波形接続型音声合成のフレーズへの適用

居村太介　村上仁一　池原悟
鳥取大学 工学部 知能情報工学科
{s042011,murakami,ikehara}@ike.tottori-u.ac.jp

1　はじめに

音声合成は古くから規則合成方式が研究されてきた．しかし，規則合成方式は音声を合成するときに信号処理が必要である．これが品質を低下させる原因になっている．

そこで音質劣化の原因とされる信号処理を出来るだけ使用しない CHATR[1] が提案された．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接続して音声を合成する方法である．しかし，この CHATR は音素波形を選ぶ際に信号処理を使用するため，最良の波形が選択されない場合がある．

そこで，CHATR とよく似た手法として音節波形接続方式 [2] が提案されている．この手法は，言語情報のみを利用して音節波形を選択する．そのため，音声を合成する際に信号処理を一切使用しない．

この手法は，過去の研究において，固有名詞，普通名詞，文節 (短文節) を対象として行われた．その結果，品質の高い合成音声が得られることが報告されている [3][4][5].

しかし，この手法は文節 (短文節) における音声合成において 1 話者のみでしか行われていない．またフレーズ (長文節) における有効性が確認されていない．そこで本研究は，この 2 つの問題点について調査を行う．

2　音節波形接続方式

2.1　音節波形接続方式による音声合成

音節波形接続方式は，波形編集型の音声合成方式の一種で，音響パラメータを使用しないで，言語的なパラメータのみで音声合成を作成する．具体的には，音節波形接続方式で文節を合成する際，収録された大量の音声データベースから以下の情報が一致する音節素片を選択する．

表 1　音節素片を選択する際の条件

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)
文節のモーラ位置
文節のモーラ数
文節のアクセント型

最後に，文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．

2.2　連続母音の扱い

音節波形接続方式で作成された合成音声は，音節素片の接続部に違和感が生じる時がある．特に母音や撥音が連続する部分で違和感が出やすい．これらの音節素片は前後の音が連続的に変化する部分であり，音節境界が明確ではない．そのため，これらの音節素片を繋げると自然性を損なうことがある．

そこで，母音や撥音や促音が連続する場合，連続母音として扱う．ただし，母音や撥音が多数連続する場合があるため，1 つの連続母音として扱うのは最大で 2 音素までとする．

2.3　波形接続方式に関する補足

波形接続方式は，接続部の違和感が自然性に大きく影響する．そのため，接続部における 2 素片間の波形の位相を考慮し，接続部の振幅の差がゼロに近づくように接続する．

2.4　音量の差

音節波形接続方式では信号処理を一切使用しないため，合成音声に使用する音節素片の音量の差が音声の品質に直接影響する．そこで，合成に使用する音節素片は，録音した時間帯が近い音節素片の組合せを選び，合成音声を作成する．

2.5　アクセント

本研究では，アクセントの高低を”NHK 日本語発音アクセント字典”[6] を参考に著者が決めた．

3　音声合成の評価方法

音声合成の評価方法にはいくつか種類があるが，本研究ではオピニオン評価と対比較実験の二種類を行う．そして，この 2 つの聴覚実験を音声研究に関わった経験のない 5 名で行う．具体的な方法は以下に示す．

(1) オピニオン評価

音声の自然性を調べるために，オピニオン評価を行う．オピニオン評価は，自然音声の文の中に比較対象となる合成音を一つ埋め込み，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．

(2) 対比較実験

作成した音声の評価のために，自然音声との対比較実験を行う．対比較実験では合成音を文に埋め込んで行うのではなく，比較対象の音声のみで行う．自然音声と合成音声の同じ内容の 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．

4　話者の依存性

過去の研究において，音節波形接続方式の文節への適用は 1 話者でしか行われていない．そこで本研究では，音節波形接続方式の文節への適用における話者の依存性について調査する．なお，実験条件は，“波形接続型音声合成の文節への適用”[5] の加藤らの実験と同一とする．

4.1　合成に使用する日本語文

日本語文の例として，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 [6]) の文を使用する．この例文集に収録されている 1000 文を使用し，一般の男性話者に文節発声で遅く発話してもらう．この収録した音声を用いて，4，5，6 モーラの文節を 100 文節作成する．

以下に文節発声で収録した日本語文の例を示す．なお，括弧内の「-」はポーズを表している．

(1) 彼女が-学校を-休んだので-がっかりした
(2) 新政権に-反対し，-市民が-暴動を-起こした
(3) 彼は-知らないうちに-その病気に-かかっていた

4.2　文節の合成音声の例

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．

(1) 聞かれて (ki‿/ka¯/re¯/te)
=危険が (**ki‿**/ke¯/N/ga¯)
+ 光の (hi‿/**ka¯**/ri/no¯)
+ 言われた (i‿/wa¯/**re**/ta¯)
+ 重ねて (ka‿/sa¯/ne¯/**te¯**)

