

フルート演奏音と管体振動の相関に基づく息音の分離と混合量の調節*

☆原明優, 吉野夏樹, 矢田部浩平 (農工大)

1 はじめに

フルートの吹鳴に伴う息の音は、楽器特有の響きや音色を構成する重要な要素である [1]。しかし、音響制作などによる録音では、特定のフレーズなどにおいて息の音量が制作者や奏者の意図と異なる場合がある。その場合、録音データの息音を後から自在に調節できれば、録音し直すことなく表現意図に沿った音色編集が可能となる。そこで、本稿ではマイクロホン信号に含まれる息音と調波音の音量をそれぞれ調節可能とするための分離手法を提案する。本手法では、息音の影響を受けにくく、管体振動から調波音を優位に取得できるピエゾセンサ信号を補助情報として利用する。具体的には、マイクロホン信号とピエゾセンサ信号の相関解析に基づいて両者を分離する。実験により、マイクロホン信号から分離された息音成分を任意の混合量で調整可能であることを示した。

2 時間周波数マスキングによる音源分離

音源分離において、時間周波数マスキングに基づく手法が広く用いられている。これは、観測信号の時間周波数領域の各ビンに重み（マスク）を乗算し、目的音信号を得るアプローチである。まず、時間周波数領域の観測信号 $X[t, f]$ をもとに目的音成分を抽出するためのマスク $M[t, f] \in [0, 1]$ を作成する。ここで、 $t, f \in \mathbb{N}$ はそれぞれ時間フレーム、周波数ビンを表すインデックスである。続いて、このマスク $M[t, f]$ を観測信号 $X[t, f]$ に要素ごとに乗算することで目的音成分の推定値

$$\hat{S}[t, f] = M[t, f] X[t, f] \quad (1)$$

を得る。従って、マスク $M[t, f]$ の精度が目的音成分の分離精度に直接影響する。

本研究では、マイクロホンによって観測されたフルート演奏音を、主に正弦波成分で構成される調波音と、それに付随して広帯域に広がる息音の2つの信号に分離するマスクの作成を扱う。これら2つの成分のエネルギーは時間的に同期するため、単一の観測信号 $X[t, f]$ のみから両成分を分離するマスクを高精度に推定することは容易ではない。そこで、調波音成分の情報が優位に含まれる補助信号を用いてマスクを設計する手法を提案する。

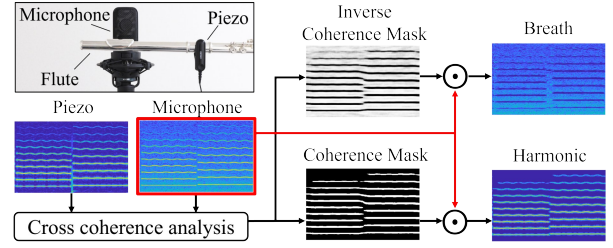


図-1 提案手法の概要。⊙ は要素ごとの積を表す。

3 提案手法

本手法では、マイクロホン信号に対する時間周波数マスキングによって、調波音成分と息音成分の分離を行う。マスクを設計するための補助情報として、息音の影響を受けにくく調波音を強く取得できるピエゾセンサ信号を用いる。この性質に基づいて、両信号間の局所相関を算出すると、調波音成分が支配的な時間周波数ビンでは高い相関、息音成分が支配的なビンでは低い相関を示す。これにより設計されたマスクによって息音成分を分離して、息音量を自在に調節した音響信号を作成する。

提案手法の概要を図-1に示す。まず、マイクロホン信号とピエゾセンサ信号に離散ガボール変換（DGT）を適用し、時間周波数表現を得る。次に、両信号間の局所相関から調波音成分を抽出するためのバイナリマスクを作成し、これをマイクロホン信号に乗算することで、分離された信号を得る。

3.1 相関の算出

調波音成分を抽出するマスク作成のため、各ビン及びその近傍領域における両信号間の局所相関を

$$R[t, f] = \frac{|P_{pm}[t, f]|}{\sqrt{P_{pp}[t, f] P_{mm}[t, f]}} \quad (2)$$

と算出する。ここで、 $P_{ab}[t, f]$ ($a, b \in \{m, p\}$) は指定した近傍領域

$$\mathcal{N}[t, f] = \{[t, f'] \mid f - \Delta f \leq f' \leq f + \Delta f\} \quad (3)$$

における局所的なクロススペクトル ($a = b$ の場合はパワースペクトル) であり、

$$P_{ab}[t, f] = \sum_{[t', f'] \in \mathcal{N}[t, f]} X_a[t', f'] \overline{X_b[t', f']} \quad (4)$$

*Separation and adjustment of breath sound in recorded flute performance based on correlation between microphone and vibration signals. By Miyu HARA, Natsuki YOSHINO and Kohei YATABE (Tokyo University of Agriculture and Technology)

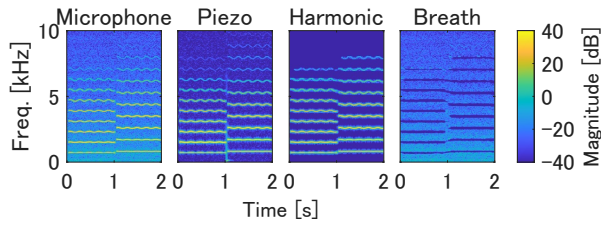


図-2 シミュレーション実験におけるテスト信号と分離後のスペクトログラム. 左からマイクロホン信号, ピエゾセンサ信号, 調波音成分, 息音成分である.

