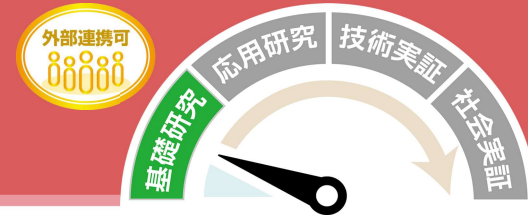




# 信号処理と深層学習を融合した音声波形生成技術

～ 声を自在に操作できる音声技術の実現に向けて ～



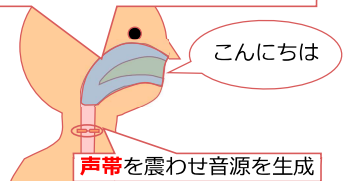
## 概要

深層学習による波形生成技術で問題であった生成時間や声の制御の困難さを克服するため、信号処理の知見を導入し、深層学習の波形生成技術の高い品質を維持しながら、高速な波形生成と自由な声の制御性を実現します。

### ソースフィルタモデル

・人の発話機構を信号処理技術でモデル化

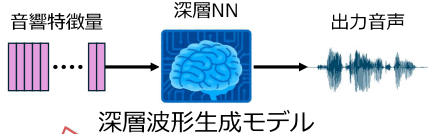
声道で音源を共鳴させ音色を生成



声帯を震わせ音源を生成

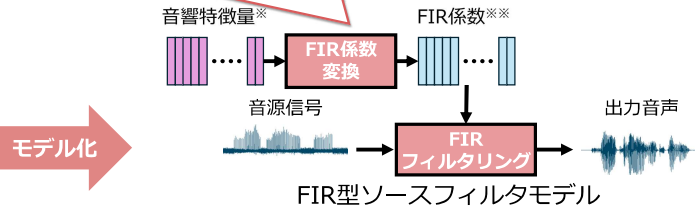
### 深層波形生成モデル

・音響特徴量から深層ニューラルネットワーク (NN) で波形を生成



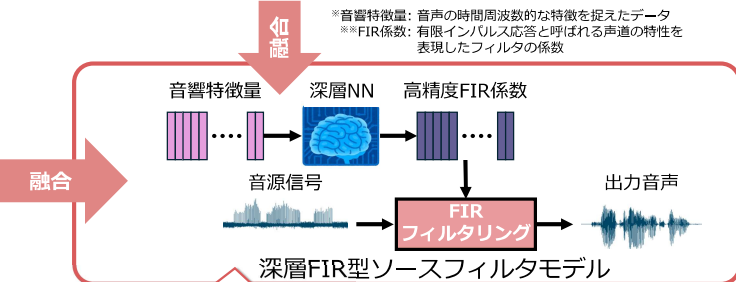
深層NNにより高音質な波形生成が可能な一方、処理が複雑であるため波形の生成速度は遅め、またデータ駆動型のため声の制御に強い制約

柔軟な声の制御や高速な音声生成が可能な一方、音響特徴量に含まれる分析誤差の影響で、生成波形の音質は低め



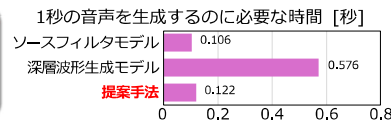
FIR型ソースフィルタモデル

\*音響特徴量: 音声の時間周波数的な特徴を捉えたデータ  
\*\*FIR係数: 有限インパルス応答と呼ばれる声道の特性を表現したフィルタの係数



深層FIR型ソースフィルタモデル

波形より簡単なFIR係数への変換に適用することで処理を単純化し、信号処理並みの速度と声の制御性を実現！さらに深層波形生成モデル並みの高音質化を両立



## 特徴

- ・従来の信号処理技術と深層学習を融合
- ・高速かつ高品質な音声生成が可能
- ・劣化を抑えつつ自在な声の制御が可能

## ユースケース

- ・同時通訳での発話者区別のための出力音声の調整
- ・歌声合成や動画コンテンツなどエンタメ向け応用
- ・組み込み向け音声合成システムへの適用

## 今後の展開

- ・本機構の音声合成システムの適用
- ・音声通訳アプリや同時通訳システムへの展開

【お問合せ先】

ユニバーサルコミュニケーション研究所 先進的音声翻訳研究開発推進センター 先進的音声技術研究室 大谷 大和  
Mail : Yamato.ohtani@nict.go.jp