社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# 音節波形接続型音声合成の文節への適用

村上　仁一†　　加藤　琢也†　　池原　　悟†

† 鳥取大学工学部，鳥取県
E-mail: †murakami@ike.tottori-u.ac.jp

**あらまし**

音節波形接続型音声合成は，波形編集型合成音声の 1 種類である．特徴として，言語的なパラメータのみを利用して波形の選択している．この音声合成方法は，過去に固有名詞と普通名詞を対象に音声合成を行っている．この結果，実用的な品質が得られたことが報告されている [2] [3]．そこで本研究では，音節波形接続方式を，文節に適用して有効性を調査した．文節は名詞に比べ韻律が複雑である．そのため，文節発声でゆっくりと発話された音声を対象にした．

また，品質向上のために，従来の音節波形接続方式に音節選択の条件を 2 つ追加した．1 つ目として，接続部において不自然さの軽減のために，母音と撥音が連続する部分を連続母音として扱った．また 2 つ目として，条件を満たす素片の中から録音した時間が近い音声を選んで音声を作成した．

作成した合成音声の品質を調査した結果，聴覚実験における合成音声の了解度は条件を追加していない場合が 98.7%，条件を追加した場合が 99.3%となった．これは自然音声の 99.3%と比べても同程度の高い値となった．また，オピニオンスコアは条件を追加しないものが 3.55 であったのに対し，条件を追加した場合は 3.83 となった．そして，対比較実験においても，60.7%が条件を追加した音声の方が自然だと判定され，条件を追加することが自然性の向上に有効であることが分かった．

一方，自然音声はオピニオンスコアが 4.75 であり，条件を追加した合成音声との対比較においても 74.3%が自然音声の方が自然だと判定された．自然性の面では合成音声は自然音声には及ばなかったが，高い品質の合成音声が作成可能であることが示された．

**キーワード**　単語音声合成，録音編集方式，音節波形接続，韻律的特徴，モーラ位置，モーラ数，アクセント型

# Phrase Speech Synthsis using Syllable Concatenative Method

Jin'ichi MURAKAMI†, Takuya KATO†, and Satoru IKEHARA†

† Department of Information and Knowledge Engineering Faculty of Engineering Tottori University
4-101, Minami Koyamachou, Tottori city, 680-8552 Japan
E-mail: †murakami@ike.tottori-u.ac.jp

**Abstract**　The "Concatenating Syllabic Components based on Positional Features and Mora Information and Accent (CSCPMA)" is a kind of corpus-based concatenative speech synthesis method. The common word synthesis using CSCPMA was already studied and showed validity to proper. The features of CSCPMA is that only the language information is used to select the wave form. So, in this paper, we applied the CSCPMA to phrase speech synthesis. The phrase speech synthesis is more complicated than the common word speech synthesis. So, we chose phrase speech with the very slow speed. And to achieve high quality voice, we added the two conditions to select wave forms. One was that double vowel. We treated the all continuous vowel as double vowel. The another was that we chose the wave form to near recoding time. As the result of experiments, we obtained the 99.3% for intelligibility and 3.83 for naturalness. For natural speech, we obtained the 99.3% for intelligibility and 4.75 for naturalness. As the result, we obtained high quality speech synthesis, although not exceed the natural speech.

**Key words**　speech synthesis, slot filling, mora position, mora length, accent, prosodic features

## 1. はじめに

現在，カーナビゲーションシステムや電車の車内アナウンスなどのように，音声ガイダンスを利用したシステムやサービスが様々な場面において利用されている．このようなシステムでは，録音編集方式が広く使われている．録音編集方式では，まず，システムやサービスに必要となる音声を，システム利用者の入力やサービスの利用される場所・時間などに依存するような比較的短い単語音声（以下,「可変部」）と，それ以外の比較的長い文節・文音声（以下,「固定部」）に区別する．そして，可変部と固定部を別々に録音しておき，必要に応じて組み合わせることで出力音声を構築する．例えばカーナビゲーションシステムにおいて,「目的地は○○○よろしいですか.」というガイダンス音声を出力したい場合，○○○の部分には，駅名や建物名などの単語音声と「で」などの助詞が挿入される．ユーザーが目的地に「東京駅」を指定した場合，ガイダンス文は「目的地は”東京駅で”よろしいですか.」となる．例の場合,「東京駅で」などの駅名や建物名などの単語音声と助詞が可変部,「目的地は～」という部分が固定部となる．

録音編集方式を用いた音声合成においては，可変部と固定部を接続した場合の違和感を軽減するために，一般に同一話者の音声が必要となる．可変部と固定部を分離して録音することにより，必要となるすべての音声を録音する場合に比べて話者に対する負担は若干軽減されるが，可変部に挿入する文節が増大した場合，同一話者から全ての音声を録音することは困難となる．さらに，録音環境の違いにより発話速度や $F_0$ 周波数にばらつきが出るため，安定した品質の音声を得ることは非常に困難となる．そこで，可変部や固定部に必要になる音声をすべて音声合成によって作成する方法が考えられる．

音声合成には古い歴史がある．そして多くの種類が提案されている．例えば，規則音声合成は，古くから TTS 音声合成において用いられてきた方法であり，基本的には，音声の特徴をパラメータとして抽出し，変形することによって合成音声を作成する．PSOLA 方式による音声合成については，現在も多くの研究がなされている．また，最近では HMM を用いて直接音声を合成する研究も行われている [5] [6]．しかし，いずれの場合においても，直接人の声を録音した自然音声のような高い品質を得ることは困難である. [7].

