

音節波形接続方式における融合ラベルの検討*

加藤 琢也, 村上 仁一, 池原 悟 (鳥取大・工)

1 はじめに

音声合成の1つの手法として音節波形接続方式が提案されている。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。音声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 [1]。

この音節波形接続方式は名詞や文節に適用され、名詞や文節の音声合成において音節波形接続方式が有効な方法であることが示されている [1][2][3][4]。しかし、文節を対象とした場合、接続部の違和感や音声素片の音量の違いによる違和感などの問題点が指摘されている。

そこで、本研究では従来の文節を対象とした音節波形接続方式に、音節選択の条件を加え合成音声を作成する。そして、合成音声の品質がどの程度自然音声に近づけるか、また、どのような問題点が現われるかを調査する。

2 従来の音節波形接続方式

2.1 従来手法の特徴

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)と、アクセント型を使用する。特定話者の単語発話においては、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 [1]。また、普通名詞の音声合成においては、モーラ情報とアクセント型が有効であることが示されている [3]。

また、文節発声で発話速度が遅い音声の場合には、文節毎に区切ることでピッチが初期化される。そのことにより、文節においても複雑な韻律情報を用いなくても品質の高い合成音声ができる。

2.2 音節波形接続方式による音声合成

文節における従来の音節波形接続方式の音声合成では、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、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素片を選択し、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。

2.3 従来手法による合成音声の例

従来手法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。

遺体を (i/ta/i/o)
= いただき (i/ta/da/ki)
+ 期待が (i/ta/i/ga)
+ ひたいを (hi/ta/i/o)
+ 気持ちを (ki/mo/chi/o)

3 問題点と改善案

3.1 問題点

従来手法で作成した音声は音声素片の接続部に違和感があり、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。これらの音素は前後の音が連続的に変化する部分であり、音素を切り分けることが困難である。その母音や撥音が連続する部分を別々の素片として扱うことが、自然性の低下に結びついていると報告されている。

また、不自然さを発生させるもう1つの大きな要因は、音声素片の音量のばらつきである。音節波形接続方式には大量の録音された音声が必要となるため、音声によって音量のばらつきがある。そして、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。

3.2 改善案

本研究では問題点解決のために2つの改善策を提案する。

(1) 接続部の違和感軽減のための改善案

従来手法では、音素境界が明確でなく、音素境界で切り離すことが困難とされる部分のみ連続母音として扱っている。改善策として、本研究では母音と撥音が連続する部分を連続母音として扱う。ただし、母音や撥音が多数続く場合があるため、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。

提案手法で作成した合成音声の例を下に示す。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。

遺体を (i/ta-i/o)
= いただき (i/ta/da/ki)
+ ひたいを (hi/ta-i/o)
+ 手紙を (te/ga/mi/o)

(2) 音量のばらつき軽減の改善案

また、従来手法では波形の選択条件に合致した音節候補の中でデータベースの上位の素片を選択している。しかし、録音した時間帯に近い音声は、音量や発話速度などが近い音声だと考えられる。そこで、本研究では音節候補の中で音声を録音した時間帯に近い素片の組み合わせを選択する。

4 評価実験

4.1 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 [5]) の文を使用する。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、従来手法による合成音声、提案手法による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。

4.2 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9名を対象に聴覚実験を行う。聴覚実験では了解度試験、オビニオン評価、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

*Examination of Contractio in Phrase Synthesis by Concatenating Syllabic Components. By Takuya Kato, Jin'ichi Murakami and Storu Ikehara (Tottori Univ)

解度試験では自然音声の文中に比較対象となる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。

そして、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提案手法の合成音声、自然音声と従来手法の合成音声、提案手法の合成音声と従来手法の合成音声の3つの組み合わせで行う。そして、同じ文節で2種類の音声を続けて出し、どちらの音声が自然に聞こえるかを判定する。

4.3 音節波形接続方式に関する補足

音節波形接続方式では接続部での違和感が自然性に大きく影響する。そこで本研究では、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。

5 実験結果

5.1 了解度、オピニオン評価の実験結果

了解度試験および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	99.3(894/900)	4.75
提案手法	99.3(894/900)	3.83
従来手法	98.7(889/900)	3.55

