

# WaveNetを用いた歌唱表現が可能な歌声合成方式の検討

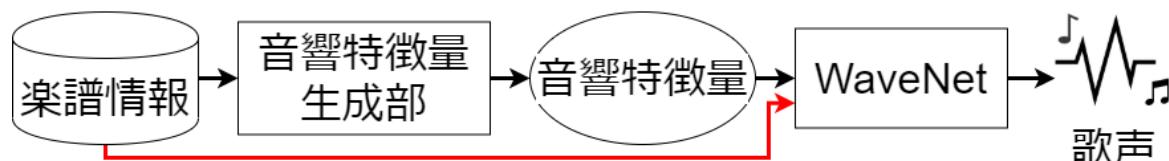
M1 金子隼人 (岡山大学 ヒューマンセントリック情報処理研究室)

## 1. 研究背景

- 歌声合成: 任意の楽譜から歌声を生成する技術
- 歌唱表現: 歌声を特徴づけるうえで重要な要素  
例: ビブラート等の音高変動
- WaveNet: 高品質で自然な音声を生成可能  
補助特徴量で波形を制御可能

## 2. 提案手法

- 歌声とそれに対応する楽譜があると仮定



- WaveNet部でも楽譜情報を用いる  
→ 歌唱表現付きの歌声合成を目指す

## 3. 実験条件

- ◆ 比較するモデル
  - AF: スペクトログラムのみを補助特徴量に使用
  - MN: スペクトログラム, 楽譜情報を補助特徴量に使用

### ◆ 使用したデータセット

東北きりたん歌唱データベース

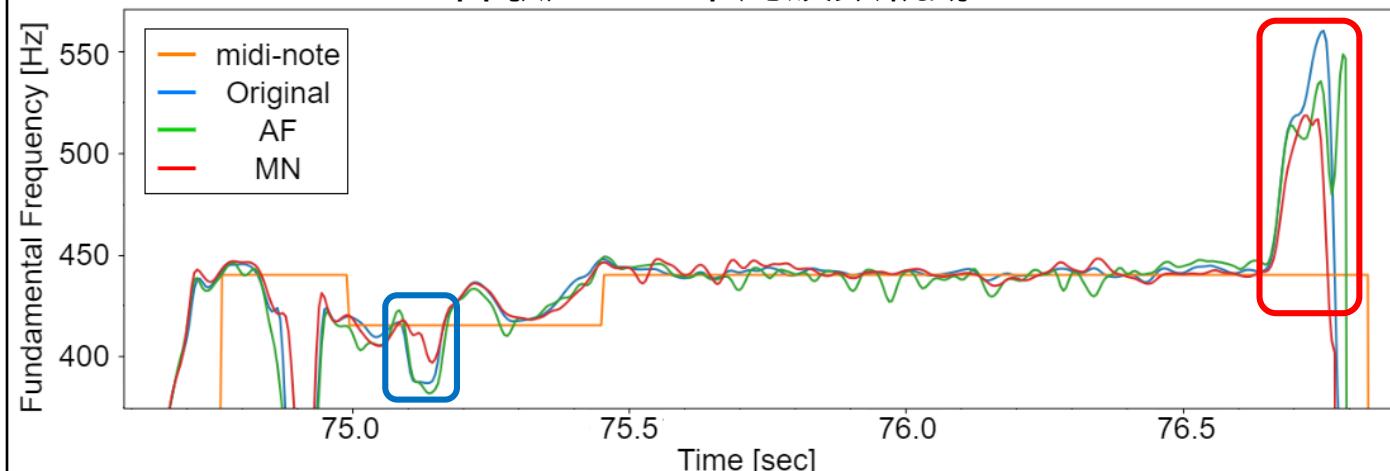
- 女性歌手1人によるJ-pop50曲(約1時間)
- 歌声に対応する楽譜データを含む

### ◆ 特徴量

- 音響特徴量: スペクトログラム80x1024
- 楽譜情報: 楽譜の音高(48次元one-hot-vector)

## 4. 実験結果

### <各歌声の基本周波数軌跡>



- 楽譜の音高から離れるような歌唱表現を再現(赤枠)  
→ 楽譜の音高を補助特徴量に使用しても問題なく再現できている

- 基本周波数軌跡が楽譜の音高に近づいている(青枠)

### <客観評価結果>

	Log-F0-RMSE[cent]
元データ	18.97
AF	18.63
MN	18.27

元データは楽譜の音高との距離が大きい  
→ 楽譜の音高から離れた表現を持つ

AFとMNは音高との距離が小さい  
→ 特徴が弱い歌声が生成されている  
→ フレーム単位の音高では不十分

## 5. まとめ

- WaveNetを用いた歌唱表現付きの歌声合成方式を提案
- フレーム単位の音高だけでは歌唱表現の再現が困難
- ◆ 今後の課題  
楽譜内での前後の情報を重視した特徴量を使用する