ところで，一般に音声信号に信号処理を加えた場合，自然性が劣化する．そこで信号処理を行わない音声合成方法が提案されている [1]．その代表的な手法が CHATR [9] である．CHATR は，あらかじめ合成したい話者の音声を録音しておき，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．

音節波形接続方式 [2], [3] は，CHATR と似た手法である．合成音声の対象は基本的に単語である．始めに，人名地名などの固有名詞を対象とした [2]，次に一般名詞を対象にした [3]．この方式は，あらかじめ録音しておいた音声波形を，音節単位で分割し，接続することで合成音声を作成する．CHATR との大きな違いは，言語的なパラメータのみを利用して波形を選択している点である．

一般に音声合成において，韻律制御は重要な課題である．そして，ToBI モデルや藤崎モデルなどの韻律モデルが使用されている [10]．しかし，地名などの固有名詞の合成音声の場合では $F_0$ 周波数のばらつきが比較的小さく，アクセント型がほぼ一意に決まるため，$F_0$ 周波数とモーラ情報の依存関係を効果的に利用できる [2]．そして，固有名詞を対象とした実験では，実用的な品質が得られたことが報告されている．また，普通名詞に適用した場合も，モーラ情報とアクセント型を考慮することで，より自然音声に近い合成音声の作成が可能である [3] [4].

本研究では，文節に対して音節波形接続方式を適用し，有効性の確認を行う．文節は名詞に比べて韻律が複雑になるため，通常の発話の音声では音声合成が困難だと考えられる．そこで，本研究では文節発声で発話速度が遅い音声の合成を試みる．また，作成した合成音声の問題点から，音声波形の選択条件を追加し，より自然音声に近い合成音声の作成を目指す.

## 2. 音節波形接続方式

### 2.1 音節波形接続型音声合成の概説

本研究で用いる音節波形接続型音声合成は，始めに，表 1 の情報が一致する音節素片を選択する．次に，情報が一致する音節候補の中から，データベースの上位の音節素片を選択し，音節の開始時間と終了時間（ラベリングデータ）から波形データを切り出す．最後に，選択された波形データを接続して合成音声を作成する．なお単語を合成する場合は，表 1 中の文節が単語になる.

表 1　音節波形接続型音声合成の音節素片の選択条件

| | |
|---|---|
| 1 | 音節 |
| 2 | 直前の音素 (前音素環境) |
| 3 | 直後の音素 (後音素環境) |
| 4 | 文節中のモーラ位置 |
| 5 | 文節のモーラ数 |
| 6 | 文節のアクセント型 |

### 2.2 韻律（$F_0$ 周波数）とモーラ情報（モーラ位置とモーラ長）とアクセント型

一般に音声合成においては，韻律の扱いは困難である．韻律を扱う場合，録音音声および出力音声の $F_0$ 周波数が必要となる．しかし，正確な $F_0$ 周波数を直接推定することは困難である．実際には ToBI モデルや藤崎モデルなどが利用されている．また，最近では HMM を用いる研究も行われている [5] [6].

しかし，音声合成の対象として地名などの固有名詞を選んだ場合，固有名詞では，アクセント型がほぼ一意に決まる．そのため $F_0$ 周波数と単語のモーラ位置と単語のモーラ長（以後モーラ情報）の依存関係を効果的に利用することが可能である [2]．そして，このモーラ情報は音素ラベリング [11] や音声認識 [12] [13] などの分野において効果があることが報告されている.

一般的な普通名詞では「雨」と「飴」のように同音異義語が多数現れるため，モーラ情報を考慮しただけでは不適切な音節

素片が選択される場合がある．そこで，音節素片の選択においてモーラ情報に加えてアクセント型を加えることで，非常に自然性の高い合成音声が得られることが示されている [4].

そこで本研究では，文節を対象とした場合に音節素片の選択にモーラ情報とアクセントを考慮することで，どの程度の合成音声の品質が得られるかを調査する．なお，文節のアクセント型については，NHK 日本語発音アクセント辞典 [14] を利用する．

### 2.3 発話速度

文節の発話は名詞のみの発話と比べて韻律が複雑となる．しかし，音節波形接続方式では信号処理を行わない．そのため，音節素片に情報を付加することで韻律の問題を解決する．

通常，文の音声合成を波形接続方式で行う場合には，ToBI モデルや藤崎モデルなどの複雑な韻律モデルが使用されている．しかし，文節発声で発話速度が遅い音声を用いる場合には，文節間で区切ることでピッチが初期化される．そのため，文節においても名詞の場合と同じように扱うことができ，ToBI モデルや藤崎モデルのような複雑な韻律モデルを使用しなくても合成音声の作成ができると考えられる．そこで，本研究では文節発声で発話速度が遅い音声を用いて文節の合成音声を作成する．

