

高品質声質変換のための特徴量分析再訪

2-Q-27 ☆ 須田仁志 △ 小谷岳 高道慎之介 齋藤大輔 (東京大学)

本研究の背景と目的

- ▶ 声質変換における「分析」と「合成」
 - ▷ 変換モデルに比べ、詳細に検討されることは少ない
- ▶ 分析合成時のパラメータを主観評価ベースでチューニングする
 - ▷ WORLD 分析 + GMM 変換 + 差分フィルタ合成の限界へ

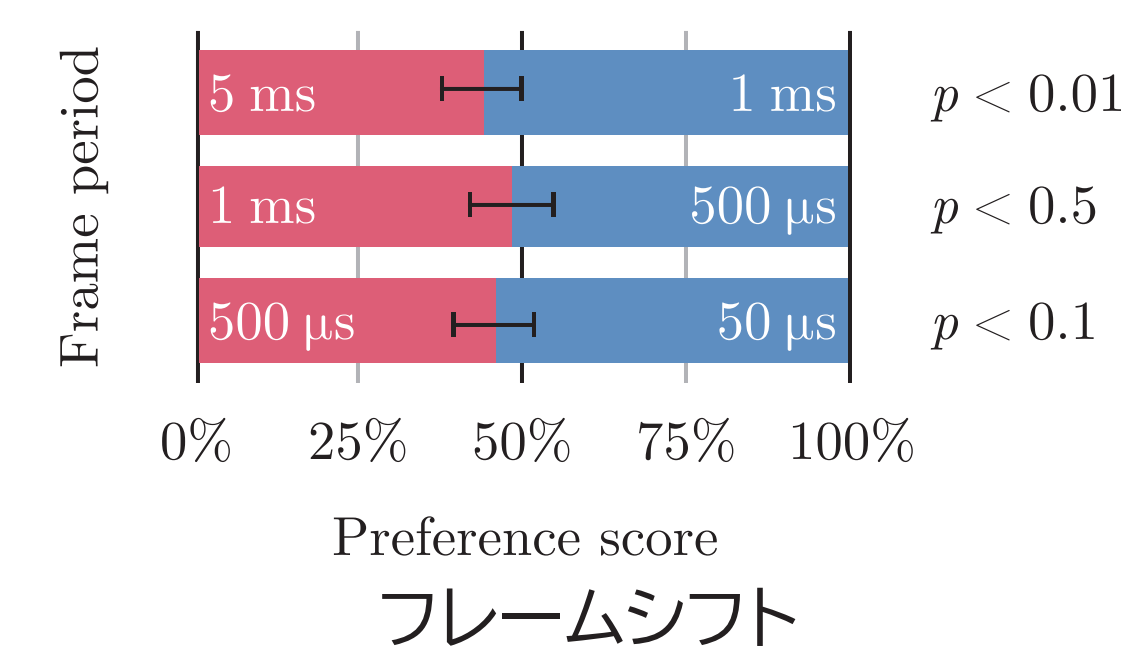
実験設定と条件

- ▶ 手順ごとに、それぞれの最適パラメータを検討する
 - ▷ 分析・再合成のみ（変換なし）
 - ▷ 変換前後のパラメータが既知の場合（理想条件での変換）
 - ▷ GMM による変換を行う場合（一般的な条件での変換）
- ▶ 各実験において自然性を AB テストにより評価
 - ▷ 3 つめのみ、くわえて話者性を ABX テストにより評価
- ▶ データセットとして ATR503 文のサブセット A（50 文）
 - ▷ 男女 2 名ずつ、サンプリング周波数は 22050 Hz
- ▶ 被験者はそれぞれの実験で 23 名、各被験者は 10 問回答

