音学シンポジウム2021 1日目(6/18)
ポスターセッション1 13:00-15:00

(10) [一般発表]

基底共有型非負値行列因子分解を用いた楽器音の音色変換

Timbre conversion of music instruments using basis-shared nonnegative matrix factorization

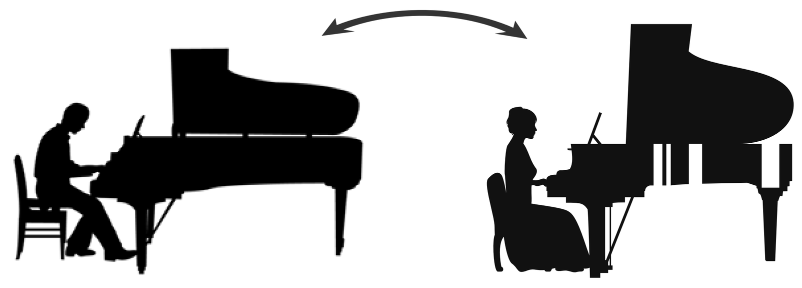
北村大地(香川高専)
香西海斗(香川高専)*

*現在，豊橋技術科学大学

研究背景

- 人は楽器音の音色の違いを感覚的に表現
  - 例:「きらびやかな音」,「やわらかい, 温かみのある音」
  - 定量的に表現する方法はあまり確立されていない

両者の違いは？



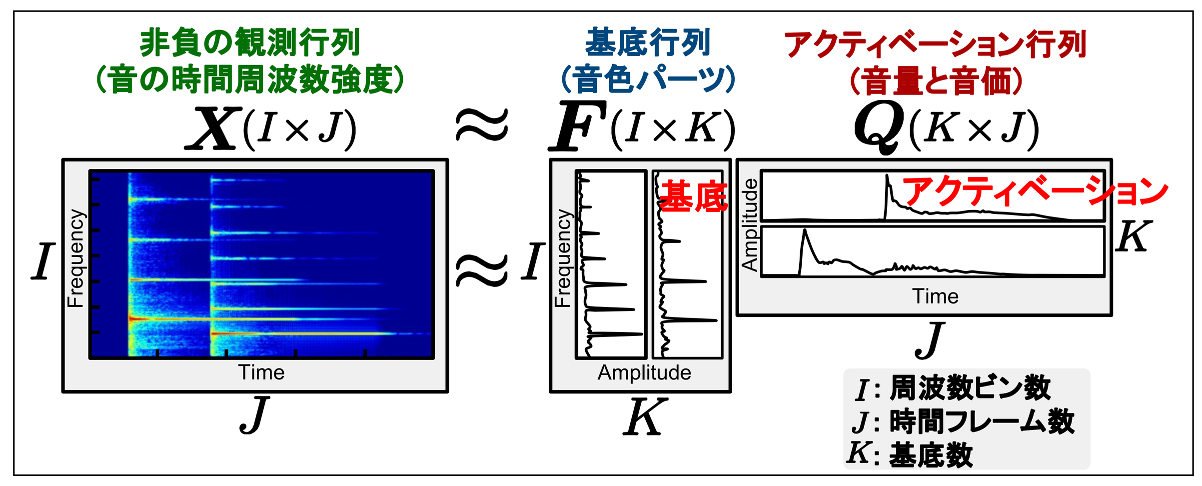
- 複数の楽器音の違いを定量的に表現
  - アマチュア奏者の上達の支援
  - 芸術的価値の高い楽器の設計製作
  - 音色変換，楽器音識別，音楽検索等にも応用可能

