

# 深層学習による音声合成

## 単語の発音の知識を与えずに、 入力文字から直接音声を合成する技術

安田 裕介 (NII)、シンワン (NII)、山岸 順一 (NII)

### どんな研究？

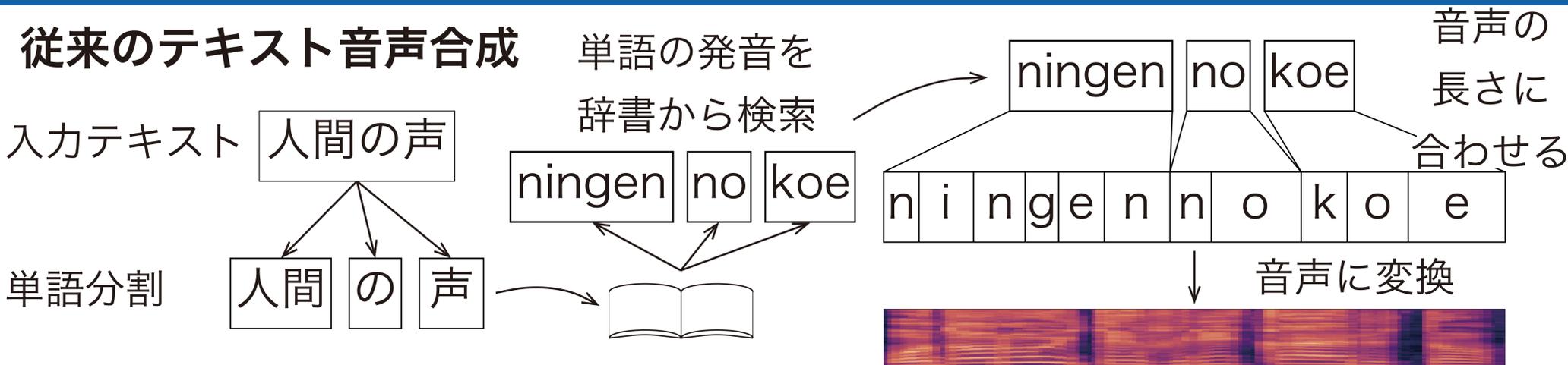
- ・ テキスト音声合成はテキストから人間の声を人工的に生成する技術
- ・ 従来、文字の読みを辞書で解決
- ・ 辞書を使わず、テキストから音声に直接変換する研究です

### 何がわかる？

- ・ テキスト音声合成がより簡単になります
- ・ 様々な言語に対して使える汎用的な手法です

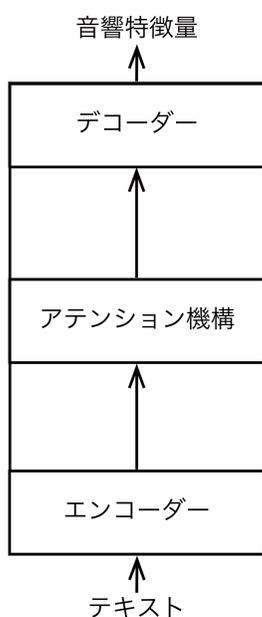
### 状況設定

#### 従来のテキスト音声合成



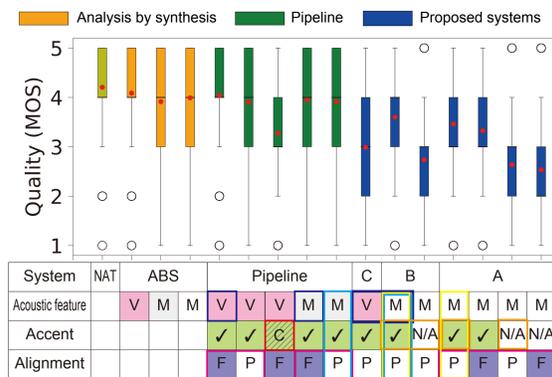
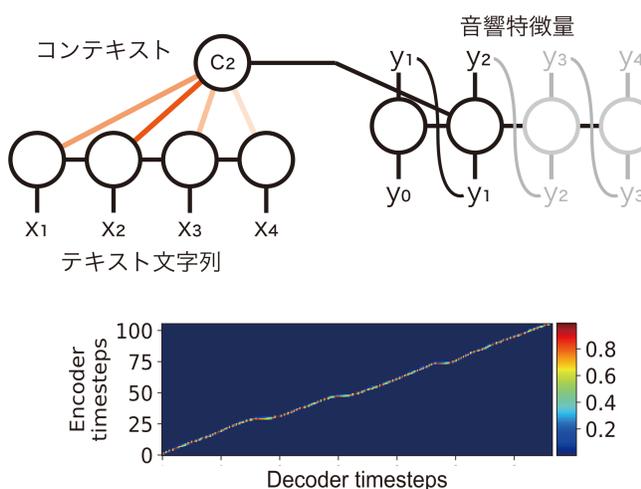
### 研究内容

#### 新しいテキスト音声合成 エンドツーエンド音声合成



1つのモデルで  
テキストから音声に  
直接変換

アテンション機構が自動で  
テキストと音声の  
対応関係を学習



#### まだまだ課題あり！

- ・ 大量の音声データが必要
- ・ 日本語のピッチアクセントの予測にはラベルが必要
- ・ 漢字かな混じり文の入力はまだ難しい
- ・ 音声波形への直接変換ではなく、その前段階の音響特徴量までを予測