分析・再合成

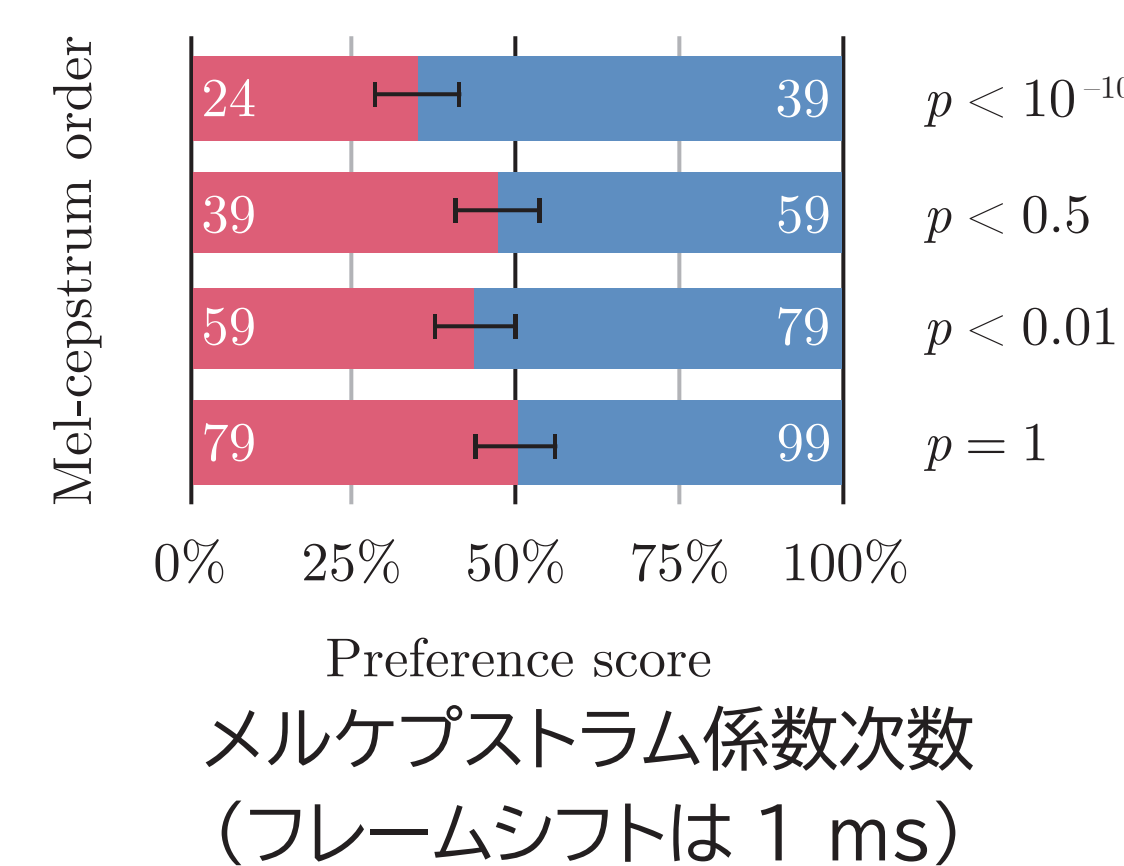
フレームシフト

- ▶ スペクトル包絡は次元圧縮しない
- ▶ $5\text{ ms} < 1\text{ ms} \leq$ それ以下
 - ▷ およそ短いほどよい
 - ▷ WORLD の F_0 分析の限界が 1 ms のためか



メルケプストラム係数次数

- ▶ スペクトル包絡をメルケプストラム係数に次元圧縮して合成
- ▶ $24 < 39 \leq 59 < 79 \leq 99$
 - ▷ 1 pt 分析では 39 で頭打ち
 - ▷ 品質の限界はあまり高い次数でない



ここでの 50 μs は 1 pt (およそ 45 μs) を表す

提案法：SP-WORLD

差分フィルタ補正による合成 [K.Kobayashi+, 2014]

- ▶ 入出力スペクトル特徴量の差分からフィルタを推定し、入力音声に畳み込むことで変換音声を作成
- ▶ 一般的には MLSA フィルタを用いる
 - ▷ メルケプストラム係数から Padé 近似を用い係数を推定

SP-WORLD

- ▶ 実ケプストラムを用いた最小位相復元法 [S.-C.Pei+, 2005] によりパワースペクトルから直接インパルス応答を生成
- ▷ WORLD [M.Morise+, 2016] の合成部と同じ手法

アフィン DTW (動的時間伸縮)

- ▶ DTW と出力話者特徴量のアフィン変換を交互に行うことで話者性の違いによる DTW の誤りを低減する
 - ▷ 1 混合 GMM での声質変換を行いつつ DTW するのと同義