本発表の概要とNMFに基づく音響信号の分解

- 非負値行列因子分解（NMF）[Lee+, 1999] に基づく解析
  - 複数楽器音信号間に共通する音色成分
  - 各楽器音信号の固有の音色成分

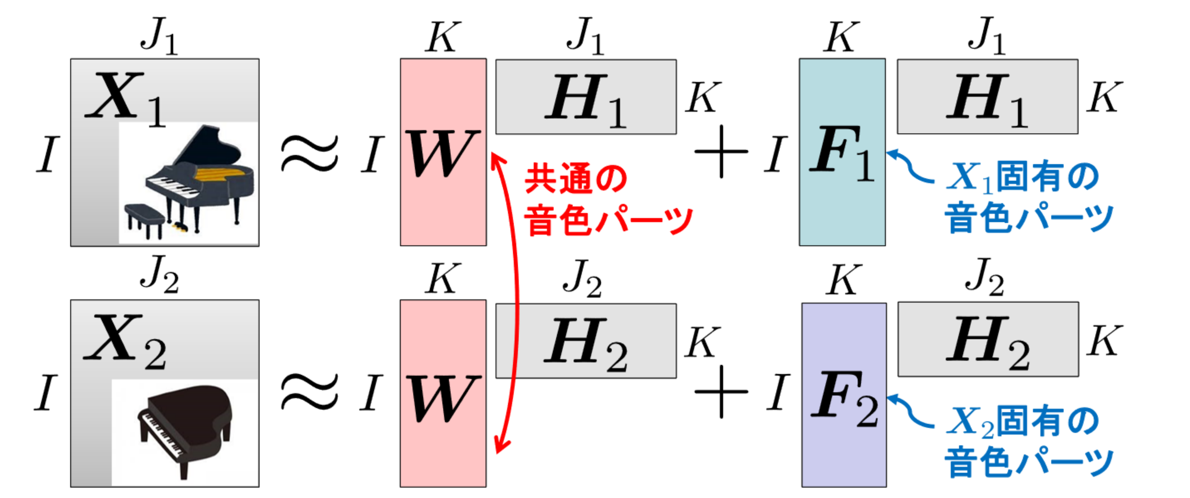
  この2成分を同時にNMFで抽出

- 音響信号におけるNMFの適用
  - 振幅スペクトログラム$\boldsymbol{X}$を基底$\boldsymbol{F}$とアクティベーション$\boldsymbol{Q}$に分解



基底共有型NMF（BSNMF）[香西ら, 2021]

- 音源$\boldsymbol{X}_1$と$\boldsymbol{X}_2$に対して共通の基底行列$\boldsymbol{W}$と固有基底行列$\boldsymbol{F}_1$と$\boldsymbol{F}_2$を用いて分解



- $\boldsymbol{W}$は2つの楽器音の共通の音色（スペクトル）
- $\boldsymbol{F}_1$と$\boldsymbol{F}_2$は各楽器の固有の音色（スペクトル）

BSNMFの最適化問題

$$\min_{\boldsymbol{W},\boldsymbol{F}_1,\boldsymbol{F}_2,\boldsymbol{H}_1,\boldsymbol{H}_2} \mathcal{D}(\boldsymbol{X}_1|\boldsymbol{W}\boldsymbol{H}_1+\boldsymbol{F}_1\boldsymbol{H}_1)+\mathcal{D}(\boldsymbol{X}_2|\boldsymbol{W}\boldsymbol{H}_2+\boldsymbol{F}_2\boldsymbol{H}_2)$$

$\boldsymbol{X}_1$と$\boldsymbol{W}\boldsymbol{H}_1+\boldsymbol{F}_1\boldsymbol{H}_1$の距離　$\boldsymbol{X}_2$と$\boldsymbol{W}\boldsymbol{H}_2+\boldsymbol{F}_2\boldsymbol{H}_2$の距離

両項を同時に小さくするような変数行列を推定

距離関数

$$\sum_n \mathcal{D}_{KL}(\boldsymbol{X}_n|\boldsymbol{W}\boldsymbol{H}_n+\boldsymbol{F}_n\boldsymbol{H}_n) = \sum_{n,i,j_n}\left[x_{ij_nn}\log x_{ij_nn} - x_{ij_nn}\log\left(\sum_k w_{ik}h_{kj_nn}+\sum_k f_{ikn}h_{kj_nn}\right) - x_{ij_nn}+\sum_k w_{ik}h_{kj_nn}+\sum_k f_{ikn}h_{kj_nn}\right]$$

[香西ら, 2021]

$$\sum_n \mathcal{D}_{EU}(\boldsymbol{X}_n|\boldsymbol{W}\boldsymbol{H}_n+\boldsymbol{F}_n\boldsymbol{H}_n) = \sum_{n,i,j_n}\left[x_{ij_nn} - \left(\sum_k w_{ik}h_{kj_nn}+\sum_k f_{ikn}h_{kj_nn}\right)\right]^2$$