(2) がっかりした (ga/q/ka¯/ri¯/shi/ta)
=がっしりした (**ga**/q/sha¯/ri¯/shi/ta)
+ しゃっきりした (sya/**q**/ki¯/ri¯/shi/ta)
+ しっかりした (shi/q/**ka¯**/ri¯/shi/ta)
+ うんざりした (uN/za¯/**ri¯**/shi/ta)
+ いきいきした (i/ki/hi¯/ki¯/**shi**/ta)
+ 専念した (seN/ne¯N/shi/**ta**)

4.3　オピニオン評価の実験結果

オピニオン評価の実験結果を表 2 に示す．

表 2　オピニオン評価の結果

評価者	自然音声	合成音声
a	4.95	4.48
b	4.96	4.43
c	4.74	4.47
d	4.99	4.75
e	4.96	3.94
平均	4.85	4.41

表 2 より文節単位の合成音声は，4.41 という高い値を得られた．

4.4　対比較実験の結果

自然音声と文節単位の合成音声の対比較実験の結果を表 3 に示す．

表 3　対比較実験の結果

評価者	自然音声 (%)	合成音声 (%))
a	79	21
b	63	37
c	83	17
d	72	28
e	70	30
平均	73.4	26.6

表 3 より文節単位の合成音声は ，自然音声には及ばないが，26.6% の文節について合成音声の方が自然に聞こえると判定された．

4.5　考察

加藤ら [5] の音声合成の実験において，オピニオンスコアが自然音声で 4.75，合成音声で 3.83 となっている．また対比較実験において，25.7% の文節について合成音声の方が自然に聞こえると報告されている．

この加藤ら [5] の実験と本研究の実験が，ほぼ同一の結果が得られたことから文節発声における音節波形接続方式は話者の依存性は少ないと考えている．

なお，加藤ら [5] の実験で使用された発話者は，女性のプロのナレーターであり，本来は一般男性話者を使用した合成音声よりも品質が良くなると思える．だが，本研究の実験結果において，自然音声，合成音声ともにオピニオンスコアが高いことや，対比較実験での結果が加藤ら [5] の実験結果と近いことから，評価者の差によるものだと考えている．

5　フレーズ (長文節) への適用

音節波形接続方式は，フレーズ (長文節) においての有効性が未だ確認されていない．そこで本研究では，フレーズにおける音節波形接続方式の有効性について調査する．

なお，本研究においてのフレーズは，単文では主部と述部で，重・複文では 2 つの単文の主部と述部である．したがってフレーズの境界は，文発声における息継ぎに相当する．以下に具体例を示す．なお，括弧内の「-」はポーズを表している．

(1) 単文
　学校は-家より遠い
(2) 重文
　群集を退散させるために-警察を呼んだ
(3) 複文
　たばこが-健康に悪影響を及ぼすというのは-定説になっている

5.1　フレーズにおける音節素片の選択

本来，音節波形接続方式では，2.1 節の表 1 で挙げた条件で音節素片の選択をするのが望ましい．しかし，フレーズのモーラ位置とアクセント型が一致する音節素片が収録された音声データベースを作成するには，莫大な録音時間が必要となる．

そこで本研究は，フレーズ単位の合成において，モーラ位置とアクセント型に関する条件をフレーズ中の文節とする．具体的には，2.1 節の表 1 と同じにし，音節素片を選択する．

5.2　フレーズの合成音声の例

本研究で作成したフレーズの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，「|」は文節の区切りを表しており，太文字は合成の際に使用された音節である．

(1) 本性を｜現した (hōN/syou/wo ｜ a/rā/wa/shi/ta)
　=本州が｜現れた (**hōN**/syuu/ga ｜ a/ra/wa/re/ta)
　+ 懸賞を｜諦めた (kēN/**syou**/wo ｜ a/ki/ra/me/ta)
　+ 柔道を｜諦めた (jūū/dou/**wo** ｜ a/ki/ra/me/ta)
+ 居ずまいを｜改めた (i/zu/mai/wo ｜ **a**/ra/**ta**/me/ta)
+ 怪獣が｜現れた (kai/juu/ga ｜ a/**ra**/wa/re/ta)
+ 肘鉄を｜くらわせた (hi/ji/te/tu/wo ｜ ku/rā/**wa**/se/ta)
+ 聖水を｜汲みだした (se/i/su/i/wo ｜ ku/mī/da/**shi**/ta)
+ 平民を｜支配した (he/i/mi/N/wo ｜ shi/hāi/shi/**ta**)