理想条件下での変換

- ▶ 変換モデルを用いず既知の平行データを利用する

合成法

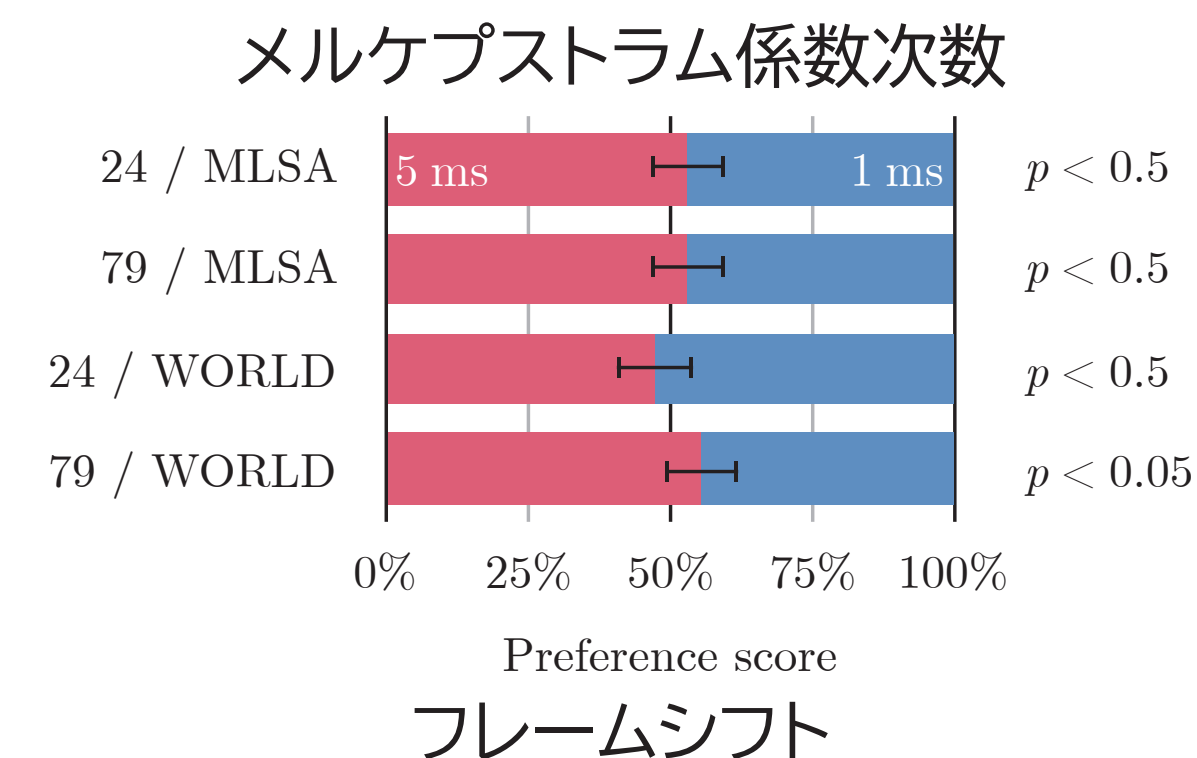
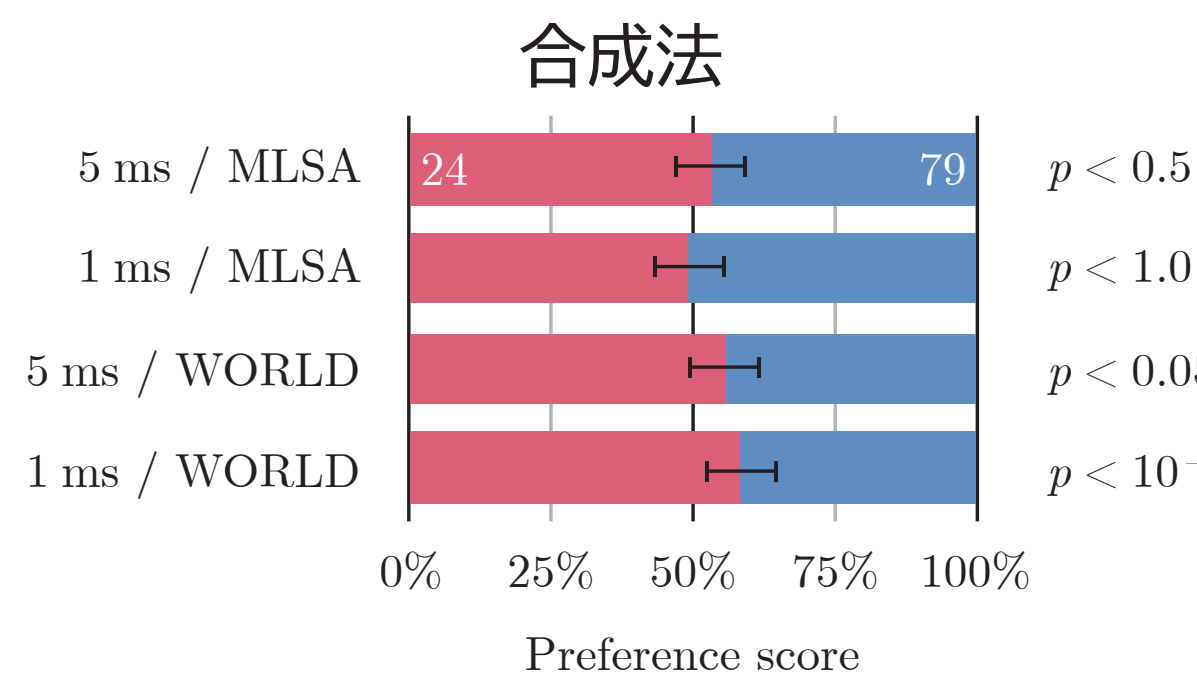