- 最適化には補助関数法 [Hunter+ et al., 2004] を用いる

二乗Euclid距離に基づくBSNMF

- 最適化関数

$$\mathcal{J} \equiv \sum_{n=1}^{N}\sum_{i=1}^{I}\sum_{j_n=1}^{J_n}\left[x_{ij_nn} - \left(\sum_{k=1}^{K} w_{ik}h_{kj_nn}+\sum_{k=1}^{K} f_{ikn}h_{kj_nn}\right)\right]^2$$
$$= \sum_{n=1}^{N}\sum_{i=1}^{I}\sum_{j_n=1}^{J_n}\left[x_{ij_nn}^2 + \boxed{\left(\sum_{k=1}^{K} w_{ik}h_{kj_nn}+\sum_{k=1}^{K} f_{ikn}h_{kj_nn}\right)^2} - 2x_{ij_nn}\left(\sum_{k=1}^{K} w_{ik}h_{kj_nn}+\sum_{k=1}^{K} f_{ikn}h_{kj_nn}\right)\right]$$

補助関数を設計

最適化更新式導出時に問題となる項

$$\mathcal{J} \le \mathcal{J}^+ = \sum_{n=1}^{N}\sum_{i=1}^{I}\sum_{j_n=1}^{J_n}\left[x_{ij_nn}^2 + \sum_{k=1}^{K}\frac{w_{ik}^2h_{kj_nn}^2}{\alpha_{ij_nkn}} + \sum_{k=1}^{K}\frac{f_{ikn}^2h_{kj_nn}^2}{\beta_{ij_nkn}} - 2x_{ij_nn}\left(\sum_{k=1}^{K} w_{ik}h_{kj_nn}+\sum_{k=1}^{K} f_{ikn}h_{kj_nn}\right)\right]$$

- α_{ij_nkn}, β_{ij_nkn}は補助変数

二乗Euclid距離に基づくBSNMF

- 導出される更新式

$$\boldsymbol{W}^+ \leftarrow \boldsymbol{W} \otimes \frac{\boldsymbol{X}_1\boldsymbol{H}_1^{\mathrm{T}}+\boldsymbol{X}_2\boldsymbol{H}_2^{\mathrm{T}}}{\boldsymbol{W}\boldsymbol{H}_1\boldsymbol{H}_1^{\mathrm{T}}+\boldsymbol{F}_1\boldsymbol{G}_1\boldsymbol{H}_1^{\mathrm{T}}+\boldsymbol{W}\boldsymbol{H}_2\boldsymbol{H}_2^{\mathrm{T}}+\boldsymbol{F}_2\boldsymbol{G}_2\boldsymbol{H}_2^{\mathrm{T}}}$$

$$\boldsymbol{F}_1^+ \leftarrow \boldsymbol{F}_1 \otimes \frac{\boldsymbol{X}_1\boldsymbol{G}_1^{\mathrm{T}}}{\boldsymbol{F}_1\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{T}}+\boldsymbol{W}\boldsymbol{H}_1\boldsymbol{G}_1^{\mathrm{T}}} \qquad \boldsymbol{F}_2^+ \leftarrow \boldsymbol{F}_2 \otimes \frac{\boldsymbol{X}_2\boldsymbol{G}_2^{\mathrm{T}}}{\boldsymbol{F}_2\boldsymbol{G}_2\boldsymbol{G}_2^{\mathrm{T}}+\boldsymbol{W}\boldsymbol{H}_2\boldsymbol{G}_2^{\mathrm{T}}}$$

