法と同量の雑音を抑圧しつつミュージカルノイズの発生も抑圧出来ることが分かる.

### 4.2.2 反復回数に基づくミュージカルノイズ発生を抑圧

次いで改良フロアリング係数 ( $\beta : 0.7$ )に基づく提案法において, ミュージカルノイズ発生を抑圧出来る反復回数について検証した. 本実験ではSSの反復回数を5~5回間隔で35回までとし, その他の実験条件は前実験と同一とした. そして被験者に合計14種類の信号(提示雑音: 7種類, 実験回数: 2回)を提示して主観評価実験を行った.

図3に実験結果を示す. SSの反復回数の増加に伴い, スコアは減少する傾向が示された. この結果はSSの反復回数の増加によって雑音を効果的に抑圧出来る一方, ミュージカルノイズが発生することを示している. ミュージカルノイズの発生を効果的に抑圧するためには, 極力少ない回数でSSを反復する必要がある. 本結果は10回程度の反復であれば, ミュージカルノイズの発生を十分抑圧できることを示している.

## 5. おわりに

本稿では改良フロアリング係数に基づいた反復SS法を提案した. 提案法の有効性を示すためにNRRを用いた客観評価, および目的音声の品質を評価するための主観評価実験を行った. 各実験結果より, 提案法を用いることで効果的に雑音抑圧および音声強調ができたと同時にミュージカルノイズの発生も抑圧出来ることを確認した.

本稿ではSS法の減算係数  $\alpha$  については検討していないため, ミュージカルノイズの発生を効果的に抑圧するための  $\beta$  以外のパラメータ最適化が今後の課題である. また今回は定常雑音を用いて評価したが, 今後は非定常雑音の抑圧も目指す予定である.

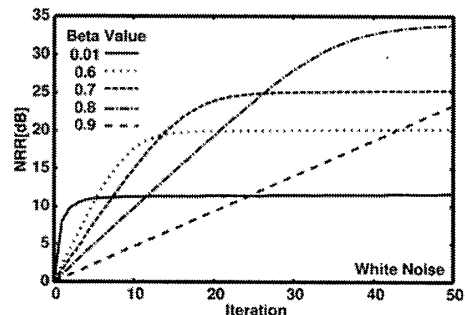


図1: NRRを用いた客観評価結果

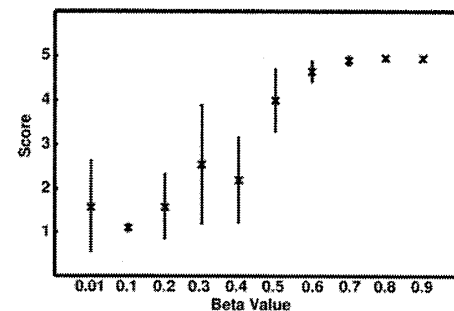


図2:  $\beta$ に関する主観評価結果

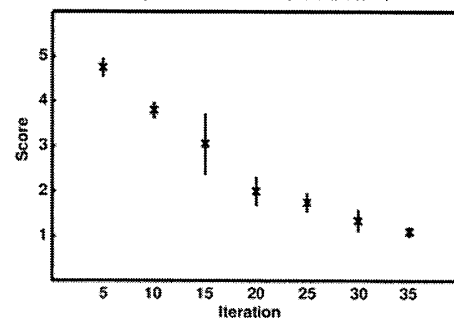


図3: 反復回数に関する主観評価結果

## 謝辞

本研究の一部は科学研究費補助金 若手(B)20700169の支援を受けて行われた.

## 参考文献

- [1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on acoustic, speech, and signal processing, vol. ASSP-27, no.2, pp. 113-120, 1979.
- [2] S. V. Vaseghi, "Advanced signal processing and noise reduction," John Wiley & Son Ltd, 1995.
- [3] Z. Goh, K. C. Tan and B. T. G. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," IEEE Transactions on Speech Audio Processing, vol. 6, no. 3, pp. 287-292, May, 1998.
- [4] T. Kawahara et al. "Free Software Toolkit For Japanese Large Oral Vocabulary Continuous Speech Recognition," Proc. ICSLP2000, vol. 4, pp. 4893-4896, 2000.