独立成分分析による残響環境下での混合音声の分離

半田 晶寛† 堤 憲亮† Leandro Di Persia‡ 柳田 益造†

†同志社大学工学部 〒610-0394 京都府京田辺市多々羅都谷 1-3

‡Universidad Nacional de Entre Rios Argentina

E-mail: †{dtc0733, bta0014}@mail4.doshisha.ac.jp, myanagid@mail.doshisha.ac.jp ‡ldipersia@hotmail.com

1. はじめに

現在の音声認識システムは，マイクに接近した位置からの音声に対してはそれなりに高い認識精度を有する一方で，マイクから離れた位置からの音声に対しては，周囲の雑音や部屋の残響の影響を受けて認識精度は著しく低下してしまう[1]. 実環境で音声認識を行うためには音声認識システムに入力される音声に対して何らかの処理を行うことが考えられる. その一つがブラインド音源分離（BSS: Blind Source Separation）である.

本稿では防音室やリビングルームにおいて，様々に条件を変えて音を収録し，その収録音に対して ICA を用いた周波数領域 BSS を行い，音声認識システム Julius での認識率の改善でその動作を評価する.

2. 研究の目的

ICA を用いた BSS は, 線形混合に対しては有効な手法であるといえるが，残響のある実環境において効果を挙げるには到っていない.

そこで本研究では 2 音源に対する周波数領域 BSS の実環境での有効性検証を目的として，残響時間の異なる収録場所，音源（スピーカ）と観測点（マイクロフォン）との位置関係，目的音源と雑音のパワー比，雑音の種類，を変えて収録した音に対して，JADE[2，3]の結果を初期値に用いた FastICA[4]による周波数領域 BSS を行い, 音声認識率の向上効果を検証する.

3. 分離実験

収録場所は防音室(490×400×290(H) cm^3)及びリビングルーム(380×230×218(H) cm^3)の 2 ヶ所である. 音を収録した環境(防音室)を図 1 に示す. 正面壁からスピーカまでの距離が 100 ㎝, 横壁からマイクａまでの距離が 200 ㎝である. リビングルームにおけるマイク･スピーカの配置は防音室と同じとした. また防音室に関しては標準状態に加えて，壁の 1 面(図 1 中の横壁)に反射板を設置した環境でも収録を行った. 残響時間は防音室の標準状態，反射板 1 面，リビングに関してそれぞれ白色ノイズで 130ms，150ms，440ms である. 信号の観測には，素子間隔 5 ㎝の 2 素子アレイ(小野測器 M1-1233)を使用した. 音源信号としては，防音室で録音し，Julius での認識率が 100%の男女各 20 フレーズ, 計 40 フレーズの音声を使用した. 雑音はコンピュータの冷却ファンノイズ，テレビ CM ノイズ，目的音声の話者から見て異性によるスピーチノイズの 3 種類である. スピーカの出力比(設定 S/N)は 0dB と 6dB の 2 種類とした.

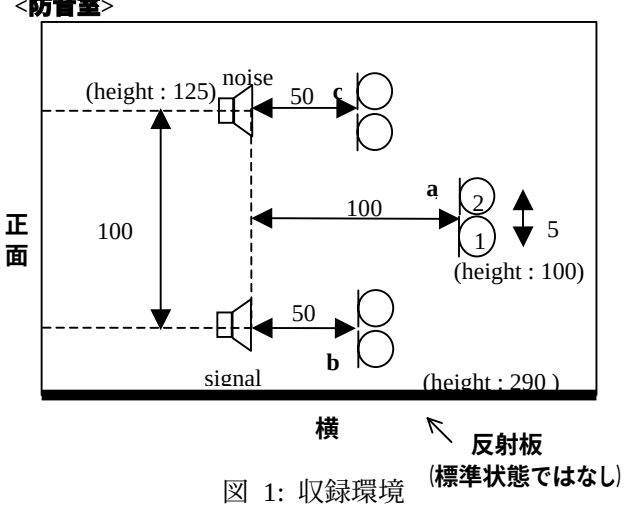

図 1: 収録環境

4. S/N の計算

受音段階及び分離後の S/N は，暫定的に以下の式により計算している.

$$SNR = 10 \log_{10} \frac{\sum_n s_n^2}{\sum_n (s_n - \alpha y_n)^2} [\mathrm{dB}]$$

ここで，s_n は音源信号，y_n は評価対象信号の各標本値である. またαは振巾に関する未定係数であり，分母が最小となるような値を最急降下法で求めた. ただし，上式を用いて実環境で収録した音の S/N の計算をするためには,音源信号と，評価対象信号のサンプル点を正確に同期させるか，もしくは正確な補間値を求める必要があり，これらは共に非常に困難である. また振幅に関してもαを用いて調整しているが，信号と雑音の間に相関があると正確な値が計算できないなどの問題がある. このようなことから，本稿での S/N の値はあくまでも参考値である.

