高品質音声合成研究の成果を紹介

ニューラルネットワークに基づくテキスト音声合成

どんな研究？

近年，様々の場面で使われ始めてきた音声合成技術．音声合成に対して，ニューラルネットワークを用いる手法の研究を行っています．

何ができる？

- 任意のテキストの音声を合成
- 音声対話システムの基礎技術
- ナレーション，アナウンス，エンターテインメントへの活用

状況設定

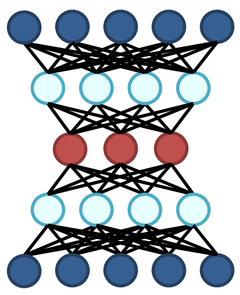
- 音声合成技術の高度化
  - 音声対話システム(カーナビ，Siri)
  - テレビのナレーター
  - 声の障碍への対策

  ➡ 様々な場面で活用

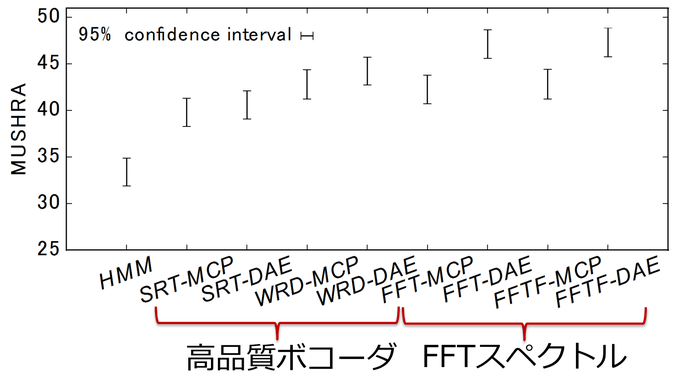
研究内容

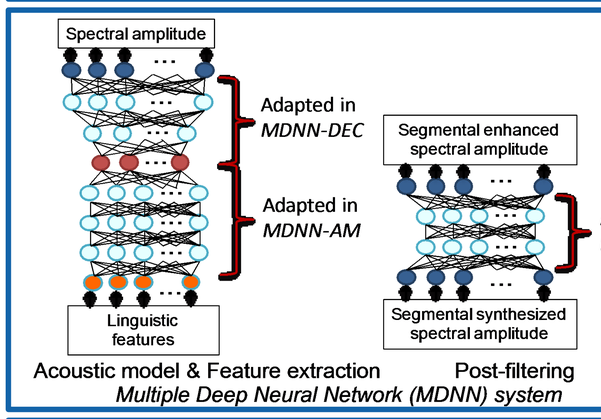
テキスト音声合成におけるニューラルネットワークの活用例

・より原信号に近い入力を用いた音声合成

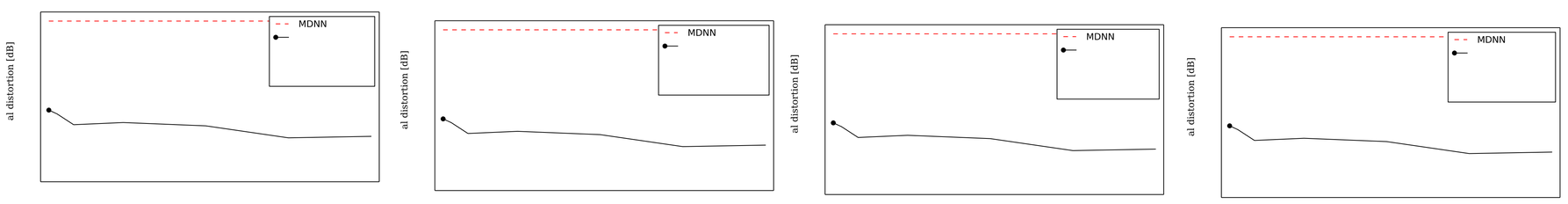

FFTスペクトルを入力

・通常，高品質ボコーダから得られたスペクトル情報を利用
・音声合成の目標：音声波形の再現
  ⇨ 原信号との誤差を少なくする方向性が考えられる
・より原信号に近い単純なFFTから得られたスペクトルの利用
・Auto-encoderに基づくFFTスペクトルからの特徴抽出