### 2.4 波形接続に関する補則

音節波形接続型音声合成では，接続部の違和感の発生が自然性に大きく影響する．本研究では，接続部における音節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調整を行う．具体的には，あらかじめラベル付けされた音節素片の開始時間と音節素片の終了時間をもとに，振幅が負から正に変わる部分を，波形が短くなる方向 (開始時間は進む方向，終了時間は戻る方向) に探し，抽出する位置を修正する．

### 2.5 合成音声の例

本研究で作成する合成音声「むかって (/mu/ka/q/te/)」および「銀行の (/gi/N/ko/u/no/)」についての例を下に示す．なお，「‾」は音の強弱（アクセント）を表している．強調部は，実際に選択される部分を示している．

むかって　(/mu/ka/q/te/)

＝ 昔の　(/mu/ka/shi/no/)
＋ 使って　(/tsu/ka/q/te/)
＋ 当たった　(/a/ta/q/ta/)
＋ 終わって　(/o/wa/q/te/)

銀行の (/gi/N/ko-u/no/)

＝ 銀行に　(/gi/N/ko-u/ni/)
＋ 深刻に　(/shi/N/ko/ku/ni/)
＋ 天候に　(/te/N/ko-u/ni/)
＋ 民謡の　(/mi/N/yo-u/no/)

## 3. 評価実験

### 3.1 収録した音声

本研究では合成音声の対象として，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 [15]) の文を使用する．この例文集は機械翻訳を目的にしたものである．この例文集に収録されている 1000 文を，女性話者 (プロのナレータ) に文節発声で遅く発声して収録した．収録した音声の発話の一部を表 2 に示す．表中の"-"は文節の区切りであり，収録時にポーズをいれて発声した．

表 2　収録した文節発声の音声の一部

| 番号 | 文例 |
|---|---|
| 1 | その男は-追いつめられて-本性を-現した |
| 2 | 彼は-長い間-政界を-歩いてきた-人だ |
| 3 | 一生に-一度-あんな-大きな-伊勢えびを-食べてみたい |
| 4 | あの二人は-親の-反対を-押し切って-結婚した |
| 5 | 彼は-わたしの-言ったことを-好意的に-解釈した |
| 6 | 彼女が-学校を-休んだので-がっかりした |
| 7 | 政府は-関税を-撤廃して-貿易を-自由化しようと-している |
| 8 | 彼女の-気迫に-満ちた-演説に-満場の-人々が-感動した |
| 9 | 事態が-急転して-両国は-戦争に-突入した |
| 10 | クーデターに-成功して-軍部が-政権を-掌握した |

### 3.2 合成する文節音声の発話内容

合成する文節の発話内容は，音節波形接続方式で作成した合成音声と，同一の発話内容の自然音声があるように決める．表 2 中の文節音声から，表 1 の条件が一致する 4，5，6 モーラの文節について，計 100 文節の音声合成を行う．音声合成を行った 100 文節は，それぞれのモーラごとの作成可能な文節数の割合から表 1 のように定める．

表 3　実験に使用する文節のモーラごとの内訳

| モーラ数 | 文節数 |
|---|---|
| 4mora | 17 |
| 5mora | 70 |
| 6mora | 13 |

### 3.3 評価方法

合成音声の評価のために，音声研究に関わったことのない 9 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聴かせ，了解度試験，オピニオン評価，対比較実験の 3 つの実験を行う．

（1） 了解度試験

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，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．自分の知識を用いず，聞こえたとおりに書き取るように指示する．

（2） 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．

（3） 対比較実験

作成した音声の評価のために対比較実験を行う．そして，対比較実験は同じ内容の自然音声と合成音声の文節を続けて聞かせ，どちらの音声が自然に聞こえたか判定する．

なお，了解度試験とオピニオン評価では，比較対象となる文節を文節発声された自然音声の文の中に埋め込み，比較対象の文節のみを評価対象とする．了解度試験とオピニオン評価で使用する文の例を下に示す．なおアンダーラインを引いた文節が埋め込み箇所である．

(例) 全部員が 優勝を 目指して 練習に 励んでいる

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．

## 4. 実験結果

### 4.1 了解度試験の実験結果

了解度試験の結果を表 14 に示す．括弧内は正解個数/調査個数である．

表 4　了解度試験の実験結果

| | 了解度　正解率 (%) |
|---|---|
| 自然音声 | 99.3(894/900) |
| 合成音声 | 98.7(889/900) |

合成音声の了解度は 98.7%と高い値が得られ，明瞭な音声が作成できたことが分かる．また，自然音声と合成音声には差が殆んどない．聞き間違えた音声を見てみると，多くの場合が発音方法の似ている子音の聞き間違えであった．