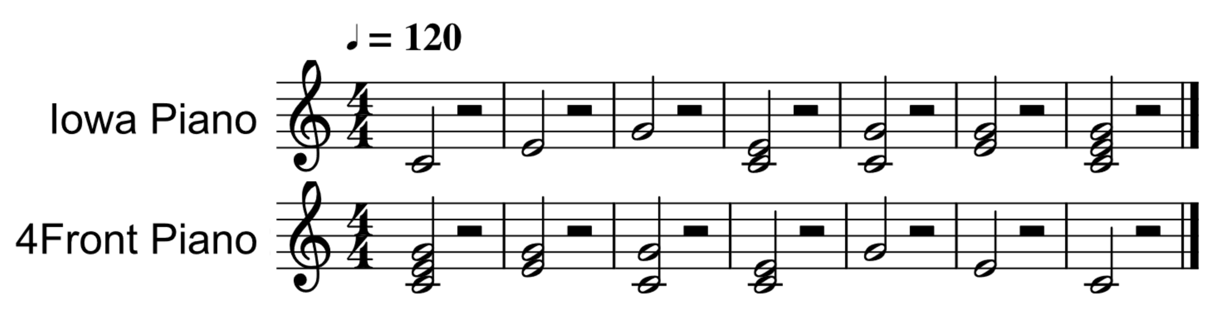
$$\boldsymbol{H}_1^+ \leftarrow \boldsymbol{H}_1 \otimes \frac{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}_1}{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}_1+\boldsymbol{W}^{\mathrm{T}}\boldsymbol{F}_1\boldsymbol{G}_1} \qquad \boldsymbol{H}_2^+ \leftarrow \boldsymbol{H}_2 \otimes \frac{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}_2}{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H}_2+\boldsymbol{W}^{\mathrm{T}}\boldsymbol{F}_2\boldsymbol{G}_2}$$

