木に基づく状態共有を利用した波形接続型音声合成法の検討*

○山形 亮，堀田 波星夫，村上 仁一，池原 悟 (鳥取大)

1　はじめに

音声合成の手法の一つとして波形接続型音声合成がある [1]．この手法は録音音声から音節を単位とし，波形素片を取り出し，信号処理をせずに接続することで，自然性の高い音声を合成できる．

しかし，任意の一般名詞を作成しようとするには大量の録音単語が必要である．音声データベースとして，ATR 単語発話データベース Aset(5240 単語) を使用した場合，5240 件中の 470 単語しか作成できない．そこで任意の音声を合成可能にするために，収録されている DB に対して木に基づくクラスタリング [3] を行い，音響パラメータが似た音節素片をグループ化する．波形接続型音声合成でこれを利用して音声を合成し音質の評価を行う．

2　波形接続型単語音声合成

波形接続型音声合成 [1] では，以下の条件の音節素片を接続して音声を合成する．特徴として言語的なパラメータだけを用いて音節素片を選択し音響的なパラメータは使用しない．そのため，自然性の高い音を合成することができる．

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

合成音声の例を以下に示す．なお「_|¯」は音の強弱（アクセント）を表し，太字の部分が音節素片である．

一帯 (/_i/q̄/t̄a/i/)= 一掃 (/**_i**/**q̄**/t̄a/i/)
+ 実態 (/**ji**/**q̄**/**t̄a**/i/)
+ 絶対 (/**ze**/**q̄**/**t̄a**/**ī**/)
+ 組合 (/**ku**/**m̄i**/**a**/**ī**/)

3　木に基づく状態クラスタリング

木に基づくクラスタリング [4] は，音声認識で学習データを効率良く使うためによく使われている．音響的特徴が類似した triphone HMMの状態集合に対して音声の決定木に基づいてクラスタリングを行う．グループ化された音節素片の情報を使うことで任意の音を合成することが可能となる．

4　評価実験

収録されたＤＢに対して木に基づくクラスタリングを行い，音響パラメータ的に似た音節素片をグループ化し，これを用いて波形接続型音声合成で合成音声を作成する．合成した音声の品質を調べるために，同一発話の自然音声と波形接続型音声合成で作成した音声を用意し聴覚実験を行う．

4.1　実験データ

本実験では音声データベースとして，ATR 単語発話データベース Aset(5240 件) を使用する．そして，Aset に含まれる 3,4 モーラ語の 100 音声 (各 50 音声) を利用する．そして以下の条件の音声を準備する．

1) 自然音声
2) オリジナルの波形接続型音声合成 (以後オリジナル合成)
3) 木に基づく状態共有を利用した波形接続型合成音声 (以後クラスタリング合成)

4.2　木に基づくクラスタリングを用いた合成音声

本研究では，前後音素環境に加えて，モーラ長，モーラ位置，アクセント型を考慮した質問を用いて木に基づくクラスタリングを行い，共有された音節素片の情報に基づいて音声合成を行う．質問の例を表 1 に示す．

Table 1　クラスタリングの質問の例

1	前音素環境は鼻音であるか?
2	モーラ数は 3 または 4 で，モーラ位置は 1 であるか?
3	アクセント型は 1，2，3 のどれであるか?

クラスタリングされた音素素片の例として, 前音素 q, 後音素 i, モーラ数 4，アクセント 0 型の単語 3 モーラ目の音素が “ta” の音素素片についての情報を表 1 に示す．なお，木に基づくクラスタリングの学習データには Aset の話者 fyn の奇数番 2620 単語を用いる．

Table 2　状態共有された音節素片

音節	前音素	後音素	モーラ数	モーラ位置	アクセント型	アクセントの高低
ta	e	i	4	3	3	高
ta	i	i	4	3	0	高
ta	q	i	4	3	0	高

*Speech synthesis by Concatenating syllabic components used treebased clustering . by Ryo Yamagata,Haseo Hotta, Jin'ichi Murakami and Satoru Ikehara (Tottori Univ.)

以下に木に基づくクラスタリングを用いた合成音声の例を示す.

