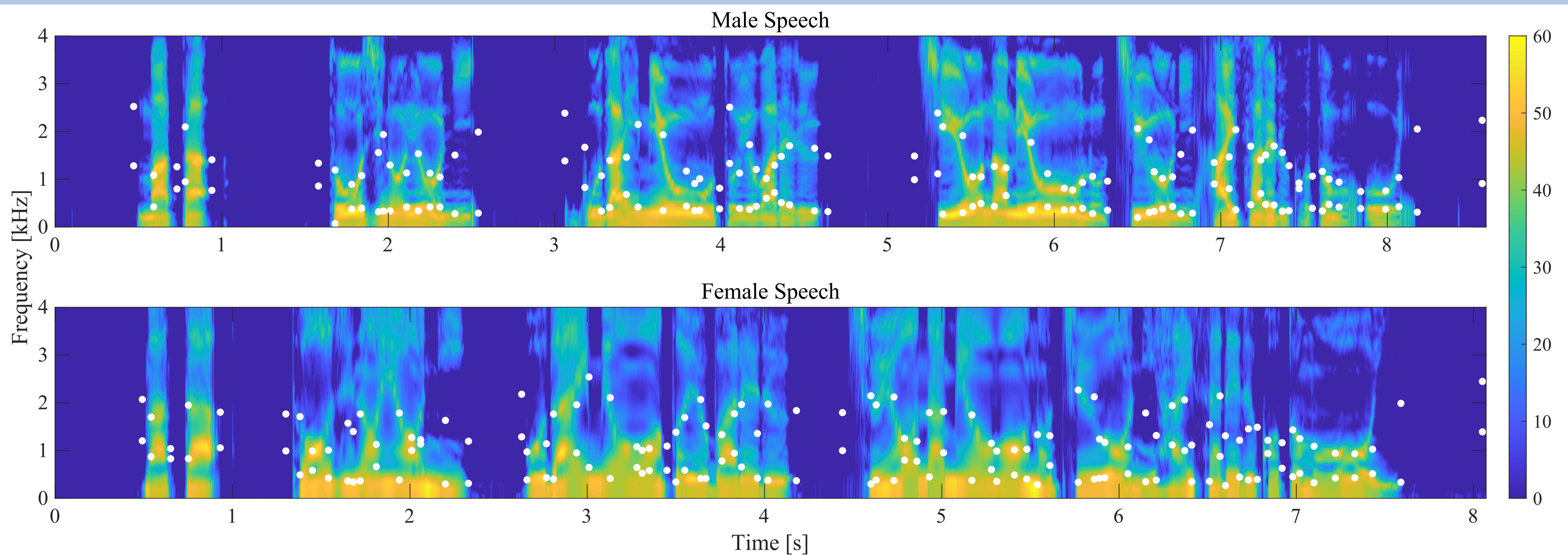


自動推定された音素境界とフォルマントによる 音声モーフィングの品質評価

☆堀部貴紀¹, 森勢将雅¹, 河原英紀²

1: 明治大学総合数理学部先端メディアサイエンス学科, 2: 和歌山大学



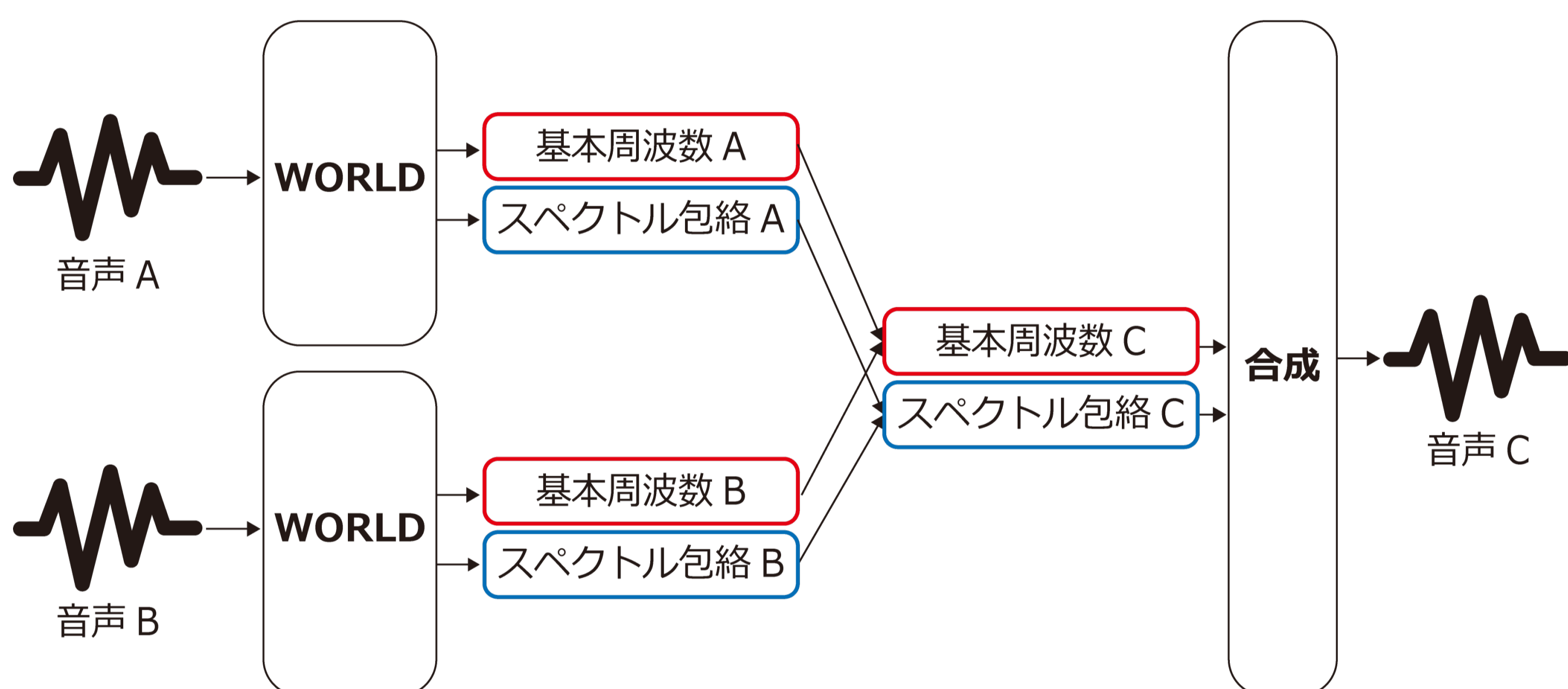
図：提案手法により得られた対応点をプロットしたスペクトログラム

発話内容「また、東寺のように、五大明王と呼ばれる、主要な明王の中央に配されることも多い。」

1. はじめに

音声モーフィング

同じテキストを発話する話者の2つの音声から、中間的な印象の音声を合成する技術。



時間軸・周波数軸で対応点を設定する必要がある。
→ モーフィング音声の品質は、作業者に依存する。

WORLDをベースとした自動設定された対応点を用いた音声モーフィングについて検討した。

時間軸：音素アライメントで推定された音素境界
周波数軸：第1, 第2フォルマント

関連研究

- STRAIGHTに基づいた音声モーフィング*
時間軸：母音の音素境界, 周波数軸：フォルマント周波数
*大西 他, 音講論 (春), pp.273-274, 2007.

2. 対応点の自動付与

Juliusによる自動音素アライメント

