

音声分析

presented in ICASSP 2017

岡本 拓磨

情報通信研究機構

Frequency-warped time-weighted linear prediction for glottal vocoding

pp. 5630-5634

■ M. Airaksinen, B. Bollepalli, J. Pohjalainen, and P. Alku

■ Aalto University

■ 背景と目的

- Glottal vocoder : 線形予測による残差信号のモデル化→位相を考慮
- TTS : ラベル→[DNN]→予測重み(声道スペクトル)→[DNN]→残差信号
→STRAIGHT vocoderよりもTTSで高音質を実現

■ 従来法

- 周波数ワーピング単独 : 聴覚特性は考慮しているが逆フィルタリングの精度が悪い
- 重み付き線形予測単独 : QCP法により精度向上→GlottDNN (interspeech 2016)

■ 提案法の特徴

- 両者を統合させた「いいとこ取り」の検討

■ 結果

- 分析合成の場合女性の方で改善あり, TTSの場合男性音声で改善あり
→IEEE Signal Process. Lett., 24(4), pp. 446-450に詳細版あり

Dereverberation based on bin-wise temporal variations of complex spectrogram

pp. 5635-5639

■ T.-H. Chen, C. Huang, and T.-S. Chi

■ National Chiao Tung University

■ 背景と目的

■ DNNに基づく1チャンネルマイクロホン入力の残響成分分解

■ 人間の聴覚特性：残響の位相成分にも敏感

■ 従来法

■ 通常のDAE：パワースペクトルの復元→位相が考慮されていない

■ STFTにおける残響成分：掛け算ではなく畳み込み

■ 提案法の特徴

■ STFTパワースペクトルをもう1回フーリエ変換：畳み込み→掛け算

■ rate領域での複素スペクトルをDNNで復元：位相成分も考慮

■ 結果

■ パワースペクトルのみ回復した従来法よりも高精度な残響成分分解を実現