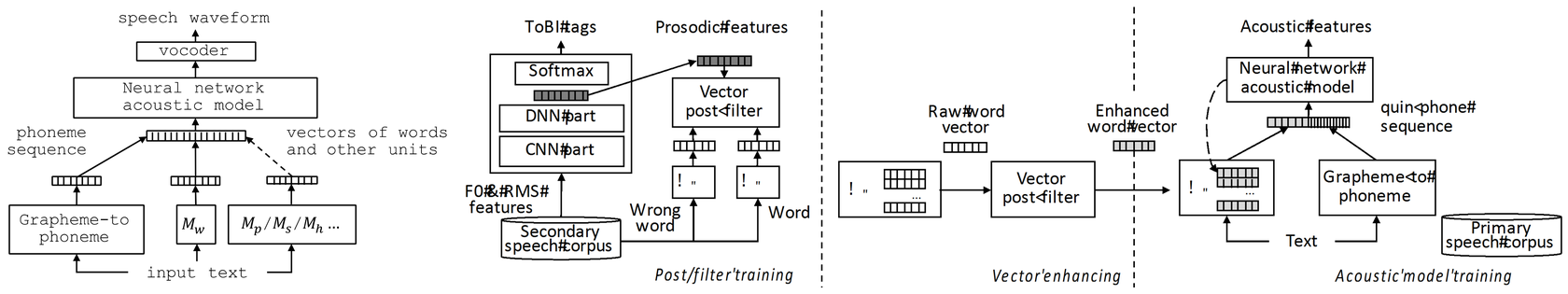


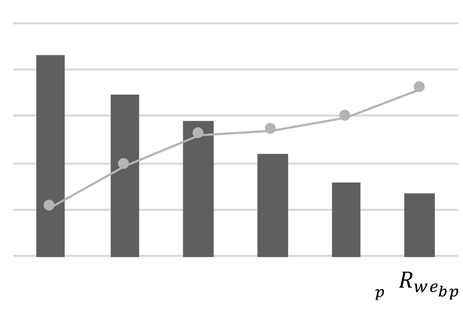


・少量のデータを用いた適応技術の検討
・音響モデル・特徴抽出・ポストフィルタをニューラルネットワークで実現し，適応技術を利用

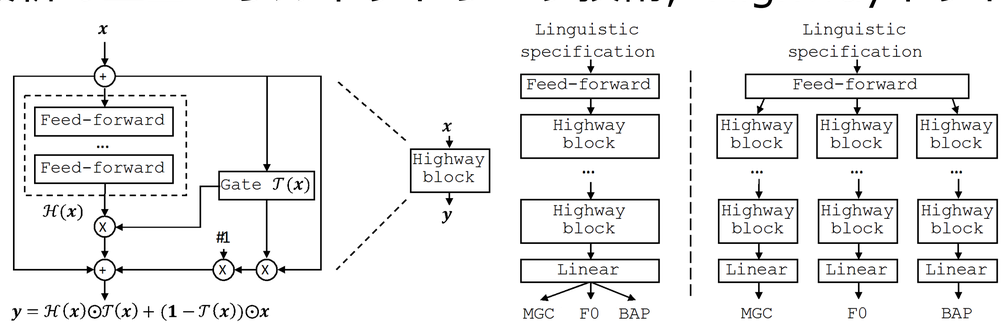


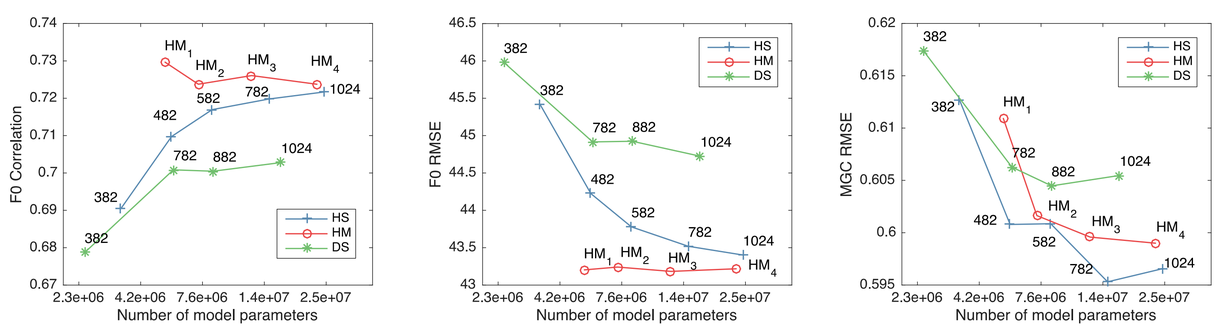
・単語低次元数値表現のための韻律情報を用いたポストフィルタリング





・最新のニューラルネットワーク技術，Highwayネットワークをテキスト音声合成へ導入





NII
連絡先：高木信二／ 国立情報学研究所 コンテンツ科学研究系
TEL：03-4212-2576 FAX：03-4212-2150 Email：takaki@nii.ac.jp