$\otimes$ はアダマール積（行列の要素ごとの乗算）
分数は$\oslash$ともされるアダマール除算（行列の要素ごとの除算）
T は行列の転置

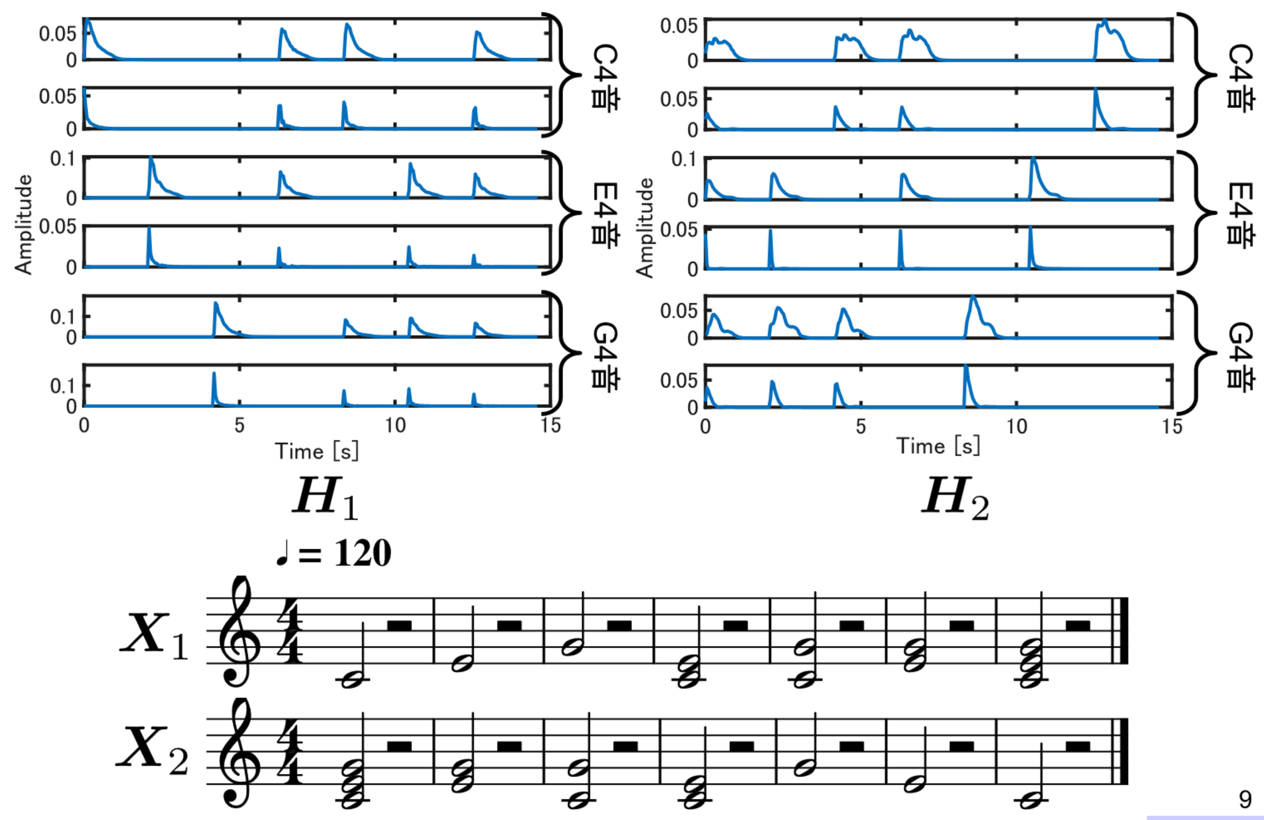
音響特徴量抽出実験：条件

- 各実験条件

	音源1	音源2
VSTプラグイン	Iowa Piano	4Front Piano
STFTの窓長	92.9 ms	
STFTのシフト長	46.4 ms	
窓関数	Hamming窓	
乖離度関数	一般化KLダイバージェンス	
反復回数	1000 回	
基底数	6 本	

