波形接続型音声合成の文節への適用*

○加藤 琢也，村上 仁一，池原 悟 (鳥取大・工)

1　はじめに

高い品質の合成音声を作成する方法として CHATR[1] がある．CHATR は合成したい話者の音声をあらかじめ録音しておき，そこから部分的に切り出した音声波形を信号処理をせずに接続して音声を合成する方法である．

そして、これと似た手法で単語の音声合成を行うのが音節波形接続方式 [2] である．音節波形接続方式は，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分割し，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をすることにより，話者性と高い自然性が保たれる特徴がある [2]．

一方，音節波形接続方式においては，韻律の扱いが問題となる．そこで，この問題の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．文献 [2] では地名を対象として実験を行い，実用的な品質が得られたことが報告されている．また文献 [3][4] において，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，さらにアクセント型を考慮することで，より自然音声に近い音声を作成可能であることが示されている．しかし，これらの研究は名詞を対象としたものであり，文や文節を対象とした研究は行われていない．

そこで本研究では，文節発声で発話速度が遅い音声を使用して，アクセント型を波形選択に利用した音節波形接続方式の文節に対する有効性の確認を行う．

2　音節波形接続方式

2.1　モーラ情報とアクセント情報

音節波形接続型方式では，韻律的な情報として，モーラ情報 (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 [2]．過去の研究では，モーラ情報は名詞の音声合成 [2][3][4] だけでなく，音素ラベリング [5] や単語音声認識 [6][7] などの分野においても効果があることが報告されている．

また，文献 [4] では音節波形接続方式を普通名詞に対して行う場合，アクセント型は波形選択において有効な情報であることが報告されている．

そこで，本研究では音節選択にモーラ情報とアクセント型の情報を利用した音声合成を文節に対して行う．

2.2　発話速度

文節発声で発話速度が遅い音声の場合には，区切ることでピッチが初期化されるため，文節の音声合成が名詞と同様に行えると考えられる．これにより，CHATR で使用されている ToBI モデルなどのような複雑な韻律情報を用いなくても品質の高い合成音声が作成できると考えた．

そこで，本研究では文節発声で発話速度の遅い音声を用いて文節の合成音声を作成する．

2.3　連続母音の扱い

本研究ではラベリングを人手によって行う．音声の中には連続母音の「エイ」や「オウ」のように音素境界が不明瞭な場合がある．今回はこのような連続母音に関しては，無理に切り分けず，一つの音素として扱う．

*Phrase Synthesis by Concatenating Syllabic Speech Synthesis. By Takuya Kato,Jin'ichi Murakami and Storu Ikehara (Tottori Univ)

2.4　音節波形接続方式による音声合成

普通名詞の場合にはアクセント型を考慮することによって，さらに品質が向上することが知られている [4]．そこで本研究においてもモーラ情報に加えてアクセント情報としてアクセント型を波形選択に使用する．

音節波形接続方式におけるアクセント型を考慮した波形選択による音声合成では，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)
- 文節のモーラ数
- 文節中のモーラ位置
- 文節のアクセント型

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

3　評価実験

3.1　実験環境

本研究では，データベースとして，複数の電子辞書から重文複文を抽出した日英対訳の例文集 (CREST コーパス [8]) の文を使用する．本来，この例文集は機械翻訳を目的としたものだが，日本語の文としては短く，音声合成に適していると思われる．そこで、この例文集に収録されている 1000 文を使用し, 文節発声で遅く発話した女性話者の音声を音声データベースとして用いる．そして，自然音声，アクセント型を考慮した合成音声，アクセント未考慮の合成音声のそれぞれの場合について 100 文節作成する．また，作成する文節は 4，5，6 モーラの文節とする．

3.2　評価方法

合成音声の評価は，音声研究に関わった経験のない人 5 名を対象に聴覚実験を行う．聴覚実験では了解度試験，オピニオン評価，対比較実験を行う．

まず，文節の明瞭性を調べるために了解度試験を行う．了解度試験では文の中に作成した文節を一つ埋め込んで行い，比較対象の文節がどのように聞こえたかを仮名で書き取らせる．