一帯 (/i/q/ta/i/)= 威張る (/i/ba/ru/)
+ 切手 (/ki/q/te/)
+ めでたい (/me/de/ta/i/)
+ 洪水 (/ko/o/zu/i/)

4.3 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 4 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，オピニオン評価と対比較実験を行う.

オピニオン評価では，自然に聞こえた度合を 5 段階 (5 が最も自然 1 が最も不自然) で評価するように指示する.

対比較実験では二種類の同じ音声を続けて流し，どちらの音声が自然に聞こえるかを判定する．本実験では自然音声，オリジナル合成，クラスタリング合成の三種類の音声を全ての組合せで評価する.

5 実験結果

5.1 オピニオン評価の実験結果

実験結果を表 2 に示す.

Table 3 オピニオン評価の結果

	オピニオンスコア 評価音節数：100
自然音声	4.90
オリジナル合成	4.31
クラスタリング合成	3.70

表 2 から，オピニオン評価で 3.7 という自然性の高い結果が得られた.

5.2 対比較実験の結果

以下の三通りの対比較実験を行った，その結果を表 3 に示す.

1) 自然音声とオリジナル合成
2) 自然音声とクラスタリング合成
3) オリジナル合成とクラスタリング合成

Table 4 対比較実験の結果

自然音声	オリジナル合成
79.25 %	20.75 %
自然音声	クラスタリング合成
90.75 %	9.25 %
オリジナル合成	クラスタリング合成
75.75 %	24.25 %

オリジナル合成とクラスタリング合成の差は自然音声とオリジナル合成との差ぐらいである，したがって自然性の高い音声であることがわかる.

6 考察

6.1 オピニオン評価の解析

波形接続型音声接続では，人手でラベリングされた情報を利用して音節素片を切り出す．しかし，ラベルに誤りがある場合がある. このとき余分な前後の音声が入ることがある.

オリジナルの波形接続型音声合成では前後の音素を考慮している為に多少前後の音があっても違和感無く音声が合成できる．しかし，木に基づくクラスタリングでは, グループ化された音素からランダムに音を選択している為に前後環境は考慮されていない，その結果，合成音声に違和感が生じ品質の低下に繋がったと考えている.

6.2 アクセントの位置

木に基づくクラスタリングでグループ化された音響パラメータが似た音節素片の中には，アクセントの高低が違うものがグループ化されている事があった．これは音質の低下を招くので本実験ではグループから除外した.

6.3 アクセントの高低

アクセントの高低は同じであるがアクセント位置が違う単語がある．例をあげると “約束” の “KU” と “隠す” の “KU” である．この二つの音節素片は木に基づくクラスタリングで音響パラメータが似ていると判断された，しかし，“隠す” は “KU” でアクセントが下がるために品質が劣化する場合がある.

7 まとめ

本実験では，波形接続型音声合成で任意の合成音声を作成するために音声が収録されているＤＢに木に基づくクラスタリングを行い，クラスタリングされた音節素片を使用した時の合成音声の音質を調査した.

聴覚実験の結果，オピニオンスコアで 3.7 が得られた．クラスタリングにより条件を緩和したことで波形接続型合成音声よりも多少オピニオンスコアが下がったものの自然性の高い音声が作成でき，任意の音が作成可能であることがわかった.

今後は，木に基づく状態共有にアクセント位置を考慮に入れるなどの検討を行い，より自然性の高い合成音声の作成を目指したい.

参考文献

[1] 石田 隆浩，村上 仁一，池原 悟. “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句への応用”，音響全体，2-Q-18, pp.1-409,410(2003-3).
[2] “NHK 日本語アクセント辞典 新版”，NHK 出版，ISBN4-14-011112-7(1998).
[3] 堀田 波星夫，村上 仁一，池原 悟. “不特定話者における同音異義語音声認識”，日本音響学会春季発表予定，NO. 2-1-11(2006).
[4] S.J. Young, J.J. Odell, and P.C. Woodland. Tree-based state trying for high accuracy acoustic modelling. Proc. ICASSP, PP.307-312(1994).