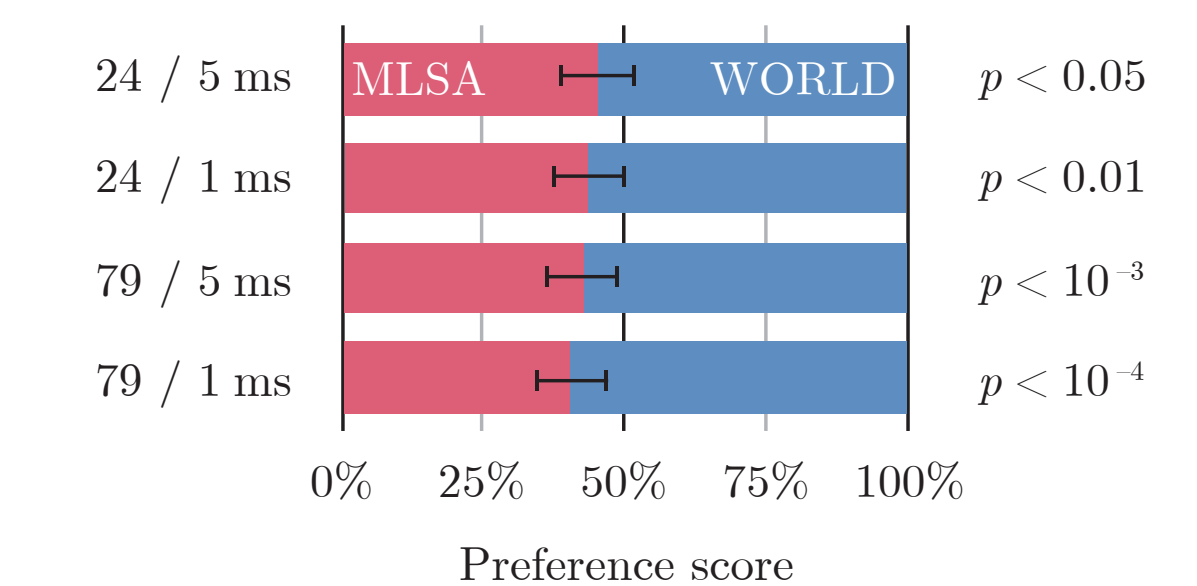
- ▶ ほぼ $\text{MLSA} < \text{SP-WORLD}$
 - ▷ 提案手法の有効性