「Separation of Mixed Speech in Reverberant Environment Using Independent Component Analysis」

Akihiro HANDA†, Kensuke TSUTSUMI†
Leandro Di Persia‡ and Masuzo YANAGIDA†
†Faculty of Engineering, Doshisha University
‡Universidad Nacional de Entre Rois, Argentina

5. 実験結果と考察

図 2 は受音段階と分離音の認識率, 図 3 は S/N の比較である．この図での値は，3 種類の雑音について平均した値である．図 2，3 の横軸の c0 は図 1 における c の位置にあるマイクロフォンで，スピーカ出力比が 0dB の音を収録した状態であること表す．他も同様である．

収録環境別の認識率の改善に関する *t* 検定，収録環境別の S/N に関する *t* 検定の結果を表 1 の上下段に示す．*，**はそれぞれ $p<0.05$，$p<0.01$ で有意水準であることを示す．なお，表内の数値は分離前後の認識率[%]，S/N[dB]の改善であり，上述した有意差が確認できれば，数値の前部のセルに「*」記号を記入した．

5．1　認識率の改善

標準状態の防音室においては表中の有意差から，認識率が改善されたと言える．反射板を用いた場合では，b6 以外で悪くとも有意水準 5%の有意差が得られたことから，概ね改善されたと言えよう．しかし，リビングルームでは有意水準 5%でも全ての状態で有意差がなく，認識率に関して改善の効果があるとは言えない．

またマイクロフォンの位置とスピーカ出力比の違いによる比較では，防音室に関しては，表 1 よりほぼすべの状態で認識率に対して改善の効果が見られると言える．ただし，b6 のようにもともと S/N が良く，認識率の高いものに対しては，改善が見られるとは言い難い．

5．2　S/N の改善

受音段階の S/N が悪い時ほど改善の割合が大きくなっている．表 2 の *t* 検定の結果で示されているように，受音段階の S/N が高い場合はどの収録環境に関しても有意水準 5%でも有意差がなく，S/N が改善されるとは言えない．

表 1: 収録環境毎の認識率の向上(上段)と S/N の改善(下段)に関する *t* 検定

[%] [dB]	防音室		防音室 （反射板 1 面）		リビング ルーム	
c0	** **	0→8 0.4→2.8	* **	0→4 0.4→2.0	 **	0→0 0.2→0.4
a0	** **	0→16 2.0→3.9	* **	1→8) 1.9→3.4	 **	0→1 0.4→0.5
a6	** **	6→22 3.1→4.1	* **	6→17 3.0→3.7	 **	1→1 0.5→0.6
b0	** **	8→18 4.3→4.9	**	7→17 4.2→4.5	 **	2→1 0.7→0.8
b6	*	16→29 5.0→4.8		15→23 4.9→4.4		6→5 0.8→0.8

6. まとめ

収録環境の違いによる比較では，残響時間が 150ms 程度までなら音声認識率，S/N 共に改善の効果があることが確認された．一方で，リビングルームにおける分離処理に関しては，S/N が改善されても，認識率は改善されなかった．

また S/N と認識率の関係についても，全体としては S/N が改善された時には認識率も改善されており，一定の相関が見られる．しかし今後，より確な S/N の評価法を開発し，再び検討を行う必要がある．

7. 終わりに

様々に条件を変えた環境で収録した音に，独立成分分析による周波数領域 BSS の処理を行い, その音声認識への有効性を検証した．

その結果, 残響時間が 150ms 程度までなら効果が得られることが確認された．ただし，残響時間としては調査した対象が 150ms から 440ms に飛んでいるので，その間の値の残響時間での音声に対する認識率に関しても調べる必要がある

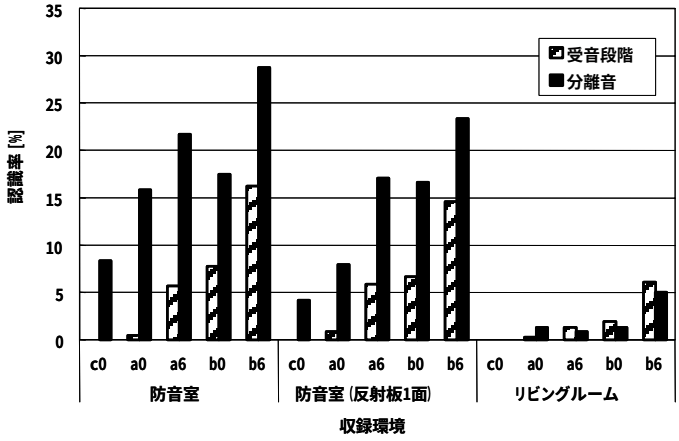

図 2: 収録環境毎の認識率の比較

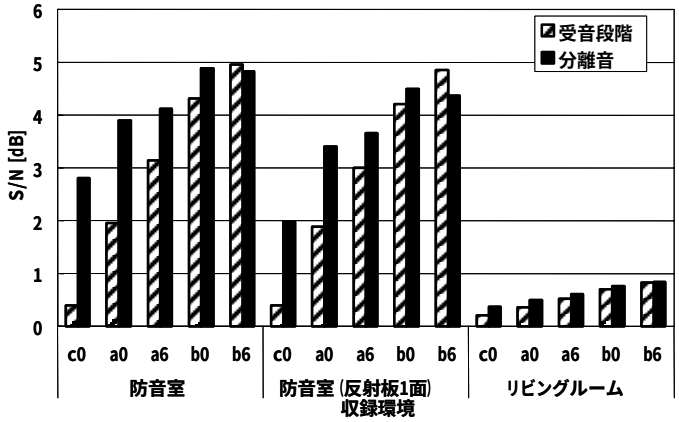

図 3: 収録環境毎の S/N の比較

文献

[1]中村哲, 実音響環境に頑健な音声認識を目指して, 電子情報通信学会 技術報告, SP2002-12, pp.31-36, 2002.
[2]甘利俊一, 村田昇. SGC ライブラリ 18:独立成分分析－多変量ﾃﾞｰﾀ解析の新しい方法,ｻｲｴﾝｽ社,東京, 2002.
[3]J.F.Cardoso, and A.Souloumiac. Blind beamforming for non Gaussian signals, IEEE Proceeding-F, vol. 140, no 6, pp.362-370, December 1993.
[4]横田康成.信号処理　~第 5 部　独立成分解析～, http://www.ykt.info.gifu-u.ac.jp/dsp/sp5.pdf