

# 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？

様々な場面で使われ始めている音声合成技術。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？

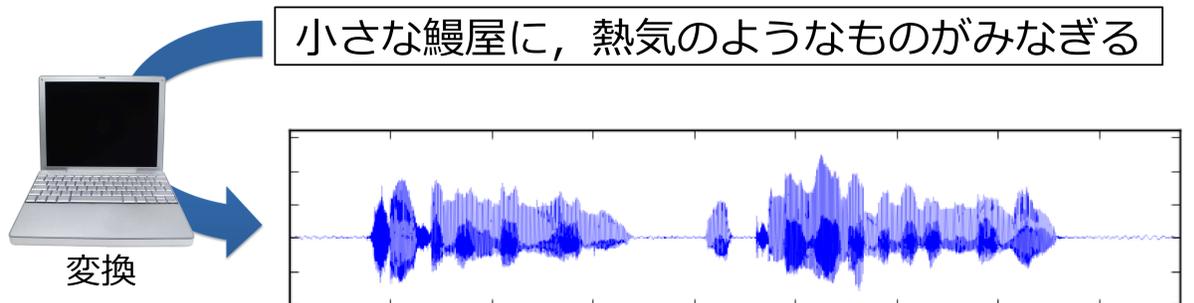
- お喋りする人工知能の実現
- ナレーション、アナウンス、エンターテインメントへの活用
- 音声対話システムの基礎技術

### 状況設定

#### • 音声合成の必要性

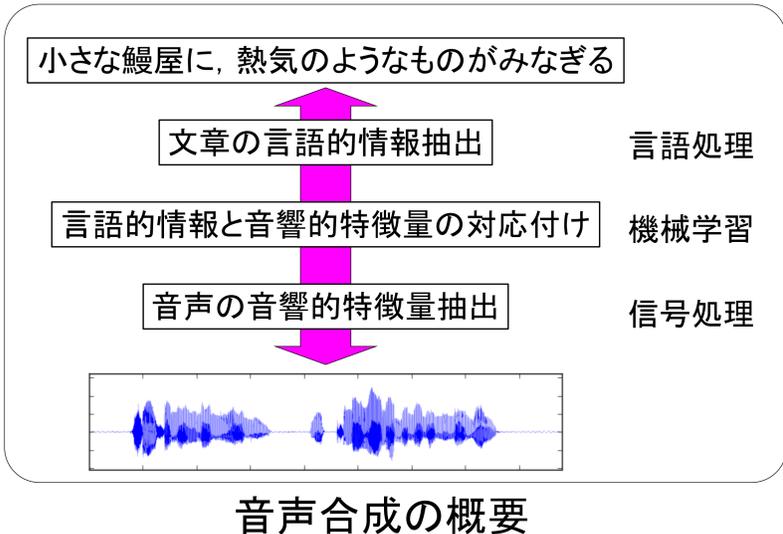
- 情報の取得 (カーナビ, Siri)
- スクリーンリーダ, 会話補助
- エンターテインメント

様々な場面で活用

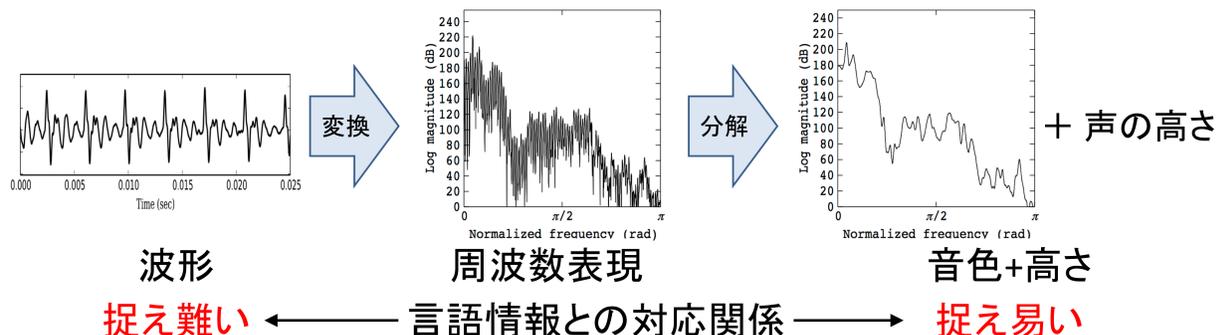


### 研究内容

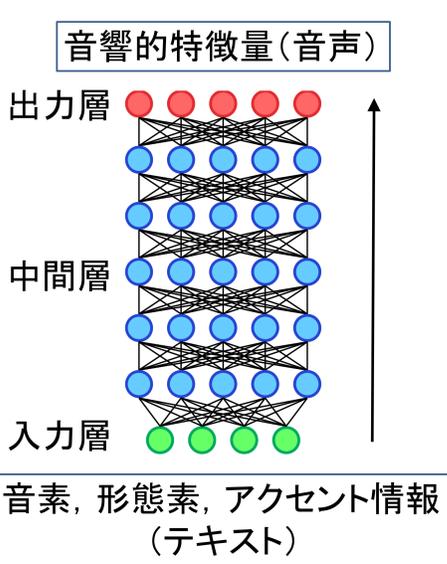
#### 音声合成の基礎



- 言語的情報抽出
  - 音素への対応付け、形態素解析(品詞, 句構造)、アクセント推定など  
赤い空 → 音素:/a/ /k/ /a/ /i/ /s/ /o/ /r/ /a/. アクセント:あかいそ | ら
- 言語的情報と音響的特徴量の対応付け
  - ニューラルネットワークの利用
- 音声の音響的特徴量抽出



#### ニューラルネットワーク音声合成の基礎



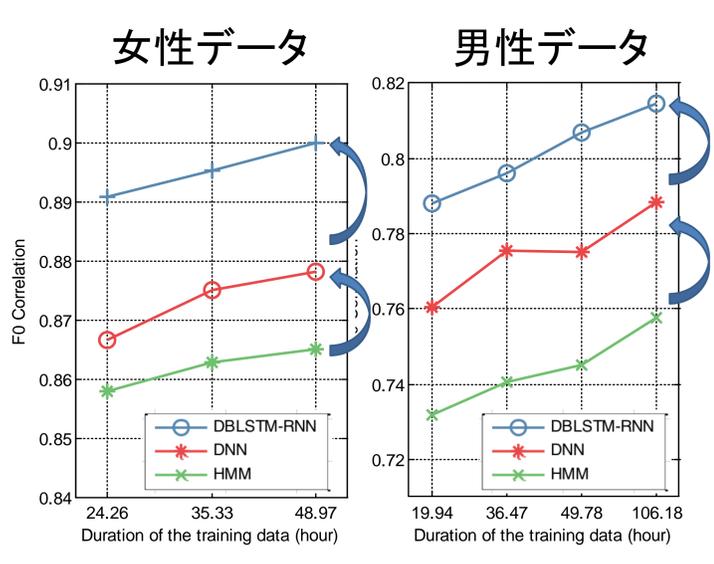
- ニューラルネットワーク
  - 関数を表現(入力 $x$ , 出力 $y$ ):  $y = f(x)$
- 音声合成: 文章(入力)を音声(出力)へ変換
 

入力: 1 0 0 ... 1  
(例: アクセント有の/a/) : /a/ /i/ /u/ ... アクセント

出力: 音声波形や音響的特徴量の値
- ニューラルネットワーク音声合成器の構築
  - 関数の中身を決めること
$$y = W' * x$$

出力 ← 変換 ← 入力

  - $W$ の値を学習する → 音声合成器の構築



実験結果(F0の相関係数)  
従来手法(HMM)より非常に高い性能