

# 話速変換を用いた音節接続型音声合成方法の検討\*

○村上仁一, 池原悟 (鳥取大学工学部)

## 1 はじめに

波形編集型音声合成 [1],[2] の方法の1つとして音節波形接続型音声合成が提案されている [3]. 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をしている. この音声合成方法は, 始めは, 固有名詞と普通名詞を対象に研究を行った [3][4]. 次に, 音節波形接続方式を, 文節に適用した [7]. この結果, 文節発声でゆっくりと発話された音声を対象にしたとき, 品質の高い合成音声を得られた. しかし, 発話速度が早い場合, 合成される品質に劣化した [8].

一方, 音声の話速を変える話速変換の技術が過去に多く研究されている [9],[10]. そして, 話速の変化が小さいとき, 音質を殆ど劣化しないことが知られている. そこで, 音節波形接続型音声合成で作成した話速が遅い文節音声に, 話速変換をおこない, 話速を速くして, 得られた音声の品質を評価する.

## 2 音節波形接続型音声合成

### 2.1 音節波形接続型音声合成の概説

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.

Table 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 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. 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 [5]. 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. つまり, 文節発声で発話速度が遅い音声を用いて, 音節波形接続型音声合成で音声を作成した場合, 自然性が高い音声を作成できる. しかし, 発話速度が速いばあい, 合成される品質が劣化する [8].

## 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 [6]) の文を使用する. この例文集は機械翻訳を目的にしている. 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.

Table 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	あの二人は-親の-反対を-押し切って-結婚した
4	彼女が-学校を-休んだので-がっかりした

また, アクセントは収録した音声データから人間が聴取実験により付与した.

### 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.

Table 3 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 話速変換

話速変換は, 録音素材の音声を高速に聞くために研究され, 多くの研究がある. 基本的な考え方は, 母音の長さを短くして, 発話速度を上げる. 具体的には, 隣あったフレームの相関係数が一定値より大きい場合, そのフレームを削除する. 話速変換には, 多くのソフトが公開されている. 本研究では, nave player [11] を使用した. また話速は1.2倍速にして実験をおこなう.

### 3.4 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聴かせ, オピニオン評価と対比較実験の実験を行う.

\*Using speech Control for Concatenating Syllabic Components based on Positional Features and Mora Information and Accent . by Jin'ichi Murakami , Satoru Ikehara (Tottori University)

## 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。オピニオン評価では、自然に聞こえた度を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。

## 2. 対比較実験

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、どちらの音声に自然に聞こえたか判定する。

なお、評価は、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。比較対象の文節は、評価者に予め提示しておく。使用する文の例を下に示す。例文中のアンダーラインを引いた文節が音声合成の箇所である。

(例) 全部員が 優勝 を目指して練習に励んでいる

## 4 実験結果

オピニオン評価および対比較実験の平均値を表4に示す。

Table 4 実験結果

	オピニオンスコア	対比較実験
自然音声	4.19	79.4%
合成音声	3.52	20.6%

合成音声のオピニオンスコアは3.52となり、高い品質の合成音声で作成できたことが分かる。しかし、自然性の面で自然音声との差がある。また、対比較試験から、合成音声の方が良い音声だと判定された文節が20%あった。

自然音声と合成音声に差のある文節を調査すると、前後の音量が変換するために違和感を感じる場合が多かった。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。

これらの結果から品質の高い合成音声で作成されているが、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 考察

### 5.1 話速

本研究では、公開されている話速変換を使用して、1.2倍速にして音声を作成した。この速度を1.4倍にしても、あまり音質の劣化がなかった。よってより速い音声合成が可能と考えている。

### 5.2 話速変換の方法

話速変換には、多くの方法がある。この中から最適な方法を選択していく必要がある。なお、過去の実験と比較すると、話速変換を行うことで、自然音声と合成音声の差が広がる傾向にある。

### 5.3 実験方法

今回の実験は、収録データベースの関係上、自然音声も含めて全ての音声に話速変換を適応した。し

かし、ナレータに文の固定部を速い音声で録音し、可変部のみを合成して評価するべきである。今後、この評価方法を試みたい。

## 6 結論

本研究では、音節波形接続方式を文節に適用したときの問題点の1つである、話速の遅さを解消するために、話速変換を利用した。その結果、高い自然性が得られた。したがって音節波形接続方式と話速変換を利用することで、話速が早い文節にも品質の高い音声合成ができることが示された。

今後の課題としては、音量が変換するために違和感を感じる場合が多いため、これを一定にする方法を考えていく必要がある。また、音節波形接続方式はパラメータが多いため、合成できる文節が少ない。そこでパラメータを選択することで、合成できる文節数を増やすことがあげられる。

**謝辞** NTTサイバースペース研究所の水野秀之氏に、基本的なアイデアをいただきました。また、鳥取大学工学部知能情報学科4年の山形亮氏に聴覚実験をお願いしました。厚くお礼を申し上げます。

## 参考文献

- [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] 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).
- [3] 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).
- [4] 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).
- [5] 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).
- [6] 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).
- [7] 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).
- [8] 村上, 他, "音節波形接続型音声合成の文節への適用", 電子情報通信学会技術研究報告 SP2005-19, pp.43-50, (2005)
- [9] 今井篤, 清山信正, 都木徹, 宮坂栄一, 小野博, "高齢者を対象とした話速変換音声の評価実験", 日本音響学会研究発表会講演論文集 Vol. 1999, 春季 pp. 367-368 (1999)
- [10] 田中章浩, 坂本修一, 鈴木陽一, "話速変換音声聴取に音声伸長量とポーズ長が及ぼす影響", 電子情報通信学会技術研究報告 SP2004-161, pp.43-48, (2005)
- [11] [http://www.seera.jp/NaveTheBK/download\\_j.html](http://www.seera.jp/NaveTheBK/download_j.html)