表1から、提案手法も従来手法も品質の高い音声を作成できた。提案手法と従来手法を比べると、提案手法の方が了解度、オピニオンスコアともに高い値が得られた。

自然音声と比べてみると、了解度では提案手法も自然音声と同程度の高い値が得られた。しかし、オピニオンスコアにおいて提案手法は自然音声と比べると自然性が低いことが分かる。

5.2 対比較実験の結果

5.2.1 自然音声と合成音声の対比較

自然音声と合成音声の対比較実験の結果を表2に示す。

表2: 自然音声と従来手法の合成音声

	自然音声 (%)	提案手法 (%)
文節数 100	74.3	25.7
	自然音声 (%)	従来手法 (%)
文節数 100	82.7	17.3

表2から合成音声は自然性の面では自然音声との間にまだ差がある。しかし、提案手法は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、従来手法は17.3%となっており、波形の選択条件を追加した効果が現われている。

5.2.2 従来手法と提案手法との対比較

表3に従来手法の合成音声と提案手法の合成音声の対比較実験の結果を示す。

表3: 従来手法の合成音声との対比較

	従来手法 (%)	提案手法 (%)
文節数 95	39.3	60.7

表3より音節の選択条件を追加することにより自然性が向上することが分かった。

6 考察

6.1 不自然な音声の解析

母音や撥音が連続する部分は、音声波形が連続的に変化する場合が多い。この母音や撥音が連続する部分を切り離すことで自然性が低下すると考えた。そこで、本研究では母音や撥音が連続する場合には、最大で2音素までを連続母音として扱い音声を作成した。そして、連続母音として扱うことで、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、今回は最大で2音素までという制限を加えたため『遺体を (i/ta/i/o)』という文節の場合には『遺体を (i/ta-i/o)』となってしまう、最後の『い』から『を』へつながる部分に不自然さが現われた。

また、文節の場合には、助詞との接続部分も音が切れずスムーズにつながるものが多く、その部分での音声の質の微妙な違いが自然性の低下に結びついている場合もあった。

この問題については、連続母音として扱う音素数の制限を撤廃することにより改善が可能であると考えられる。

6.2 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、録音した時間帯に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とも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も作成あった。

音量についてさらに品質を上げるには、接続部分の音量の同程度の素片を使用するなどという手法が考えられる。

6.3 素片単位に関する考察

今回は母音や撥音が連続した場合に連続母音として扱って音声を作成し、作成した合成音声の品質は向上した。しかし、連続母音の数が多くなるため、作成可能な文節数が減少する。

本研究では日英対訳の例文集に含まれる1000文を使用し、従来手法に母音や撥音が連続した場合は連続母音として扱うという条件を加えた。そして、従来作成で作成できる文節数が382文節だったのに対し、提案手法では323文節へと減少した。品質向上のために制御を増やすと、作成可能な文節数はさらに減少してしまう。

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素片の種類数を少なくすることで解決していくことが可能であると考えている。

7 まとめ

本研究では、音節波形接続方式を文節に適用した従来手法で問題となっている2つの問題点を取り上げ、新しい手法を提案した。問題点は、母音や撥音が連続する部分での接続部の違和感、そして、素片の音量の違いによって現われる違和感である。そして、母音や撥音が連続する場合には連続母音として扱い、録音した時間帯に近い素片の組み合わせを選んで合成音声を作成し、合成音声の品質を調査した。

その結果、聴覚実験において提案手法の合成音声は了解度が99.3%、オピニオンスコアは3.83が得られた。そして、提案手法と従来手法を比較すると了解度、オピニオンスコアともに高い値となった。また、対比較実験においても60.7%が提案手法の方が自然であると判定された。このことから、音声波形の選択条件を追加することで音声の品質が向上し、追加した音節の選択条件が音節波形接続方式に有効であることが示された。

今後は、接続部の違和感を軽減など様々な制御を導入し、音声の品質を高めることが必要である。また、それと同時に、音声コーパスをより有効に利用できるように、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重要である。

参考文献

- [1] 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).
- [2] 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).
- [3] 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).
- [4] 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, 3-2-12, pp.1-339,340 (2004-09).
- [5] 村上, 池原, 徳久, "日本語英語の文対訳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)