また，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文の中に作成した文節を一つ埋め込んで行い，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．

作成した音声の比較のために，対比較実験を行う．対比較実験では文節を文に埋め込んで行うのではなく，文節の音声のみで行う．対比較は自然音声とアクセント型を考慮した合成音声，そしてアクセント型を考慮した合成音声と考慮しない合成音声の 2 つの組み合せで行い，同じ文節で 2 種類の音声を続けて流し，どちらの音声が自然に聞こえるかを判定する．

3.3　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，括弧内の「_¯ ¯ ¯ ¯」はアクセントを表しており、太文字は合成の際に使用された音節である．

発言が (h**a**_/t**su**_¯/g**e**_/N_/g**a**_¯)
＝発音に (**ha**_/t**su**_¯/g**e**_/N_/n**i**_¯)
＋脱獄を (d**a**_/**tsu**_¯/g**o**_/k**u**_/o_¯)
＋失言が (sh**i**_/t**su**_¯/**ge**_/N_/g**a**_¯)

＋体験が (ta/i/ke/**N**/ga)
＋水準が (su/i/ju/N/**ga**)

政治家の (se-i/ji/ka/no)
＝誠実で (**se-i**/ji/tsu/de)
＋政治家は (se-i/**ji**/ka/ha)
＋アメリカに (a/me/ri/**ka**/ni)
＋横綱の (yo/ko/zu/na/**no**)

3.4 波形接続方式に関する補足

本研究では，情報が一致する音節候補が複数ある場合，データベースの上位のものから選択する．また，波形を切り出す位置は，前後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くように調整を行う．

4 実験結果

4.1 了解度，オピニオン評価の実験結果

了解度試験，オピニオン評価の実験結果を表 1 に示す．

表 1：了解度試験，オピニオン評価の結果

	了解度　正解率 (%)	オピニオンスコア
自然音声	99.6(498/500)	4.96
アクセントあり	98.4(492/500)	4.08
アクセントなし	96.2(481/500)	3.61

表 1 から，アクセント型を考慮した合成音声は了解度，オピニオンスコアともに高い値となっており，文節に対しても実用性のある合成音声が作成されたことが分かった．自然音声と比べると了解度は同程度であったが，オピニオンスコアの値には開きがあった．また，アクセント型未考慮の音声よりも考慮した音声の方が，了解度と自然性のどちらにおいても高い値となった．

4.2 対比較実験の結果

4.2.1 自然音声との対比較

自然音声とアクセント型を考慮した合成音声の対比較実験の結果を表 2 に示す．

表 2：自然音声との対比較

	自然音声 (%)	アクセントあり (%)
文節数 100	87.0	13.0

表 2 から，アクセント型を考慮した合成音声と自然音声との差はまだ大きいことが分かった．

4.2.2 アクセント型未考慮の合成音声との対比較

表 3 に波形選択にアクセント型を考慮した合成音声と未考慮の合成音声の対比較実験の結果を示す．

表 3：アクセント型未考慮との対比較

	アクセントあり (%)	アクセントなし (%)
文節数 69	74.2	25.8

表 3 より，アクセント型を考慮した場合の方が品質の高い音声であると判定されており，文節を対象とした場合にもアクセント型を利用することで品質が向上することが示された．

5 考察

5.1 不自然な音声の解析

オピニオン評価で評価が低かった音声を調べると大きく分けて二つの種類があった．一つは「ン」につながる部分での不自然さであり，もう一つは今回は音素境界が明瞭だとした「アイ」などの連続母音の部分での不自然さである．