### 4.2 オピニオン評価の実験結果

オピニオン評価の全被験者の平均を表 5 に示す．

表 5　オピニオン評価の実験結果

| | オピニオンスコア |
|---|---|
| 自然音声 | 4.75 |
| 合成音声 | 3.55 |

合成音声のオピニオンスコアは 3.55 となり，高い品質の合成音声が作成できたことが分かる．しかし，自然音声が 4.75 となっており，自然性の面で自然音声との差がある．差のある文節を調査すると，素片ごとの音量の違いがオピニオンスコア低下の原因となっている場合が多かった．また，スムーズに次の音素に移行していく部分での，微妙な声質の違いによる違和感もオピニオンスコアに影響を与えていた．

### 4.3 対比較実験の実験結果

自然音声との対比較実験の結果を表 6 に示す．

表 6　対比較実験の実験結果

| | 自然音声 (%) | 合成音声 (%) |
|---|---|---|
| 文節数 100 | 82.7 | 17.3 |

対比較実験では合成音声の方が良い音声だと判定された文節が 17.3%であった．この結果から品質の高い合成音声が作成されているが，自然性の面では自然音声との間にまだ差があるということが分かる．

## 5. 音節波形接続方式の問題点

### 5.1 接続部の違和感

音節波形接続方式で作成した合成音声は音節素片の接続部の違和感が問題となる．特に違和感を感じるのは母音や撥音が連続する部分である．違和感を感じた例として図 1，2 に「膨大な」の自然音声と合成音声のスペクトラムと音声波形を示す．この文節では「アイ」の部分に違和感が感じられた．波形データおよびスペクトログラムの出力には Wavesurfer [16] を使用する．

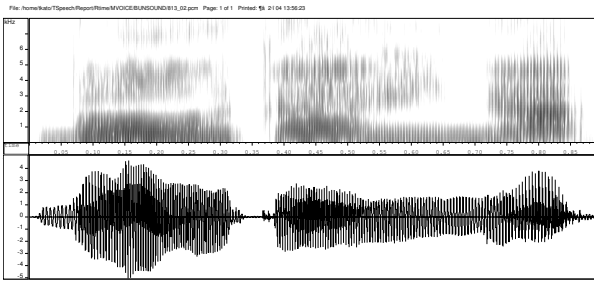

図 1　「膨大な」(自然音声)

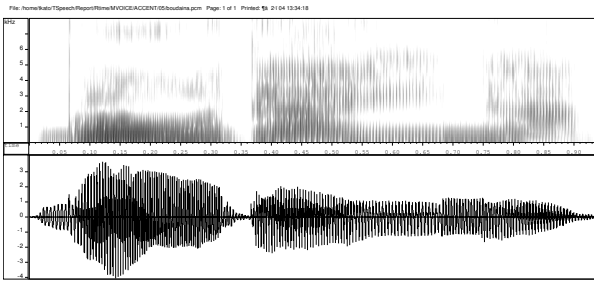

図 2　「膨大な」(合成音声)

これらの音素はスペクトルを見ると連続的に変化している．したがって音素境界位置が不明確である．そのため，音節素片の接続部において違和感が出やすい．これが，自然性の低下に結びついていると考えている．

### 5.2 データベースの音量のばらつき

作成した合成音声の中には音節素片ごとに音量のばらつきがあった．音節波形接続方式には録音された音声が大量に必要となる．そのため，録音された音声によって音量のばらつきが出る．音節波形接続方式は録音した音声に信号処理を加えないため，音量のばらつきが作成した合成音声に反映する．例として今回作成した音声の中では「部長の」がある．この「部長の」という音声では，最後の「の」に使用した「不況の」の音声が他の音節素片よりも音量が大きく，最後の「の」が強調されて違和感が現われた．

図 3,4 に「部長の」の自然音声と合成音声のスペクトラムと音声波形を示す．

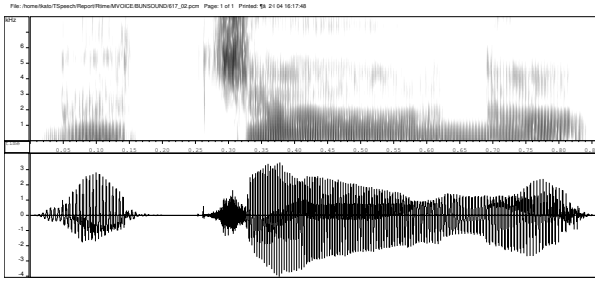

図 3 「部長の」(自然音声)

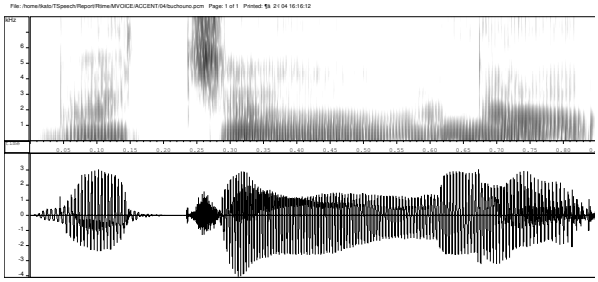

図 4 「部長の」(合成音声)

## 6. 音節波形接続型音声合成法の改良

本章では音節波形接続型音声合成法の改良のために 2 つの方法を提案する.