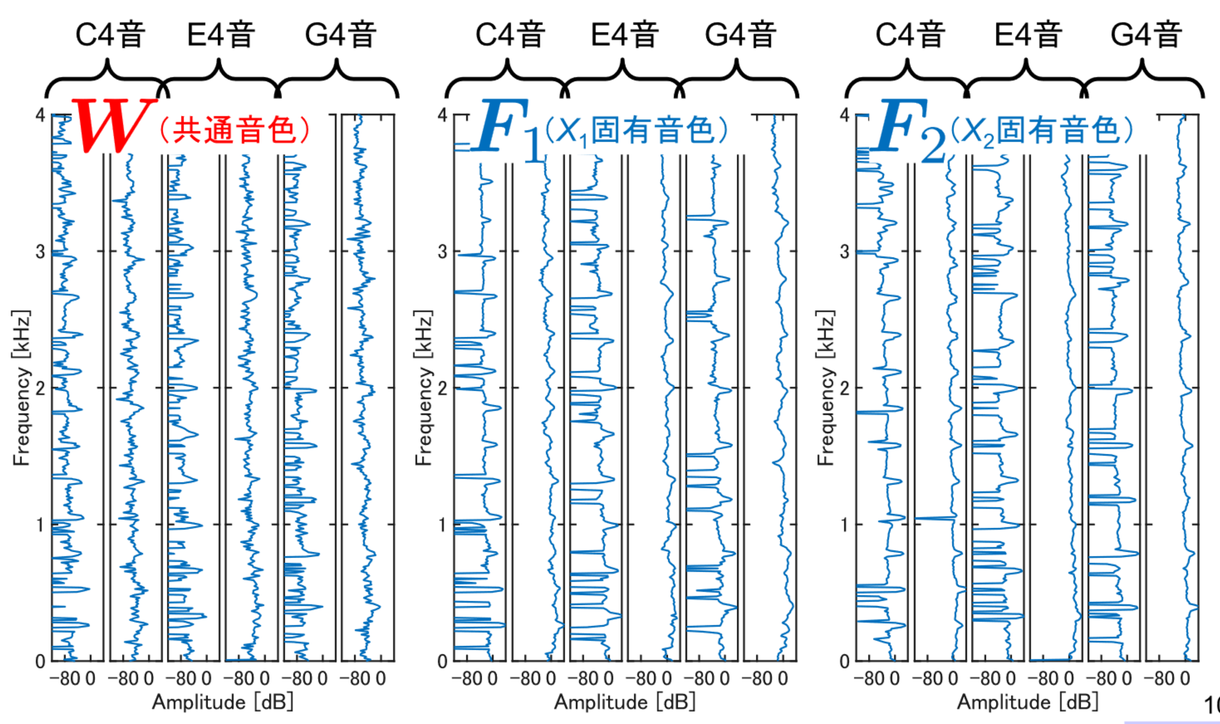


音響特徴量抽出実験：結果



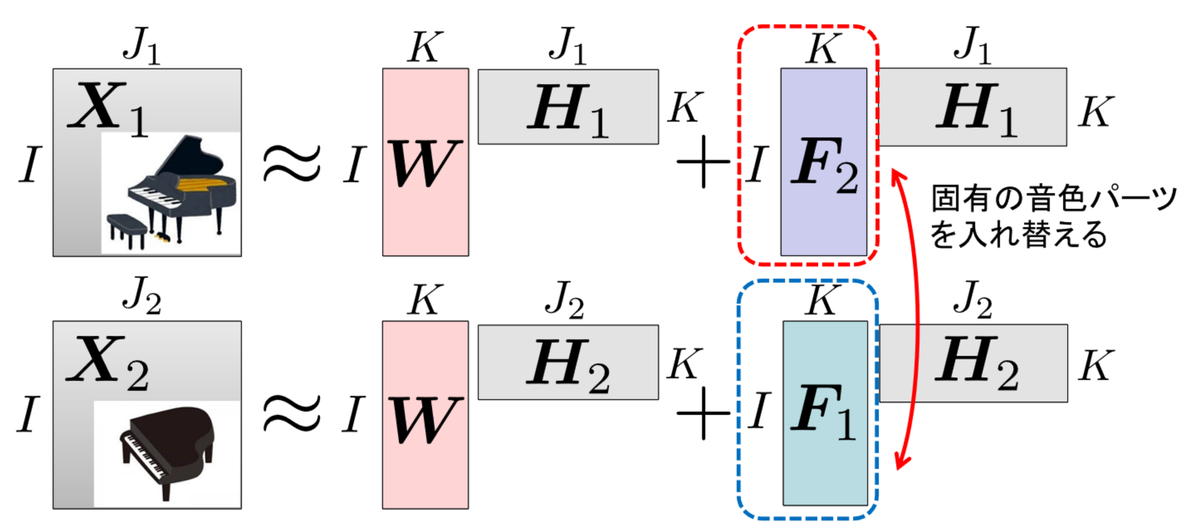
音響特徴量抽出実験：結果

- 打撃部分に明確な差異



音色変換への応用：方法

- 固有基底行列$\boldsymbol{F}_1$及び$\boldsymbol{F}_2$のみを入れ替えることで，音色の変換を実現



音色変換への応用：スケールフィッティング

- $\boldsymbol{W}\boldsymbol{H}_n$と$\boldsymbol{F}_n\boldsymbol{H}_n$のパワー比は楽器音$n$によって異なる
  - $\boldsymbol{W}$と$\boldsymbol{H}_n$の間にはスケールの任意性が存在するため
  - 単に固有基底行列$\boldsymbol{F}_1$及び$\boldsymbol{F}_2$を入れ替えただけでは音が歪む

固有基底行列の列ベクトルをスカラー倍できる変数を導入（対角行列との積）

対角行列変数

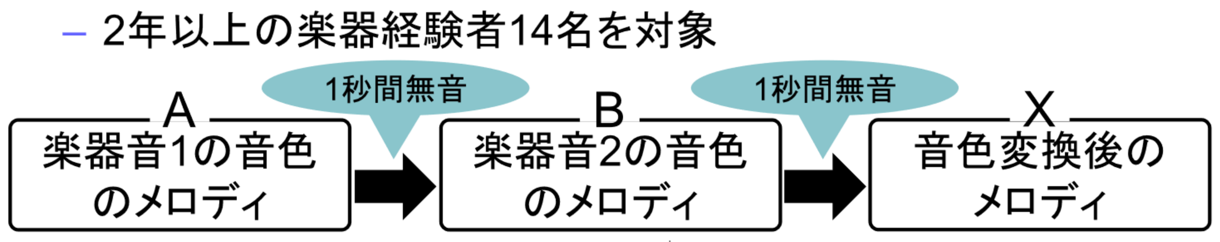
$$\begin{cases}\boldsymbol{W}\boldsymbol{H}_1+\boldsymbol{F}_2\boldsymbol{H}_1\\ \boldsymbol{W}\boldsymbol{H}_2+\boldsymbol{F}_1\boldsymbol{H}_2\end{cases} \Rightarrow \begin{cases}\boldsymbol{W}\boldsymbol{H}_1+(\boldsymbol{F}_2\boldsymbol{D}_1)\boldsymbol{H}_1\\ \boldsymbol{W}\boldsymbol{H}_2+(\boldsymbol{F}_1\boldsymbol{D}_2)\boldsymbol{H}_2\end{cases}$$

