

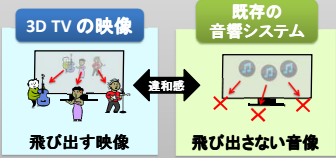
アクティベーション共有型非負値行列因子分解を用いた音像深度推定

◎宮内 智, 北村 大地, 猿渡 洋, 中村 哲 (奈良先端大・情報科学研究科)

1. 研究背景と目的

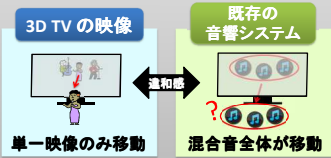
音像定位操作問題

- 3D テレビ等によって立体感のある奥行きを持たせた映像を表示する事が可能
- 既存の音響システムではディスプレイより前方に定位する音像の再現は困難
- 音像と映像の知覚上の位置が一致せず視聴者は違和感を感じてしまう



混合音源問題

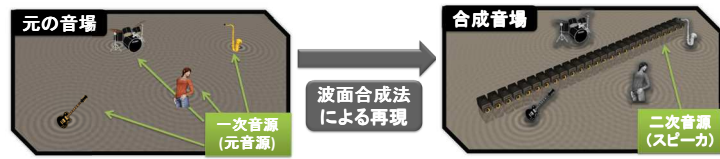
- 既存のコンテンツでは各音源が2chに混合されて収録されているため、音像ごとに操作することが出来ない
- ある特定の映像だけが移動した場合に、映像に合わせて特定の音源だけを操作する事ができない



本研究の最終目的 立体映像に適した音像を表示できる立体音響システムの構築

2. 波面合成法

- 一次音源により生じる一次音場を複数のスピーカアレイを用い再現する技術
- スピーカアレイよりも前方に音像を配置する(物理的な波面を合成する)ことも可能



駆動関数

$$S_{2n}(\omega) = S_P(\omega) \sqrt{\frac{\text{sign}(\zeta)}{2\pi j}} C(\zeta) \frac{\exp(\text{sign}(\zeta) j k r_{pn})}{\sqrt{r_{pn}}} \cos \theta_{pn} \Delta x$$

$$C(\zeta) = \sqrt{\frac{\zeta}{\zeta-1}}, \quad \zeta = \frac{y_P}{y_P}$$

ζ : 一次音源と参照受聴線のy座標の比

- 立体映像に一致する奥行きのある音像を表示することが可能となる
- × 音像の位置をユーザ側から指定する必要がある → 元音源の位置は分からない

本発表の目的 波面合成法で必要となる混合音源中の音像深度(奥行き)を推定する

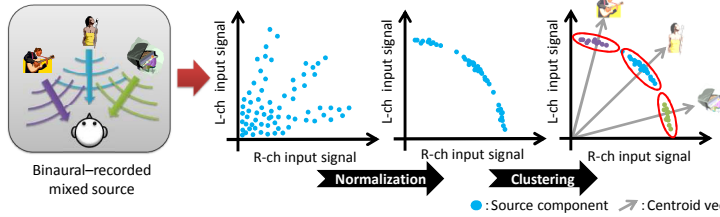
3. k-meansクラスタリングに基づく方位分解

- 各時間周波数グリッドにおいて知覚される方位は1つの方向のみ(第一波面の法則)
 - ある方向に知覚される音の塊をオーディオオブジェクトと呼び、分離結果はこれに相当する
- チャンネル間の音圧比を音源の方位情報とクラスタリングによって方位毎の音源へ分離する

$$Z(f, \tau) = Y(f, \tau) H(f, \psi)$$

$$H(f, \psi) = [H_1(f, \psi), H_2(f, \psi)]^T$$

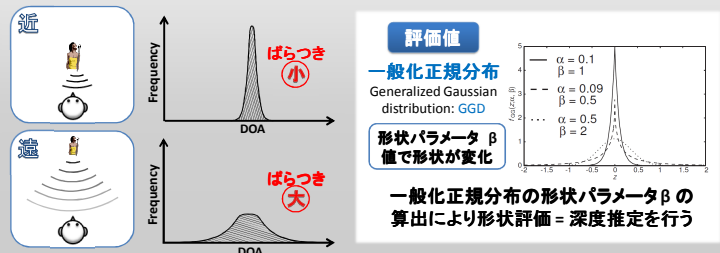
分離信号: Z , モノラル合成信号: Y , 空間情報: H , ψ : オブジェクト番号



