

圧電素子を用いた目的音抽出法

高橋直也[†] 橋本周司[†]

早稲田大学院理工学研究科物理学及応用物理学専攻

1. はじめに

コンサートで複数の楽器を同時に演奏する際、音量のバランス等を調整するために、特定の楽器の音を取り出す必要が生じる。マイクロフォンアレイ[1][2]は目的音を抽出する有効な方法であるが、コンサートホールの様な反響のある部屋での適用は困難である。独立成分分析もまた効果的に目的音を抽出する手法であるが、非定常な音源には適用できない[3]。そのため、現状では圧電素子を楽器に取り付けて目的音を抽出する手法が用いられているが、得られた信号は圧電素子の周波数特性や、楽器が共鳴する前の音を抽出するという理由から、本来の音色とは大きく異なり、不自然な音色となる。本研究では、圧電素子の信号から目的音の存在する周波数領域を求め、適当な位置に置いたマイクロフォンの信号から他の楽器の音が存在する周波数領域のパワーを下げるこことにより、目的音を抽出する方法を提案する。

2. 提案手法

2.1. 圧電素子の特徴

提案手法ではマイクロフォンと圧電素子を用いる。圧電素子は楽器の表面に圧電素子を貼り付けることで楽器自体の振動を電気信号として捕らえることができる。そのため、他の楽器の音は混入しにくい。圧電素子から得られる信号のスペクトルは、聴衆が聞くものとは異なるが、貼付された楽器と同じ周波数成分を持っている。

2.2. 目的音抽出法

図1に提案手法の流れを示す。マイクロフォンと圧電素子を適当な位置に設置し、目的の楽器音を取得する。マイクロフォン及び圧電素子から得られる信号をそれぞれ $x(t), y(t)$ とすると、

$$x(t) = m_1(t) * o(t) + \sum_i m'_i(t) * n_i(t) \quad (1)$$

$$y(t) = p_1(t) * o(t) + \sum_i p'_i(t) * n_i(t) \quad (2)$$

と表すことができる。但し、 $o(t), n_i(t)$ はそれぞれ目的の楽器音と目的楽器以外の i 番目の楽器音、

Sound focus combining microphone and piezoelectric device

[†]Naoya Takahashi · Shuji Hashimoto · School of Science and Engineering, Waseda University

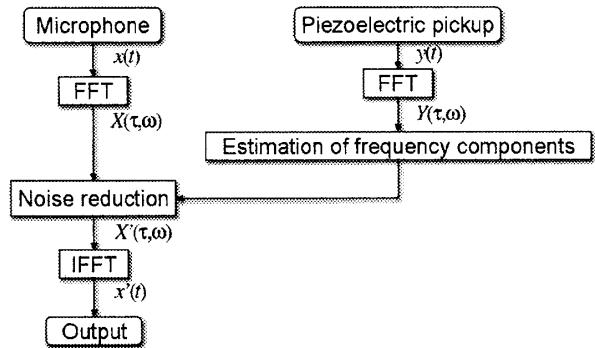


図1. 提案手法の流れ

$m_1(t), p_1(t)$ は、目的の楽器からマイクロフォン及び圧電素子までのインパルス応答、 $m'_i(t), p'_i(t)$ は、 i 番目の楽器からマイクロフォン及び圧電素子までのインパルス応答である。 $*$ は畳み込みを表している。我々の目標は目的信号 $s(t) = m_1(t) * o(t)$ を $x(t)$ から抽出することである。以下では簡単のために $i=1$ としても説明する。

$x(t), y(t)$ の短時間フーリエ変換 $X(\tau, \omega), Y(\tau, \omega)$ はフレーム時間 τ 及び角速度 ω を用いて次式で表される。

$$X(\tau, \omega) = M_1(\tau, \omega)O(\tau, \omega) + M_2(\tau, \omega)N(\tau, \omega) \quad (3)$$

$$Y(\tau, \omega) = P_1(\tau, \omega)O(\tau, \omega) + P_2(\tau, \omega)N(\tau, \omega) \quad (4)$$

但し、 $O(\tau, \omega), N(\tau, \omega)$ はそれぞれ $o(t), n(t)$ のフーリエ変換である。 $M_1(\omega), M_2(\omega)$ は、目的楽器及び他の楽器からマイクロフォンまでの伝達関数である。同様に $P_1(\omega), P_2(\omega)$ は、目的楽器及び他の楽器から圧電素子までの伝達関数である。

まず、目的音と他の音の周波数成分が重なっていない状況として以下の関係を考える。

$$O(\tau, \omega)N(\tau, \omega) = 0 \quad (\forall \omega) \quad (5)$$

圧電素子から得られる信号には目的の楽器音以外に他の楽器音も含まれるが、そのパワーは目的音に比べて十分に小さいため閾値処理により分離することができる。つまり、

$$\begin{cases} X'(\tau, \omega) = \begin{cases} X(\tau, \omega) & Y(\tau, \omega) \geq th \\ 0 & Y(\tau, \omega) < th \end{cases} \\ = M_1(\omega)O(\tau, \omega) \end{cases} \quad (6)$$

但し、 $X'(\tau, \omega)$ は処理後のスペクトルを表す。

一方、目的音と他の楽器音の周波数成分が重

なっているときの出力は次のようになる。したがって、この処理により部分的に他の楽器の音を除去できることがわかる。

$$X'(\tau, \omega) = \begin{cases} M_1(\omega)O(\tau, \omega) + M_2(\omega)N(\tau, \omega) & \text{if } Y(\tau, \omega) > th \\ 0 & \text{if } Y(\tau, \omega) < th \end{cases} \quad (7)$$