メルケプストラム係数次数

- ▶ SP-WORLD では $24 > 79$
 - ▷ 高次では DTW のミスが強く現れてしまうためか

フレームシフト

- ▶ 有意差はない
- ▶ SP-WORLD/24 次では 1 ms のほうがわずかに優れる



差分フィルタ補正における逆フィルタ近似精度

- ▶ 入力音声から残差波形を取り出すフィルタに相当



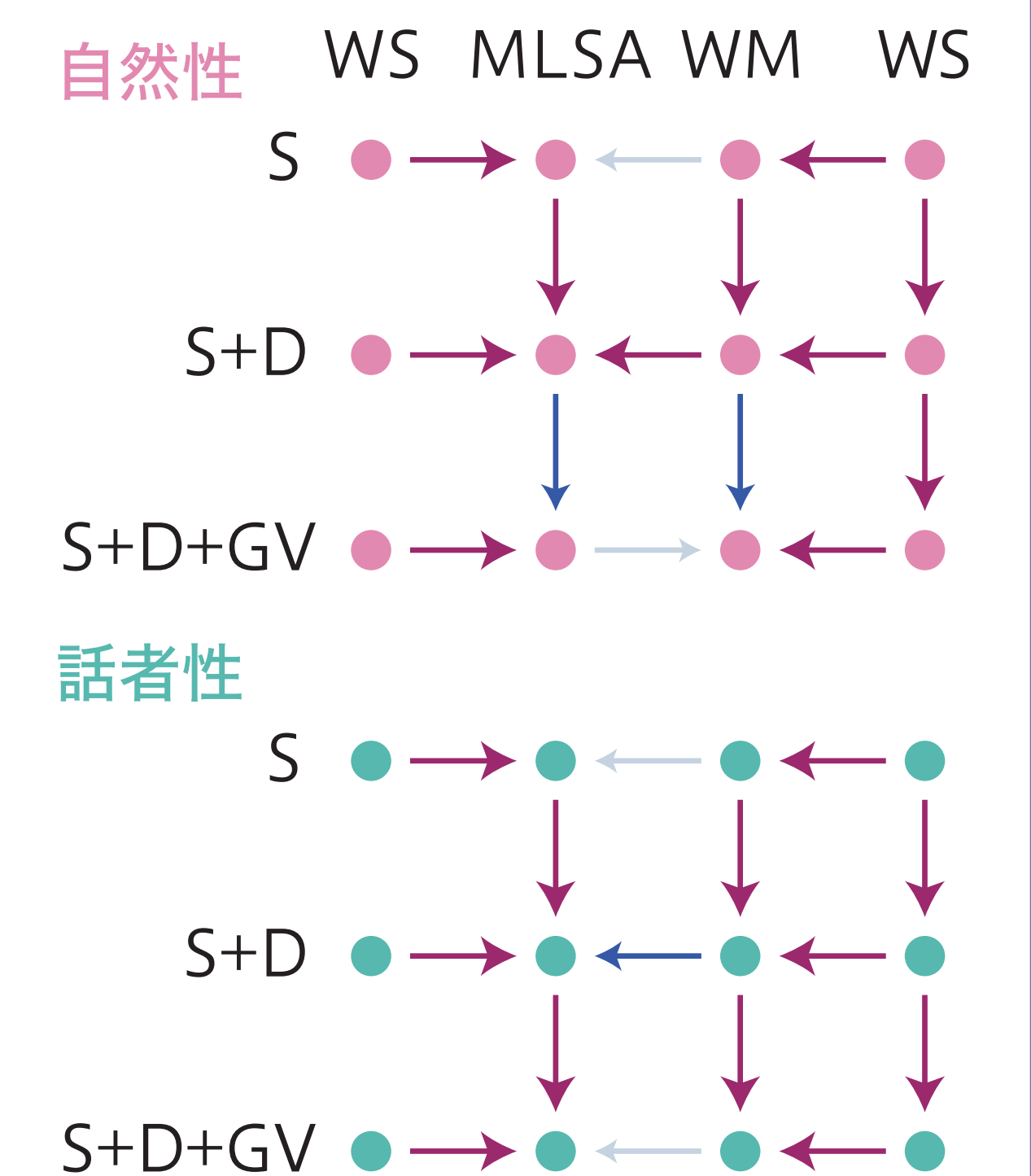
- ▶ MLSA では出力音声と同じ次数のメルケプストラム係数
- ▶ SP-WORLD では分析時の非圧縮スペクトル包絡が使える
 - ▷ ここでは、メルケプストラム係数を用いたものを WM、スペクトル包絡を用いたものを WS とよぶ

GMM を用いた声質変換

- ▶ 動的特徴量・系列内変動 [T.Toda+, 2007] の効果も検討
- ▶ GMM の混合数は、静的特徴量のみを用いる場合は 64、静的動的特徴量を用いる場合は 128
- ▶ フレームシフトは 1 ms、メルケプストラム係数は 24 次

合成法

- ▶ MLSA と WM は話者性・自然性ともに同等の性能
- ▶ WS は他 2 手法に劣る
 - ▷ 逆フィルタが緻密で高次の成分が抜け落ちるためか
 - ▷ ボコーダの音声に近くなる



動的特徴量と系列内変動

- ▶ どの合成法においても有効

まとめと今後の展望

- ▶ 提案合成法（SP-WORLD）は特定条件下で有効
 - ▷ GMM 変換においても既存手法（MLSA）と同等の性能
- ▶ 1 ms 以下分析の実装や GMM 声質変換におけるメルケプストラム係数の次数などの影響についての検討

謝辞

本研究の一部はJSPS科研費 16H06681 およびセコム科学技術支援財団の助成を受け実施した。