モーラ情報とアクセント情報を用いた波形接続型音声合成の普通名詞への応用*

○石田 隆浩，村上 仁一，池原 悟 (鳥取大・工)

1 はじめに

波形接続型音声合成は，あらかじめ録音しておいた音声波形を，音素単位や音節単位などで接続することによって合成音声を作成する方法である．波形に信号処理を行わずに接続することにより，話者性と高い自然性を保てる特徴がある [1].

一方，波形接続型音声合成においては，韻律の扱いが問題となる．この問題に対する一つの解決法として，モーラ情報を用いる方法が提案されている．参考文献 [1] で提案されている音節波形接続方式では，地名を対象に実験した結果，実用的な品質が得られたことが報告されている．また参考文献 [2] において，同様の方法を普通名詞に適用した場合も，明瞭性の高い合成音声が作成できたことが示されている．しかし，普通名詞に適用した場合，アクセント型の未考慮による合成音声の品質の悪化が問題となった．

そこで本研究では，アクセントを考慮することによって合成音声の品質をどの程度改善できるかについて調べる．

2 アクセントを考慮した音節波形接続方式

2.1 モーラ情報とピッチ情報

音節波形接続方式では，韻律的な情報として，モーラ情報 (単語のモーラ数とモーラ位置) を使用する．特定話者の単語発話においては，単語のモーラ数とモーラ位置が決まれば，単語によらずピッチ周波数がほぼ決定されることが知られている [1]．過去の研究では，モーラ情報は音素ラベリング [3] や音声認識 [4][5] などの分野において効果があることが報告されている．

2.2 アクセントを考慮した音節波形接続方式による音声合成

音節波形接続方式による普通名詞の合成音声の作成において，モーラ情報を用いることで，明瞭性の高い合成音声を作成できたことが報告されている．しかし一方で，アクセント型の未考慮による品質の悪化が指摘されている [2]. そこで本研究では，NHK 日本語発音アクセント辞典 [6] を元に音声データにアクセント情報を付加し，波形選択においてアクセントを考慮する．

アクセントを考慮した音節波形接続方式における音声合成では，以下の情報が一致する音節部品を選択する．

- 音節
- 直前の音素 (前音素環境)
- 直後の音素 (後音素環境)
- 単語中のモーラ位置
- 単語のモーラ数
- 単語のアクセント型

そして，音節の開始時間と終了時間を元に波形データを切り出し，接続して合成音声を作成する．

*Common Noun for Word Synthesis by Concatenating Syllabic Components with Mora and Accent. By Takahiro Ishida, Jin'ichi Murakami and Satoru Ikehara (Tottori Univ.)

2.3 自動音素ラベリング

波形接続方式による音声合成においては，一般に，人手によるラベルデータが必要となる．しかし，コストがかかる点が問題となる．

そこで本研究では，基本的な HMM によるセグメンテーションの方法を用いて自動的に音素ラベリングを行い，そのラベルデータを用いた場合について品質を調査する．

3 評価実験

3.1 実験環境

本研究では，音声データベースとして，ATR の単語発話データベース Aset(5,240 件) を使用する．そして，Aset に含まれる 4 モーラ単語のうち 50 個について，自然音声，アクセント型未考慮の合成音声 (ac 無)，アクセント型を考慮した合成音声 (ac 有), 市販の合成機 [7] の合成音声 (市販の音声) を作成する．

音節波形接続方式による合成音声 (ac 無, ac 有) については，人手によるラベルデータ (手動ラベル) を用いた音声 (ac 無/有 (手動)) と，自動音素ラベリングによるラベルデータ (自動ラベル) を用いた音声 (ac 無/有 (自動)) の 2 種類の音声を準備する．

話者には，ピッチ周波数のばらつきが比較的小さく，収録された音声にエコーが少ない，FTK と FYN の 2 話者を選ぶ．

3.2 評価方法

合成音声の評価のために，音声研究に関わった経験のない人 5 名を対象に，自然音声と合成音声をランダムにヘッドフォンから被験者に聞かせ，了解度試験とオピニオン評価を行う．評価は，作成した単語を文に埋め込んで行うのではなく，単語音声のみで行う．