2.3. 時間的に滑らかな除去処理

式(7)で与えられる処理は $Y(\tau, \omega)$ の値によって通過させる周波数をスイッチングしている。急激なスイッチングはフレーム間で不連続性を生み、ノイズを生じさせる。そこで滑らかにスイッチングを行う必要がある。これを実現するため透過係数 $a(\tau, \omega)$ を次式の様に導入する[4]。

$$\left\{ \begin{array}{l} X'(\tau, \omega) = a(\tau, \omega) \cdot X(\tau, \omega) \\ a(\tau, \omega) = \begin{cases} d \cdot a(\tau-1, \omega) & Y(\tau, \omega) < th \\ 1 - r + r \cdot a(\tau-1, \omega) & Y(\tau, \omega) \geq th \end{cases} \end{array} \right. \quad (8)$$

where $0 < d < 1$, $0 < r < 1$, $a(0, \omega) = 1$

$a(\tau, \omega)$ は 0 から 1 の間の値を取り、指数関数的にフレーム毎に更新される。 d , r は減少及び増加の速さを決める定数である。処理後の出力 $x'(t)$ は $X'(\tau, \omega)$ を逆 FFT することにより得られる。

$$\left\{ \begin{array}{l} x'(t) = F^{-1}[X'(\tau, \omega)] = F^{-1}[a(\tau, \omega) \cdot X(\tau, \omega)] \\ \equiv G[x(t)] \end{array} \right. \quad (9)$$

但し $F^{-1}[\cdot]$ は逆フーリエ変換を、 $G[\cdot]$ は提案手法の関数を表す。

3. 実験

3.1. 定量的性能評価

提案手法の性能を定量的に評価するため、SNR、及び SDR を求めた。目的音 $s(t)$ は箏とし、他の楽器の音 $n(t)$ としてドラム、ギター、エレキベースを含む録音音源をスピーカから約 100dB で再生した。配置を図 2 に示す。データは箏、及び再生音源を別々にマイクロフォンと圧電素子を用いて録音した。マイクロフォン位置の任意性を示すため、箏の上部、前方、下部の位置においてそれぞれ評価した。は処理前、処理後の SNR、及び SDR はそれぞれ次式で表される。

$$SNR_{Original} = 10 \log_{10} \frac{\sum_{t=0}^T s(t)^2}{\sum_{t=0}^T n(t)^2} \quad (10)$$

$$SNR_{Proposed} = 10 \log_{10} \frac{\sum_{t=0}^T G[s(t)]^2}{\sum_{t=0}^T G[n(t)]^2} \quad (11)$$

$$SDR = 10 \log_{10} \frac{\sum_{t=0}^T s(t)^2}{\sum_{t=0}^T (s(t) - G[s(t)])^2} \quad (12)$$

但し T はデータ長である。表 1 に SNR、SDR を示す。提案手法は全ての位置で高い SDR を示しながらも、SNR を 4.9~9.2dB 改善している。

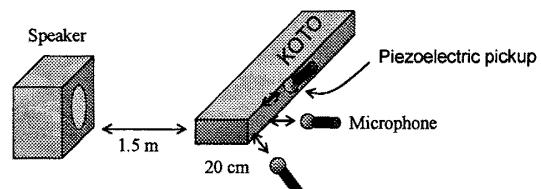


図 2. 実験環境

表 1 SNR と SDR

	Upper	Front	Bottom
SNR _{Original}	-0.20	-4.46	2.91
SNR _{Proposed}	9.03	2.74	7.84
SDR	20.0	19.7	17.1

表 2 FFT のパラメータ

Window size	2,048
Overlap	75%
Window	Hamming
Sampling frequency	44.1kHz

3.2. 実環境実験

Max/MSP によりリアルタイムシステムを作成した。3.1 節と同じ環境で、同時に演奏し本手法を適用したところ、聴覚的に十分な音量と自然な音色を得ることが可能であることを確認できた。

4. まとめと今後の展望

圧電素子を用いた目的音抽出法を提案し、作成した。高い SNR、SDR を得ることを実験により確認した。提案手法は完全に目的音を分離するものではないが、マイク位置を気にせず、例えば箏とドラム、エレキベースなど音量差の大きい楽器間でのセッションを可能にし、音楽の幅を広げるものである。今後はパラメータの自動決定など実用化への検討をさらに進めたい。

本研究の一部は CREST 「デジタルメディア作品の製作を支援する基盤技術」によっている。

参考文献

- [1] 野村、金田、小島 “近接音場型マイクロホンアレイ”，日本音響学会誌，53巻2号，pp.110-116,1997
- [2] Y. Kaneda and J. Ohga, "Adaptive microphone array system for noise reduction," IEEE Trans. Acoust. Speech Source Process, Vol.ASSP-34, No.6, pp.1391-1400, 1986.
- [3] A. J. Bell and T. J. Sejowski, "An information maximization approach to blind separation and blind deconvolution," Neural Comput, vol.7, pp.1129-1159, 1995.
- [4] N. Takahashi, M. Matsumoto and S. Hashimoto, "Noise reduction combining microphone and piezoelectric pickup", ICA 2007,CD-Proc., MUS-07-007,2007