(2) 仮説を｜検証した (ka/se/tū/wo |¯ kēN/syou/shi/ta)
　=火災を｜体験した (**ka**/sāi/wō ｜ tai/keN/shi/ta)
　+ 左折を｜強要した (sa/**se**/tū/wō ｜ kyou/you/shi/ta)
　+ 砂鉄を｜強奪した (sa/te/**tū**/wō ｜ gou/da/tu/shi/ta)
　+ 個室を｜契約した (ko/shī/tu/**wo** ｜ kei/ya/ku/shi/ta)
　+ 神社を｜建設した (jiN/ja/wo ｜ **kēN**/se/tu/shi/ta)
　+ 神社が｜延焼した (jiN/ja/ga ｜ eN/**syou**/shi/ta)
　+ 不満を｜解消した (fu/maN/wo ｜ kāi/syou/**shi**/ta)
　+ 部隊を｜編成した (bu/tai/wo ｜ heN/sei/shi/**ta**)

5.3　合成に使用する日本語文

本研究では，日英対訳の例文集 (CREST コーパス [6]) の文と人手で作成した例文を用いる．この例文を使用し，一般男性にフレーズ単位で発話してもらう．この収録した音声を用いて，4 モーラから 14 モーラのフレーズを 100 文用意する．なお，本節で使用した一般男性話者は，4 節の実験で使用した男性話者と同じである．

5.4　オピニオン評価の実験結果

オピニオン評価の実験結果を表 4 に示す．

表 4　オピニオン評価の結果

評価者	自然音声	合成音声
a	4.55	3.16
b	5.00	3.83
c	4.55	3.90
d	4.99	4.24
e	4.32	3.36
平均	4.68	3.71

フレーズの合成音声は, オピニオンスコアが 3.71 と高い結果が得られた．しかし，文節の合成音声と比べると

かなり値が低くなった.

5.5 対比較実験の結果

自然音声とフレーズの合成音声の対比較実験の結果を表 5 に示す.

表 5 対比較実験の結果

評価者	自然音声 (%)	合成音声 (%))
a	97	3
b	94	6
c	88	12
d	93	7
e	93	7
平均	93.0	7.0

表 5 よりフレーズの合成音声は，自然音声には及ばないが、7 %のフレーズについて合成音声の方が自然に聞こえると判定された.

5.6 フレーズへの適用の考察

5.6.1 評価の低い音声

オピニオン評価の低かった音声を見ると，合成するフレーズのモーラ数が多いほど品質が低下する傾向にあった．本研究は，フレーズのモーラ位置やアクセント型を一致させるのは困難だと考え，フレーズ中の文節ごとにモーラ位置とアクセント型を一致させた．そのため，自然性が低下したと考えている.

なお，フレーズのモーラ位置とアクセント型を一致させた条件下で合成音声を作成していないため，どの程度品質に差が生じるのかが不明である．そこで今後の課題としては，フレーズでモーラ数やモーラ位置を揃えて実験を行いたい.

5.6.2 評価のばらつきについて

今回，5 人の被験者でオピニオン評価を行った．合成音声において一番高い結果を出した人で 4.03，一番低い結果を出した人で 3.16 と評価が分かれた．しかし，各被験者における自然音声とフレーズの合成音声のオピニオンスコアの差に違いがあまり見られない．また, 対比較実験の結果からも合成音声の品質が高いと考えている.

5.6.3 データベースの音量のばらつき

従来手法では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．今回の実験では音量のばらつきを抑えるために，全て同時期に収録した音声を選んで音声を作成した．その結果，音量のばらつきは少なくなり品質の高い音声の作成ができた.

しかし，完全に音量の統一はできず，不自然さが残る音声があった．特に「が，を，に」などの助詞の音節素片に音量のばらつきが見られた．これは文中における第 2 アクセントなどが原因であると考えられる．文の内容によっては，名詞や動詞の後の助詞が強調される場合があり，その強調された助詞によって音量のばらつきが生じた．今後，助詞の音節部分の品質を上げるには，接続部分の音量が同程度の音節素片を使用する手法が考えられる.

6 まとめ

本研究では，文節発声の話者依存性と，フレーズにおける音節波形接続方式の有効性について調査した．聴覚実験において，文節発声はオピニオンスコアが 4.41，対比較実験でも 26.6% が合成音声の方が良い音だと判定された．このことから, 話者の依存性は少ないと言える.

また音節波形接続方式をフレーズへ適用して作成した音声合成は，オピオンスコアが 3.71，対比較実験で 7% が良い音だと判定された．このことから，音節波形接続方式のフレーズへの適用は，文節発声の合成音声と比較すると自然性では劣るが，品質の高い合成音声が作成できることが確認された.

今後は，フレーズのモーラ位置やアクセント型を揃えた合成音声の作成と，音節部品を選ぶ際にどこまで言語的なパラメータの条件を緩和して良いかを調査していく.

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”, 音響全体, 2-Q-18, pp.1-409,410 (2003-03).

[5] 加藤, 村上, 池原，” 波形接続型音声合成の文節への適用”, 音響全体,

[6] 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)

[7] NHK 放送文化研究所，”NHK 日本語発音アクセント字典”, (1998)