まず，単語音声の明瞭性を調べるために了解度試験を行う．了解度試験では，どのように聞こえたかを仮名で書き取らせる．自分の知識などは用いず，聞こえたとおりに書き取るように指示する．

また，単語音声の自然性を調べるために，オピニオン評価を行う．オピニオン評価では，自然に聞こえた度合を 5 段階 (1 が最も不自然，5 が最も自然) で評価するように指示する．

3.3 自動音素ラベリング

自動音素ラベリングでは，同一話者において，人手によるラベルデータがあることを前提に，自動セグメンテーションを行うこととする．

具体的には，自動音素ラベリングのツールとして，HTK[8] を使用する．そして，Aset の奇数番のデータを利用して HMM を学習し，偶数番のデータのセグメンテーションを行う．また，偶数番のデータを利用して HMM を学習し，奇数番のデータのセグメンテーションを行う．

なお，特定話者に対する自動セグメンテーションにおいて，人手による音素境界位置に対する標準偏差は，参考文献 [3] による方法で 20ms 程度であったと報告されている．

3.4 合成音声の例

本研究で作成した合成音声の一部を以下に示す．なお，「_|‾‾‾‾」はアクセントを表している．

幹旋 (/a_|q‾/se‾/N‾/)= 圧倒 (/a_|q‾/to‾/u‾/)
+ 雑草 (/za_|q‾/so‾/u‾/)
+ 実践 (/ji_|q‾/se‾/N‾/)
+ 国電 (/ko_|ku‾/de‾/N‾/)

印刷 (/i_|N‾/sa‾/tsu‾/)= 引率 (/i_|N‾/so‾/tsu‾/)
+ 申請 (/shi_|N‾/se‾/i‾/)
+ 診察 (/shi_|N‾/sa‾/tsu‾/)
+ 推察 (/su_|i‾/sa‾/tsu‾/)

3.5 波形接続方式に関する補則

本研究では，情報が一致する音節候補が複数ある場合，ランダムに選択する．また，波形を切り出す位置は，2 つの音節波形の位相を考慮し，接続部分の振幅の差がゼロに近づ??茲Δ膨汗阿鮃圓Ａ?

4 実験結果

実験結果を表 1 に示す．

表 1: 実験結果

	了解度　正解率 (%) 評価音節数: 1,000			オピニオンスコア 評価単語数: 200		
	FTK	FYN	平均	FTK	FYN	平均
自然音声	95.6	96.0	95.8	4.9	4.8	4.85
ac 無 (手動)	90.4	95.6	93.0	3.7	4.0	3.85
ac 有 (手動)	94.4	98.4	96.4	4.3	4.3	4.30
ac 無 (自動)	90.8	96.4	93.6	3.5	4.3	3.90
ac 有 (自動)	94.0	96.4	95.2	4.1	4.2	4.15
市販の音声	79.2	78.4	78.8	2.1	2.2	2.05

表 1 から，手動ラベル，自動ラベルのいずれの場合においても，アクセント未考慮の場合よりもアクセントを考慮した方が，了解度，オピニオンスコアとも高くなっていることが分かる．

また，手動ラベルと自動ラベルの結果にはあまり違いはなく，自動ラベルでも品質の高い合成音声を作成できたことが分かる．

しかし，自然音声と比較すると，了解度は同程度であったが，オピニオンスコアは及ばなかった．

5 考察

5.1 了解度試験の解析

了解度試験において，手動ラベルでアクセントを考慮した場合に，被験者が間違えた音声の一部を表 2 に示す．なお，表中の下線部は，被験者が間違えた箇所を示す．

表 2: 間違いの例

	正解	間違いの例
f1	はいせき (排斥)	か いせき，た いせき
f2	さいだい (最大)	さい が い
f3	せつだん (切断)	せつ が ん
f4	かいだん (階段)	かい が ん
f5	ぐんかん (軍艦)	ぶ んかん
f6	れんさい (連載)	め んさい，え んさい
f7	ちょっかく (直角)	しょっかく