### 6.1 連続母音の拡張

前述の実験では「アア」「イイ」「ウウ」「エエ」「オオ」「エイ」「オウ」のみを連続母音として扱って合成音声を作成した. しかしラベリングしてみたところ, 文節では, 母音連続の多くが, 音素境界が明瞭ではなかった. そこで改良方法として, 母音と撥音が連続する部分を全て連続母音として扱う.

ただし, 母音や撥音が連続する場合がある. 例えば,「病院を (byo/u/i/N/o)」という文節の場合には, "ouiNo"の連続母音がある. そこで母音や撥音が連続する場合には, 最大 2 音素までを, 1 つの連続母音として扱う.「病院を (byo-u/i-N/o)」の場合は 3 つの音節 ("byo-u", "i-N", "o") として作成する.

### 6.2 録音時間による選択

録音した時間が近い音声は, 音量や発話速度が同程度となると考えられる. したがって, 録音した時間が近い音節素片で合成音声を作成した場合, 音声合成に使用する音節素片の音量や発話速度のばらつきが抑えられると考えられる. 本研究で使用する音声データベースは, 音声が録音した順番に文番号が割り当てられている. そこで, 音節素片の含まれる文の文番号の値の差が小さい音節素片の組み合せを選択し, 音声を作成する.

### 6.3 合成音声の例

「むかって (/mu/ka/q/te/)」および「銀行の (/gi/N/ko/u/no/)」について, 従来の音節波形接続方式を文節に適用した手法 (3. 章) と, 改良した手法 (6. 章) で作成した場合の音節素片の選択の具体例を下に示す. なお,「‾|_」は音の強弱 (アクセント) を表している. 括弧内の太字の部分は, 実際に選択される部分を示している. また, 括弧の右にある数字は文番号を表している.

(1) 従来法 (3. 章) の合成音声

むかって (/mu/ka/q/te/)

= 昔の (/**mu**/ka/shi/no/) .831
+ 使って (/tsu/**ka**/q/te/) .223
+ 当たった (/a/ta/**q**/ta/) .178
+ 終わって (/o/wa/q/**te**/) .416

銀行の (/gi/N/ko-u/no/)

= 銀行に (/**gi**/N/ko-u/ni/) .595
+ 深刻に (/shi/N/**ko**/ku/ni/) .054
+ 天候に (/te/N/**ko-u**/ni/) .027
+ 民謡の (/mi/N/yo-u/**no**/) .001

(2) 改良法 (6. 章) の合成音声

むかって (/mu/ka/q/te/)

= 昔の (/**mu**/ka/shi/no/) .831
+ 使った (/tsu/**ka**/q/ta/) .890
+ 止まった (/to/ma/**q**/ta/) .854
+ 歌って (/u/ta/q/**te**/) .791

銀行の (/gi/N/ko-u/no/)

= 銀行に (/**gi**/N/ko-u/ni/) .595
+ 健康に (/ke/N/**ko-u**/ni/) .545
+ 劇場の (/ge/ki/jo-u/**no**/) .558

図 5,6 に「向かって」の従来法と改良法のスペクトラムと音声波形を示す.

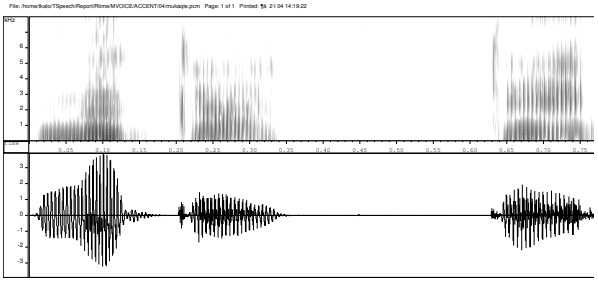

図 5 「むかって」(従来法 (3. 章))

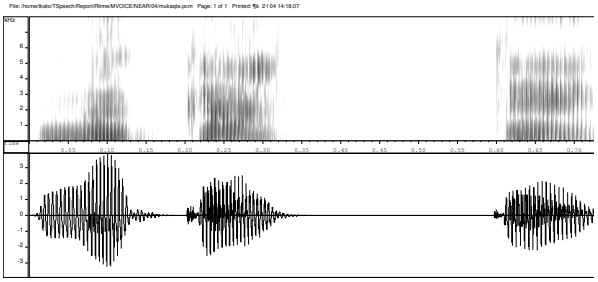

図 6 「むかって」(改良法 (6. 章)

## 7. 改良法 (6. 章) の評価実験

作成した音声の評価のため, 聴覚実験を行う. 聴覚実験は前述した 3.3 節と同じ条件とする. また, 対比較実験は, 自然音声と, 3. 章で作成した音節波形接続型音声合成法 (以後従来法) と, 改良法 (6. 章) の合成音声の 3 種類の組み合せで行う.