表-1 分離前後の SI-SDR

評価対象成分	分離前 [dB]	分離後 [dB]	Δ SI-SDR [dB]
調波音成分	30.36	37.24	+6.88
息音成分	-24.37	6.52	+30.89

と定義される. ただし, Δf は周波数近傍幅のパラメータであり, $X_m[t, f]$, $X_p[t, f]$ は両信号の時間周波数表現, $\overline{(\cdot)}$ は複素共役を表す. また, m, p はそれぞれマイクロホンとピエゾセンサを示す添え字である. 両信号に共通して含まれている調波音成分は相関が高く, マイクロホン信号にのみ多く含まれている息音成分の相関は低くなる.

3.2 マスクによる分離と調節

算出された相関 $R[t, f]$ に対して, ミュージカルノイズの抑制のため, メディアンフィルタを適用した. 平滑化された相関に基づき, 閾値を $\theta \in [0, 1]$ として, 高い相関を持つ調波音成分を 1 として抽出するためのバイナリマスク

$$M[t, f] = \begin{cases} 1 & (\text{if } R[t, f] > \theta) \\ 0 & (\text{otherwise}) \end{cases} \quad (5)$$

を作成する. 推定されたマスク $M[t, f]$ 及び $1 - M[t, f]$ をそれぞれマイクロホン信号 $X_m[t, f]$ に乗算し, その結果を逆 DGT することで, 分離された調波音成分と息音成分を得る. 最終的に, 分離した両成分を任意の割合で混合することにより, 息音の量を自在に調節した音響信号を生成する.

4 実験

シミュレーションと実測データによる実験を行った. 両実験で共通して, サンプル周波数は 44.1 kHz, Hann 窓を用いて窓長 2048, シフト長 512 として DGT を適用した. $\Delta f = 2$, $\theta = 0.85$, メディアンフィルタサイズは 8×15 とした.

4.1 シミュレーション実験

図-2 に示すような 2.0 秒間の調波音成分と息音成分から成る模擬信号によるシミュレーションを行った. また, 分離性能は SI-SDR を用いて評価した.

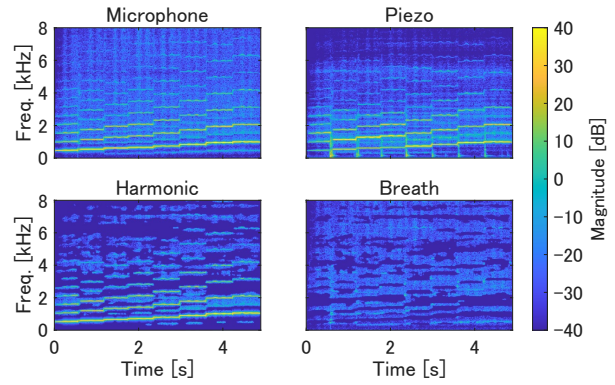


図-3 実測信号と分離後のスペクトログラムの拡大図. 上段左からマイクロホン信号, ピエゾセンサ信号. 下段左から調波音成分, 息音成分である.

実験に用いたマイクロホン信号, ピエゾセンサ信号, 及び分離された各成分のスペクトログラムを図-2 に示す. 分離後はマイクロホン信号とピエゾセンサ信号で共通に含まれている調波音成分と, それ以外の息音成分が分離されていることが分かる. 各成分の分離前後における SI-SDR とその改善量 (Δ SI-SDR) を表-1 に示す. Δ SI-SDR が両成分において増加していることから, 提案手法により両者の劣化を防ぎつつ分離可能であることが示された.

4.2 実測実験

提案手法の有効性を実際の演奏環境において検証するため, 実録音データを用いて分離実験を行った. 録音はヤマハの防音室 AMDC35C で行い, 図-1 左上のようにマイクロホンは吹口から約 15 cm 離れた位置, ピエゾセンサは管体中部にそれぞれ設置した.

実測データに対する各成分の分離結果のスペクトログラムを図-3 に示す. マイクロホン信号に対してピエゾセンサ信号では息音の影響が少なく, 調波音成分の構造が明瞭に観測される. また, マイクロホン信号に混在していた調波音成分と息音成分が分離されていることが確認できた. これにより, 息音成分が任意の混合量で調節可能となった.

5 むすび

本稿では, マイクロホン信号とピエゾセンサ信号の相関に基づく分離手法を提案し, 実験により息音成分の分離と任意の混合量の調節が可能であることを示した. 今後の課題として, より高精度なマスク推定法の検討が挙げられる.

参考文献

- [1] B. Fabre, J. Gilbert and A. Hirschberg, "Springer handbook of systematic musicology," in *Springer Handbook of Systematic Musicology*, (Springer Berlin Heidelberg, Berlin, Heidelberg, 2018), Chap. 7, pp. 121–139.