表 2 からも分かるように,「は」と「か」,「だ」と「が」など，似た音韻を間違える場合が多かった．また，特に「a」の音を持つ場合に多く間違えていた．

5.2 オピニオン評価の解析

オピニオン評価においては，波形の接続部分が不自然に聞こえる部分を含む音声において評価が悪かった．実験では，例えば「最大」という音声で「祭日」の第 1 音節と「階段」の第 2 音節の接続部分が不自然に聞こえた．

また，FYN においてはあまりアクセントの効果が見られなかった．これは，FYN の音声の抑揚があまり明確でないためと考えている．

5.3 データベースにおける品質のばらつき

音声データベースに使用した Aset では，5,240 単語の音声の音量や音質に多少ばらつきがあった．そのため，例えば音量の大きい波形と音量の小さい波形をつないだ結果，自然性が落ちることがあった．

今回の実験では，例えば「解散」という音声で第 2 音節に利用した「財政」と第 3 音節に利用した「計算」の音量差による自然性の低下が見られた．

この問題に対しては，録音環境 (日時，場所など) が分かる場合，それらが近い音声を優先的に用いることで多少改善することができると考えられる．

5.4 音節の選択方法に関する考察

今回の実験では，音節選択方法については，複数の候補が残った場合にはランダムに選択する方法を採った．

アクセントを考慮した場合，考慮しない場合に比べ候補数が減少するため，ランダムに選択した場合でも，品質にそれほど大きなばらつきは生じなかった．

しかし，さらに候補を絞り込む手法としては，参考文献 [2] で提案されている方法以外に，接続部分の FFT スペクトルを比較して近いものを選択する方法や，各音節の発話時間が最も近くなるようなものを選択する方法などが考えられる．

6 まとめ

本研究では，音節波形接続方式を普通名詞に応用する場合において，アクセント情報の有効性を調査した．聴覚実験における合成音声の単語了解度は，手動ラベルの場合で 96.4 %，自動ラベルの場合で 95.2 %が得られた．また，オピニオンスコアはそれぞれ 4.30，4.15 が得られた．アクセント情報を考慮することで了解度は 3.4 %，オピニオンスコアは 0.45(手動ラベルの場合) の向上となり，アクセント情報が有効であることが分かった．

一方，自然音声の単語了解度は 95.8 %，オピニオンスコアは 4.85 であった．オピニオンスコアは自然音声には及ばないものの，より自然音声に近い音声を作成可能であることがわかった．

今後は，波形候補が複数残った場合の絞り込み手法や考察で述べた手法の検討を行い，さらに自然音声に近い合成音声の作成を目指したい．

参考文献

[1] 村上, 水澤, 東田, ”音節波形接続による単語音声合成,” 電子情報通信学会論文誌 D-II Vol.J85-D-II No.7 pp.1157-1165 (2002.7).

[2] 石田, 村上, 池原, ”音節波形接続型音声合成の普通名詞への応用,” 信学技報, SP2002-25, pp.7-12 (2002.5).

[3] 前田, 村上, 池原, ”モーラ情報を用いた音素ラベリング方式の検討,” 信学技報, SP2001-53, pp.25-30 (2001.8).

[4] 妹尾, 村上, 池原, ”モーラ情報を用いた単語音声認識の検討,” 信学技報, SP2002-130, pp.55-61 (2002.12).

[5] 谷口, 村上, 池原, ”モーラ情報を用いたフィルタバンクによる孤立単語認識,” 信学技報, SP2002-131, pp.63-68 (2002.12).

[6] ”NHK 日本語発音アクセント辞典 新版”, NHK 出版, ISBN4-14-011112-7 (1998).

[7] 富士通株式会社, ”Linux 版 日本語音声合成ライブラリー” http://www.createsystem.co.jp/

[8] Hidden Markov Model Toolkit (HTK) http://htk.eng.cam.ac.uk/