前処理① 8 kHzへのダウンサンプリング

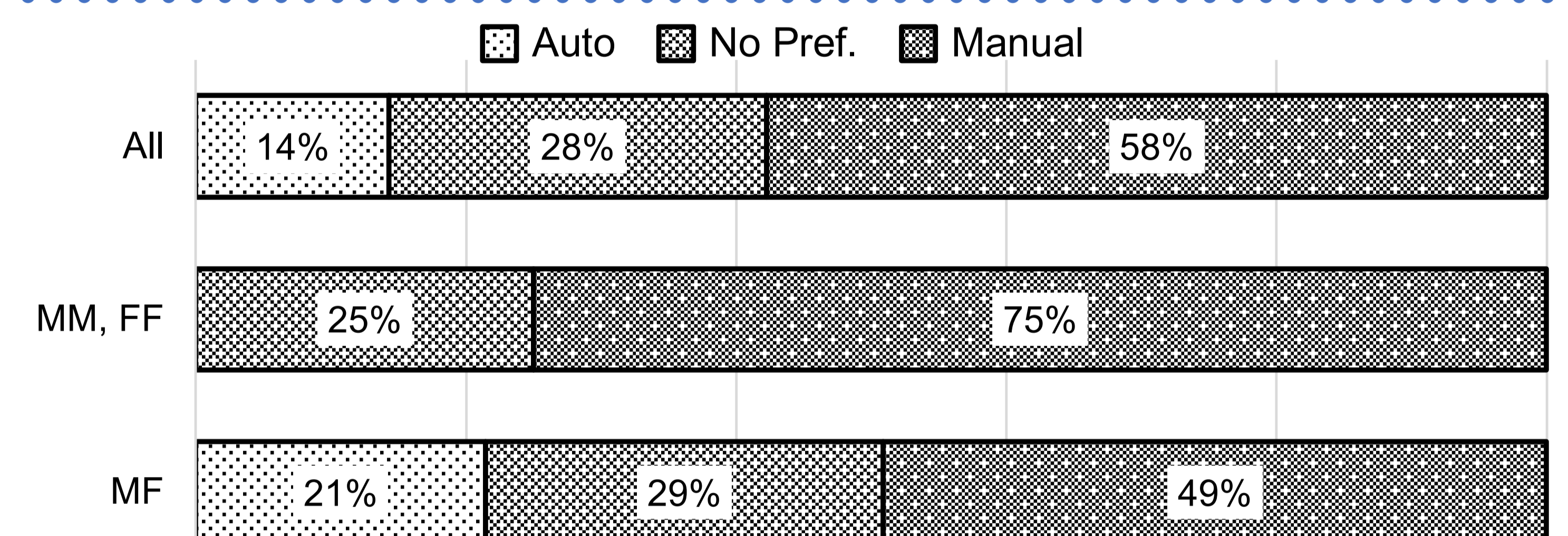
前処理② 高域強調

8次のLPC係数からフォルマント推定

3. 実験：Preference評価

モーフィング率	50 %
使用音声DB	JVS/parallel100コーパス (男女2話者ずつランダムに選択)
試行数	24ペア：6ペア* × 2文 × 2順番 (注：4話者からできる組合せ)
実験参加人数	7名 (学生)

4. 結果と考察



Auto：提案手法, Manual：手作業

- 40%程度の音声において同等の品質

同性による音声モーフィング

→ 同等の品質あるいはManualの方が自然性が高い。

- 一般に, 同性の音声モーフィングは異性の場合と比較して劣化しにくい。

= フォルマント推定の誤差による劣化が相対的に大きくなることが考えられる。

異性による音声モーフィング

→ 同性と比較して相対的にAutoが高い。

- 一般に, スペクトル包絡の周波数伸縮量の差が大きくなるため品質が低下しやすい。
- 使用した音声サンプルにフォルマント周波数の差が見られなかった。

5. 今後の展望

- 品質向上 (フォルマント推定誤差の解決)
- 音素アライメントに依存しない対応点の検討