### 7.1 了解度試験の実験結果

了解度試験の実験結果を表 7 に示す.

表 7　了解度試験の実験結果

| | 了解度　正解率 (%) |
|---|---|
| 自然音声 | 99.3(894/900) |
| 改良法 (6. 章) | 99.3(894/900) |
| 従来法 (3. 章) | 98.7(889/900) |

改良法 (6. 章) の了解度は 99.3%であり従来法（3. 章）の 98.7%に比べると改良法の方が高い値が得られた．自然音声と比較してみても，改良法は自然音声と同程度の高い値が得られた．

### 7.2　オピニオン評価の実験結果

オピニオン評価の全被験者の平均を表 8 に示す．

表 8　オピニオン評価の実験結果

| | オピニオンスコア |
|---|---|
| 自然音声 | 4.75 |
| 改良法 (6. 章) | 3.83 |
| 従来法 (3. 章) | 3.55 |

この表から，改良法のオピニオンスコアは 3.83 となり，自然音声には及ばないものの，高い品質の音声が作成されていることが確認できる．また，従来法（3. 章）の 3.55 より高い値が得られ，音節素片の選択の条件を追加することで品質が向上していることが分かる．

### 7.3　対比較実験の結果

#### 7.3.1　自然音声との対比較

従来法と自然音声および改良法と自然音声の対比較実験の結果を表 9 に示す．

表 9　自然音声との対比較の結果

| | 自然音声 (%) | 改良法 (6. 章)(%) |
|---|---|---|
| 文節数 100 | 74.3 | 25.7 |
| | 自然音声 (%) | 従来法 (3. 章) (%) |
| 文節数 100 | 82.7 | 17.3 |

この結果から改良法は自然性の面では自然音声との間にまだ差があるが，品質の高い音声が作成されていることが分かる．また，改良法は 25.7%の文節が自然音声よりも良い音声だと判定されているのに対し，従来法（3. 章）は 17.3%となっており，音節波形の選択条件を追加した効果が現われている．

#### 7.3.2　従来法（3. 章）と改良法 (6. 章) との対比較

従来法（3. 章）と改良法 (6. 章) の合成音声の対比較実験の結果を表 10 に示す．

表 10　従来手法の合成音声との対比較

| | 従来法 (3. 章)(%) | 改良法 (6. 章)(%) |
|---|---|---|
| 文節数 95 | 39.3 | 60.7 |

従来法（3. 章）との対比較では，改良法 (6. 章) の方が良いと判定された文節が 60.7%となった．このことから，音節の選択条件を追加することで自然性が向上したことが分かる．

## 8.　考　　察

### 8.1　了解度試験の解析

了解度においては，改良法の合成音声は自然音声と同程度の高い値となった．了解度試験において，合成音声で多くの被験者が間違えた音声を表 11 に示す．

多くの誤りは，「**か**たった」を「**た**たった」や「あった**が**」が「あった**ら**」等の発音方法が似ている子音であった．また，最終モーラの聞き逃しが，従来法（3. 章）では 3 つあるが，録音時間を考慮した改良法では 1 つ減少した．

表 11　聞き間違えた音素

| 原因 | | 改良法 (6. 章) | 従来法 (3. 章) | 合計 |
|---|---|---|---|---|
| 最終モーラの欠落 | | 1 | 3 | 4 |
| 母音部の誤り | [i] → [-] | 1 | | 1 |
| | [a] → [u] | | 1 | 1 |
| 撥音部の誤り | [N] → [bu] | | 1 | 1 |
| 促音部の誤り | [q] → [-] | 1 | | 1 |
| 子音部の誤り | [g] → [r] | | 1 | 1 |
| | [k] → [h] | | 2 | 2 |
| | [k] → [t] | 1 | 1 | 2 |
| | [ky] → [ry] | | 1 | 1 |
| | [n] → [r] | 1 | | 1 |
| | [sh] → [j] | 1 | | 1 |
| | [t] → [k] | | 1 | 1 |

### 8.2　オピニオン評価の解析

了解度では自然音声も合成音声も同程度の値が得られた．しかしオピニオンスコアでは自然音声は 4.75 の対して従来法（3. 章）では 3.55，そして改良法（6. 章）でも 3.83 となり，自然音声には及ばなかった．オピニオンスコアの評価が悪かった合成音声の多くが接続部での違和感が原因となっていた．これは，音素ごとの微妙な音量の違いから，接続部が強調されてしまい不自然に聞こえたと考えている．オピニオン評価の音声の種類による比較を表 12 に示す．

表 12　オピニオンスコアの平均の種類による比較

| | 自然音声 | 同値 | 従来法 (3. 章) |
|---|---|---|---|
| 文節数 100 | 96 | 2 | 2 |
| | 自然音声 | 同値 | 改良法 (6. 章) |
| 文節数 100 | 87 | 6 | 7 |
| | 従来法 (3. 章) | 同値 | 改良法 (6. 章) |
| 文節数 95 | 20 | 5 | 70 |