これらの音素は前後とのつながりが強い部分であり，音量や音の高さの違いが大きく自然性の低下に結びついたのではないかと考えられる．

そこで，これらの連続母音に関しても一つの音素として扱うことで改善ができるのではないかと考えられる．

5.2 アクセントに関する考察

文節は普通名詞よりもアクセントの多様性があり，複雑だと考えられるが，本研究ではアクセント情報としてアクセント型のみを使用した．アクセント型に加えてアクセント核の情報も波形選択に使用してみたところ，今回作成した 100 文節全てにおいて同じ波形が選択され，アクセント核の情報は意味をなさなかった．

この結果から，波形接続方式を文節に適用した場合にも，普通名詞の場合と同様にアクセント型のみを考慮すれば十分であると思われる．

5.3 データベースの音量のばらつき

普通名詞を対象に行われた研究 [4] でも指摘されていたように，本研究でも音量のばらつきが問題となり，自然性が損なわれることがあった．とくに，今回の実験では作成した音声を文の中に挿入して実験を行ったため，作成した音声自身の不自然さだけでなく，同一文内の文節との音量の違いによる不自然さも表われた．

音量のばらつきについては，録音された時間の近い波形から優先に選ぶことで多少改善ができると考えられる．

5.4 波形選択に関する考察

今回の実験では複数の候補があった場合に，最初に出てくる候補を利用して合成音声を作成した．しかし，音節候補は複数あるため，候補の絞り込み手法を考えることでさらに品質の高い音声の作成が可能である．

そこで，波形候補をさらに絞る手法としては，参考文献 [3] で提案されているような継続時間による絞り込みや品詞での絞り込みが考えられる．

5.5 発話速度に関する考察

通常の速度で文節発声している ATR の単語発話データベースの中の DSB を用いて同様の実験を行った．その結果，今回の実験で得られた音声に品質では及ばなかった．これは通常の発話速度で発話した音声では，文節の区切りでピッチが初期化しきれず，アクセントのばらつきがあったためだと考えられる．したがって，今回のような波形撰択を通常の発話速度の音声に適用することは困難だと考えている．

6 まとめ

本研究では，文節発声で発話速度が遅い音声を用いたときの音節波形接続方式の文節における有効性を調査した．聴覚実験においてアクセント型を考慮した合成音声は了解度が 98.4 %，オピニオンスコアは 4.08 が得られ，文節を対象とした場合にも音節波形接続方式が有効であることが分かった．また，アクセント型を考慮しない音声と比べると，明瞭性と自然性の両方で高い値となり，アクセント型は品質が高い音声を得るために有効な情報であることが分かった．

一方，自然音声の了解度は 99.6 %，オピニオンスコアは 4.96 であり，対比較実験では 87.0 %が自然音声の方がいい音だと判定された．合成音声も了解度とオピニオンスコアでは高い値を得たが，自然音声と比較すると，その品質の差はまだ大きいことが分かった．

今後は，波形候補の絞り込みの手法やデータベースの音量の問題についての検討を行い，さらに品質の高い合成音声の作成を目指したい．

参考文献

[1] N.Campbell and A.Black”CHATR:自然音声波形接続型任意音声合成システム”，信学技法, SP96-7,pp45-52 (1996-05).

[2] 村上, 水澤, 東田，” 音節波形接続による単語音声合成”, 信学技報, SP99-2, pp.45-52 (1999-05).

[3] 石田, 村上, 池原，” 音節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 (2002-05).

[4] 石田, 村上, 池原，” 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”, 音響論, 2-Q-18, pp.1-409,410 (2003-03).

[5] 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”，信学技法，SP2001-53,pp.25-30(2001-08).

[6] 妹尾, 村上, 池原，” モーラ情報を用いた単語音声認識の検討”，信学技法，SP2002-130,pp.55-61(2002-12).

[7] 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”，信学技法，SP2002-131,pp.63-68(2002-12).

[8] 村上，池原，徳久，” 日本語英語の文対応の対訳データベースの作成”,「言語，認識，表現」第 7 回年次大会，(2002-12)