スカラー倍をスケールフィッティングにより最適化

$$\min_{\boldsymbol{D}_1,\boldsymbol{D}_2} \mathcal{D}(\boldsymbol{X}_1|\boldsymbol{W}\boldsymbol{H}_1+\boldsymbol{F}_2\boldsymbol{D}_1\boldsymbol{H}_1)+\mathcal{D}(\boldsymbol{X}_2|\boldsymbol{W}\boldsymbol{H}_2+\boldsymbol{F}_1\boldsymbol{D}_2\boldsymbol{H}_2)$$

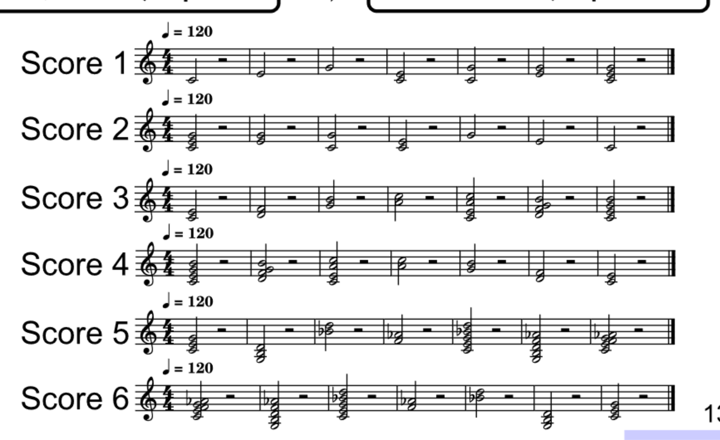
音色変換への応用：実験条件

- ABX法を用いて音色変換の精度を主観評価
  - Xは「X_1の音色をX_2に変換したもの」または「X_2の音色をX_1に変換したもの」のいずれかをランダムに提示
  - 2年以上の楽器経験者14名を対象



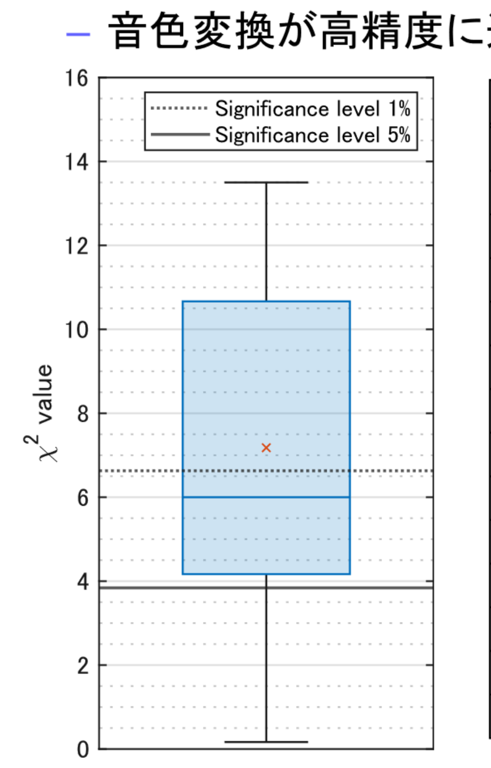
1. **Iowa Piano**
（グランドピアノVST）
2. **Sketch Upright Piano**
（アップライトピアノVST）

の2種類のピアノ音源間で音色を変換



音色変換への応用：実験結果

- 有意水準5%で帰無仮説「AとBには差がない」を棄却可
  - 音色変換が高精度に達成されていることを裏付けている



ID	年代	性別	正答率 [%]	カイ二乗値
被験者1	20代	女性	70.83	4.17
被験者2	30代	女性	66.67	2.67
被験者3	30代	女性	75.00	6.00
被験者4	20代	男性	70.83	4.17
被験者5	20代	男性	87.50	13.50
被験者6	10代	男性	83.33	10.67
被験者7	10代	男性	83.33	10.67
被験者8	10代	男性	70.83	4.17
被験者9	10代	女性	75.00	6.00
被験者10	10代	女性	54.17	0.17
被験者11	10代	男性	79.17	8.17
被験者12	20代	女性	75.00	6.00
被験者13	20代	男性	87.50	13.50
被験者14	20代	女性	83.33	10.67

デモンストレーション

- 音色変換の結果例

変換元 Sketch Upright Piano	変換先 Iowa Piano	音色変換結果 (Sketch → Iowa)

変換元 Iowa Piano	変換先 Sketch Upright Piano	音色変換結果 (Iowa → Sketch)