この結果を対比較実験と比較すると，自然音声の自然性が非常に高くなっている．本研究ではオピニオン評価は比較対象となる文節を自然音声の文の中に埋め込んで行った．そのため，自然音声の文と比較対象の文節との音量差から，合成音声が不自然に聞こえてしまうことがあった．そのため，合成音声は自然性が低く評価されてしまい，自然音声のオピニオンスコアの値が際立った結果になったと考えられる．

### 8.3 通常の発話速度における合成音声

この節では，通常の発話速度における音節波形接続型音声合成の文節への適用を試みる．用いるデータベースは ATR の Aset の中の DSB である．このデータベースは文を文節ごとに区切って発話されているが，区切る時間が短く，普通の発話に近い．また，収録されているデータ量が 115 文と少ないため，作成できた文節音声は 12 文節であった．合成方法は，従来法（3. 章）である．作成した合成音声の評価のために，8 人の被験者について，了解度試験とオピニオン評価を行った．他の条件は，3. 章と同じである．

評価の結果を表 13 に示す．この表から，得られた合成音声は了解度，オピニオンスコアとも 3. 章で得られた音声に品質では及ばないことがわかる．

この原因として，通常の発話速度で発話した音声では，文節間での区切りの時間が短いため，ピッチが初期化しないために，$F_0$ が複雑になったためと考えている．したがって，音節波形接続型音声合成方法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．

表 13　通常の発話速度における実験

| | 了解度　正解率 (%) | | | オピニオンスコア | | |
|---|---|---|---|---|---|---|
| | FTK | FYN | 平均 | FTK | FYN | 平均 |
| 自然音声 | 96 | 99 | 98 | 4.5 | 4.7 | 4.6 |
| 合成音声 | 96 | 98 | 97 | 3.2 | 3.0 | 3.1 |

### 8.4 アクセント型とアクセント核

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，作成したほとんどの文節において同じ波形が選択された．この結果から，波形接続方式を文節に適用した場合は，アクセント型のみを考慮すれば十分であると思われる．

### 8.5 連続母音の効果

従来法（3. 章）で作成した合成音声は音節素片の接続部の違和感が問題となる．特に違和感を感じるのは母音や撥音が連続する部分である．そこで改良法（6. 章）では最大で 2 音素をまとめて連続母音として扱い，音声を作成した．その結果，波形の接続部での違和感が軽減でき，品質が向上した．

しかし，改良法（6. 章）でも最大で 2 音素までという制限を加えたため，例えば『遺体を (i/ta/i/o)』という文節の場合には『遺体を (i/ta-i/o)』となってしまい，最後の『い』から『を』へのつながりが不自然になっているということがあった．また，文節の場合には，助詞との接続の部分も音が切れずスムーズにつながることが多く，その部分での音声の質の違いが自然性の低下に結びついている場合もあった．

改良法（6. 章）では，最大で 2 音素までという制限を加えて合成音声を作成した．しかし同じ音声から使用する音声の制限をなくすことで，さらに品質が高くなると考えている．

### 8.6 録音時間が近い音節素片を選択した効果

改良法では，母音や撥音が連続した場合に連続母音として音声を作成した．この連続母音を作成せずに，録音時間による制御のみを使用して作成した音声と従来法の対比較の結果を表 14 に示す．表 14 より，録音した時間による制御でも，合成音声の品質向上に大きく影響していることがわかる．ただし，不適切な音節素片が選択され，品質が低下する場合もあった．

表 14　録音時間による制御の効果

| | 録音時間による選択 | 従来法（3. 章） |
|---|---|---|
| 文節数 50 | 59.2(266/450) | 40.8(184/450) |

### 8.7 音量のばらつき

改良法（6. 章）では音量のばらつきを抑えるために，録音した時間が近い音声を選んで音声を作成するようにした．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，完全に音量の音量の統一はできず，不自然さが残る音声も作成された．

さらに品質を上げるには，接続部分の音量の同程度の音節素片を使用するなどという手法が考えられる．しかし，自然性に関わる要素は音量だけではないため，信号処理を用いる必要があると考えている．

### 8.8 合成できる文節音声の数

改良法では，母音や撥音が連続した場合に連続母音として音声を作成した．この結果，作成した合成音声の品質は向上した．しかし，音節素片の種類の数が増加するため，作成可能な文節数が減少する．母音や撥音が連続した場合に，連続母音として扱わずに作成する従来法（3. 章）で作成できる 4，5，6 モーラの文節数は，382 文節だった．これに対し，連続母音として作成する改良法（6. 章）の場合は，323 文節へと減少した．品質向上のために条件を増加すると，作成可能な文節数はさらに減少する．

この問題については，特に後音素環境において似た子音をグループ化し，音素環境を代替して音節素片の種類数を少なくすることで，解決できると考えている．今後音声品質を減少することなく音節素片の数を削減する方法を考慮する必要があろう．なお，環境などは異なるが，文献 [17] において，短い刺激音声では自然性が劣化していることが報告されている．