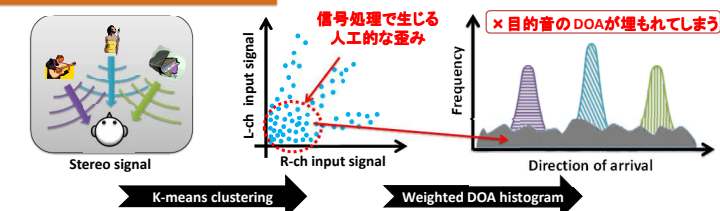
4. 提案手法の概要

提案手法 音の到来方向 (direction of arrival: DOA) 分布に基づく深度推定

音源の DOA でヒストグラムを作成すると音源の距離によって形状が変化する

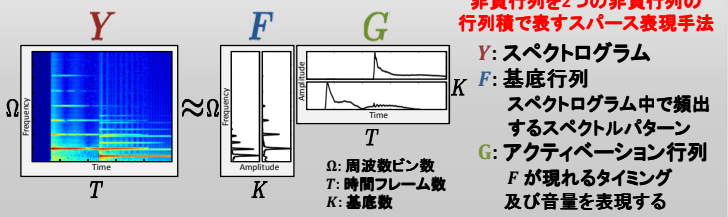


実際の信号処理における問題点



5. 提案手法における特徴量抽出

非負値行列因子分解 (nonnegative matrix factorization: NMF)



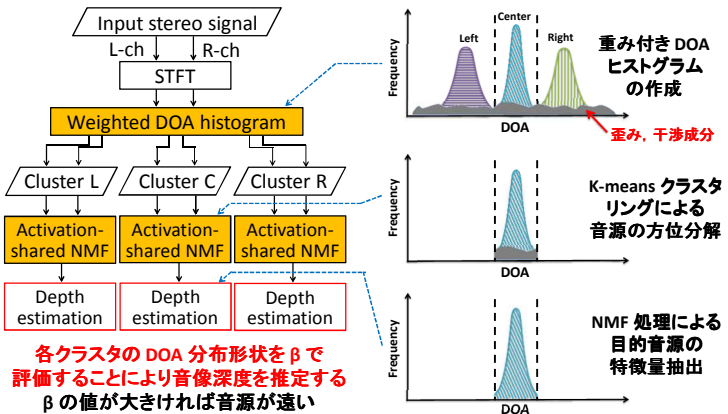
アクティベーション共有型 NMF

- スパース表現による次元圧縮
- 信号の方位情報を保ちつつ目的音の特徴量抽出を行う
- デジタル信号処理過程で生じる歪みや背景雑音の影響を取り除き目的音の DOA 分布を取得できる

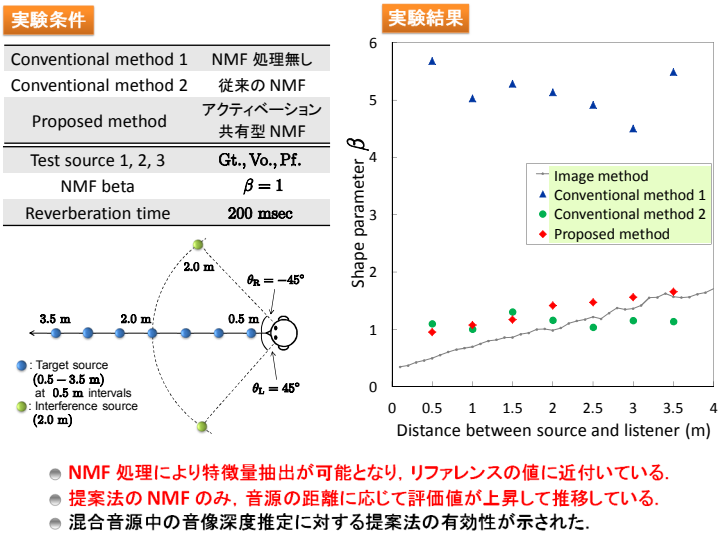
コスト関数 $\arg \min_{F^{(L)}, F^{(R)}, G} \mathcal{J} = D(Y^{(L)} | F^{(L)} G) + D(Y^{(R)} | F^{(R)} G)$

s.t. $f_{\omega, k}^{(L)}, f_{\omega, k}^{(R)}, g_{k, t} \geq 0$

6. 提案手法における信号のフロー



7. 実験条件および結果



8. まとめと今後の方針

- まとめ
- 波面合成法の再現には「音像深度」が必要だが、混合音源では情報が失われている。
 - 音像深度の推定を目的として、DOA 分布に基づく音像深度推定を提案した。
 - 信号処理問題の改良手法としてアクティベーション共有型 NMF を提案し適用した。
 - 実験から、音像深度推定に対する提案手法の有効性が示唆された。
- 今後の方針
- 他条件での追加実験の実施. 実音源との比較