様々な合成単位における End-to-end 逐次音声合成の検討*

○柳田 智也 (NAIST), サクティ サクリアニ，中村 哲 (NAIST/RIKEN AIP)

1 はじめに

同時音声翻訳システムは，元言語の音声を目標言語の音声へ逐次翻訳し出力する．そのシステムは，音声認識・機械翻訳・音声合成から構築される．通常の音声合成は，文全体の入力後に処理を行うため，一発話が長い状況において出力に深刻な遅延が生じる．従って，同時通訳用の音声合成は，入力を逐次処理し出力する必要がある．逐次音声合成は，隠れマルコフモデルによる方法が広く研究されている [1, 2]．近年，End-to-end 音声合成が，高品質な音声合成を実現している [3]．より高品質な逐次音声合成のため，本研究は，End-to-end 逐次音声合成に取り組む．先行研究の End-to-end 逐次音声合成の検討 [4] では，逐次音声合成の合成単位を，全て短文として扱う．従って，合成単位間の音響特徴の変化や，テキスト中の位置情報が考慮できない．本論文は，これらを考慮する方法を提案し，提案する End-to-end 逐次音声合成の音声品質を様々な合成単位で調査する．

2 End-to-end 音声合成

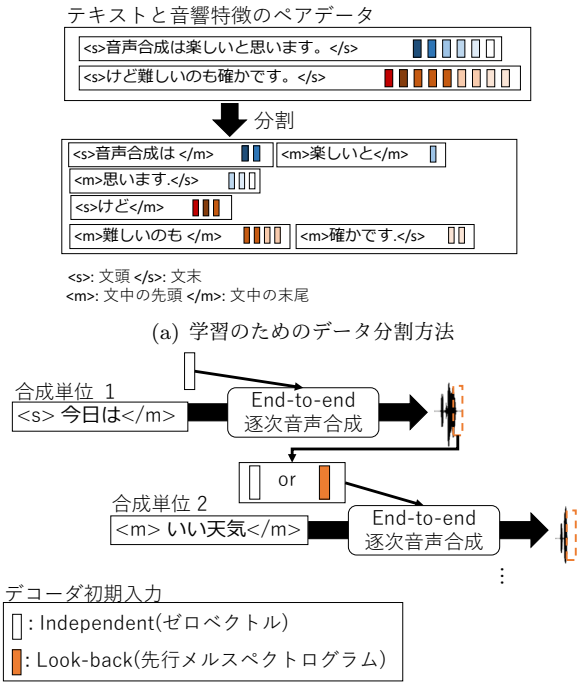
深層学習に基づくエンコーダーデコーダ型 End-to-end 音声合成として，Tacotron[3] 使用する．先行研究の Tacotron は、表層文字からメルスペクトログラムとリニアスペクトログラムとを推定し、リニアスペクトログラムと推定した位相から音声を復元する．本論文では、読み推定を容易にするため，音素記号系列を入力とする．更に，日本語の高低アクセントを表現するため，アクセント句のアクセント型も使用する．本研究の Tacotron には [5] と同様に，音響特徴の出力停止を示す止フラグを出力する層を追加する．

3 提案方法

3.1 逐次音声合成の学習と合成方法

Fig. 1(a) に，学習時のデータ分割について示す．逐次音声合成では，文より短い単位を合成単位として扱う．先行研究 [4] では，合成単位のテキスト全てに，文頭を表す記号 (<s>) と文末を表す記号 (</s>) を付与している．これは，合成単位を，全て短文として扱っており，合成単位間における音響特徴の時系列変化を考慮できない．この問題を改善するため，初めに，データセットのテキストを，無作為に三分割し，Fig. 1(a) に示すように，文頭と文末を表す記号の他に，文中の先頭を表す記号 (<m>) と文中の末尾を表す記号 (</m>) をテキストに付与する．学習は，分割したデータ及び文単位のデータを両方を用いる．

合成時は，Fig. 1(b) に示すように，文単位より短い合成単位毎に行う．ここで，メルスペクトログラムの生成は，停止フラグが 1 を出力するまで生成し，フレーム長を決定する．その後，出力された各音声波形を接続し，文単位の合成音声を作成する．合成単位の選択については，次節に示す．



(a) 学習のためのデータ分割方法

(b) 合成手順とデコーダ初期入力の選択

Fig. 1 学習と合成方法の概要

3.2 合成単位の選択および初期入力の提案

Fig. 1(b) に示す合成単位としては，アクセント句が適当な単位として調査されている [2]．本研究では，アクセント句を含む以下の合成単位を用いる．

１アクセント句: １アクセント句毎に合成する．

２アクセント句: ２アクセント句毎に合成する．

３アクセント句: ３アクセント句毎に合成する．

半文: 文長の半分となるアクセント句毎に合成する．

１文: １文毎に合成する．

実利用時は，半文の検出は困難であり，この条件は実験的設定として用いる．また，１文は，通常の音声合成と同様の実験条件である．

End-to-end 音声合成の学習および合成時に，メルスペクトログラム推定用デコーダの初期値は，”Go frame ”と呼ばれるゼロベクトルを用いる．音響特徴の時系列変化を考慮し，合成単位間の韻律の不連続を改善するため，図 1(b) に示すように，学習時と合成時において，初期入力に対する以下の提案を行う．

Independent: 各合成単位の初期入力としてゼロベクトルを用いる．この場合，音響特徴の時系列変化は，文単位の入力時のみ学習される．

Look-back: 初めの合成単位のみゼロベクトルを用いる．それ以降の初期入力は，先行合成単位のメルスペクトログラム系列における最後方の出力を使用する．この場合，各合成単位毎に音響特徴の時系列変化を考慮する可能性がある．

*End-to-end incremental speech synthesis in various synthesis units, by YANAGITA, Tomoya, SAKTI, Sakriani, NAKAMURA, Satoshi (Nara Institute of Science and Technology/RIKEN AIP).