## 9. ま と め

本研究では文節発声で発話速度が遅い音声を合成する場合の音節波形接続方式の有効性を調査した．

音節波形接続方式は，音節・直前の音素・直後の音素音素・文節中のモーラ位置・文節のモーラ数・文節のアクセント型を一致している音節素片を接続して合成音声を作成する．この手法で作成した合成音声は，了解度が 98.7% でオピニオンスコアも 3.55 が得られ，音節波形接続方式が文節に対しても有効であることが分かった．

また，音節波形接続方式の改良方法として，音節選択に 2 つの条件を追加した．この改良方法では，了解度は 99.3%，オピニオンスコアが 3.83 となり音声の品質が向上したことがわかった．対比較実験でも 60.7% の文節が従来法（3. 章）よりも品質の高い音声と判定され，追加した 2 つの条件が音節素片の選択に有効であることが分かった．一方，自然音声は了解度が

99.3%，オピニオンスコアは 4.75 となった．自然性の面では自然音声と合成音声の間にまだ差がある．

今後は，接続部の違和感を軽減するため様々な制御を導入していくつもりである．また，多くの文節を合成可能にするため，接続する音素の条件を緩和する方法，例えば，後音素環境における子音のグループ化を検討していくことが必要である．

## 謝辞

録音音声の収録に対し，株式会社アイアール，アルトの山本修さん，柴田葉子さんに御世話になりました．またナレータの轟美穂さんに，無理な収録をお願いしました．最後に，鳥取大学工学部知能情報工学科計算機 C 講座の方々に聴覚実験をしてもらいました．これらの方々に感謝いたします．
なお，合成した音声 100 文節は，http://unicorn.ike.tottori-u.ac.jp/2002/tkato/demo/demo.html においてあります．

## 文　献

[1] 広川智久, "波形辞書を用いた規則合成法" , 電子情報通信学会技術研究報告, SP88-9, pp.65-72,1988.

[2] 村上仁一, 水澤紀子, 東田正信,"音節波形接続方式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).

[3] 石田隆浩, 村上仁一, 池原悟, "音節波形接続型音声合成の普通名詞への応用", 電子情報通信学会技術研究報告, SP2002-25, pp. 7–12 (2002).

[4] 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への応用", 日本音響学会 2003 年春期研究発表会, 2-Q-18, pp. 1-409,410 (2003).

[5] 益子貴史, 徳田恵一, 小林隆夫, 今井,"動的特徴を用いた HMM に基づく音声合成", 電子情報通信学会論文誌 D-II, Vol.J79-D-II, No.12, pp. 2184–2190 (1996).

[6] 徳田恵一, "HMM による音声合成の基礎", 電子情報通信学会技術研究報告, SP2000-74, pp. 43–50 (2000).

[7] Jan P.H. van Santen, Richard W. Sproat, Joseph P. Olive and Julia Hirschberg, "Progress in Speech Synthesis", Springer, ISBN 0-387-94701-9 (1996).

[8] 戸田智基, 河井恒, 津崎実, 鹿野清宏, "素片接続型日本語テキスト音声合成における音素単位とダイフォン単位に基づく素片選択", 電子情報通信学会論文誌 D-II, Vol.J85-D-II, No.12, pp.1760–1770 (2002).

[9] Nich Campbell and Alan W.Black, "CHATR:自然音声波形接続型任意音声合成システム", 電子情報通信学会技術研究報告, SP96-7, pp. 45–52 (1996).

[10] 石川泰, "音声合成のための韻律制御の基礎", 電子情報通信学会技術研究報告, SP2000-72, pp. 27–34 (2000).

[11] 村上仁一, 前田智広, 池原悟, "モーラ情報を用いた音素ラベリング方式の検討", 電子情報通信学会技術研究報告，SP2003-137，pp. 145–150 (2003).

[12] 妹尾貴宏, 村上仁一, 池原悟, "モーラ情報を用いた単語音声認識の検討", 電子情報通信学会技術研究報告, SP2002-130, pp. 55–61 (2002).

[13] 谷口勝則, 村上仁一, 池原悟, "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", 電子情報通信学会技術研究報告, SP2002-131, pp. 63–68 (2002).

[14] NHK 出版, "NHK 日本語発音アクセント辞典 新版", ISBN 4-14-011112-7 (1998).

[15] 村上仁一，池原悟，徳久雅人, "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).

[16] Kåre Sjölander and Jonas Beskow: Wavesurfer, http://www.speech.kth.se/wavesurfer/

[17] 河井恒, 津崎実, 舛田剛, 岩澤秀紀,"波形素片接続時の音素環境代替による自然性劣化の知覚的評価", 電子情報通信学会技術研究報告, SP2001-22, pp. 51–57 (2001).

[18] 加藤琢也, 村上仁一, 池原悟,"波形接続型単語音声合成の文節への適用", 日本音響学会 2004 年秋期研究発表会, 3-2-12, pp.1-339,340 (2004-09).

[19] 加藤琢也, 村上仁一, 池原悟, "波形接続型単語音声合成における融合ラベルの検討", 日本音響学会 2005 年春期研究発表会, 3-P-22, pp.1-283,284 (2005-03).