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普通名詞への応用

石田　隆浩† 　村上　仁一† 　池原　悟†

† 鳥取大学工学部知能情報工学科
〒 680-8552 鳥取市 湖山町南 4-101
0857-31-6788
E-mail: †{tisida,murakami,ikehara}@ike.tottori-u.ac.jp

あらまし　録音編集型音声合成においては、可変部のために大量の単語を同一話者から録音する必要がある。そこで、この問題に対する 1 つの解決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、可変部に必要となる単語のうち一部の単語を録音し、残りの単語については録音した音声の音節波形を部分的に切り出し、信号処理を行うことなく接続することにより単語音声を合成する手法である。しかし、過去の研究では固有名詞に対する有効性しか示されていない。また、音節波形を切り出す際に使われる音素ラベリングデータは通常人手によって作成されているため、ラベリング作業に大きな負担がかかる。そこで、自動的に音素ラベリングを行う方法が提案されている。しかしその精度は人手によるものに比べると低い。そこで本研究では、まず、ATR の単語発話データベース Aset を用いて普通名詞の単語音声を作成し、聴覚実験により普通名詞に対する評価を行った。その結果、普通名詞にも十分な有効性があることを確認した。また、既存の自動音素ラベリングの手法を用いて作成した音素ラベリングデータを用いた場合に、合成音声にどの程度の影響があるかを調べた。その結果、既存の自動ラベリングの精度であっても十分な品質が得られることが分かった。

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

Common Noun for Word Synthesis by Concatenating Syllabic Components

Takahiro ISHIDA†, Jin'ichi MURAKAMI†, and Satoru IKEHARA†

† Department of Information and Knowledge Engineering Faculty of Engineering Tottori University
4-101, Minami Koyamachou, Tottori city, 680-8552 Japan
+81-857-31-6788
E-mail: †{tisida,murakami,ikehara}@ike.tottori-u.ac.jp

Abstract　For the speech synthesis using slot filling method, it is need to record a lot of words from the single speaker. So, word synthesis by concatenating syllabic components is proposed over this problem. This technique is compounding a word sound by connecting syllable waveform from recorded sounds which are several parts of all of requiring sounds. However, only the validity to a proper noun is shown by the past research. And, in generally, the labelling data for cutting syllabic wave is made by hand, it requires a big burden for labelling work. Then, the automatic labelling method is proposed. However, the accuracy is low compared by hand. In this research, first, the common noun is synthesized using the word utterance database (ATR Aset), and hearing test is carried out. Consequently, it obtained that there was sufficient validity for a common noun. And, when the automatic labelling method is used, it investigated what it would have the influence of a synthetic sound. As a result, even if it was the accuracy of existing automatic labelling, it turns out that sufficient quality is obtained.

Key words　word synthesis, slot filling method, syllable, prosodic features, mora position, mora length

1. はじめに

1.1 研究の背景

近年、カーナビゲーションシステムの音声ナビゲーションのように、ユーザーの入力に対して音声で応答するシステムが様々な場面で使われている。このようなシステムで出力される音声ガイダンス文の作成には、録音編集方式が広く使われている。

録音編集方式は、システムを利用するユーザーの入力に依存する単語（以下、「可変部」）と、それ以外の部分（以下、「固定部」）から文が構成される場合に、可変部と固定部を別々に録音しておき、ユーザーの入力に対して、固定部に可変部を挿入することで音声ガイダンス文を作成する方法である。

例えば、カーナビゲーションシステムでは「目的地は○○でよろしいですか。」といったガイダンス文が出力される。この場合、○○の部分には日本の地名、駅名、建物名などの単語が挿入される。たとえば目的地として「東京駅」をユーザーが入力した場合、「目的地は ”東京駅 ”でよろしいですか。」というガイダンス文を作成する。

1.2 録音編集方式の問題点と音声合成

録音編集方式では、固定部と可変部に分けて録音しておき、実際の出力時に合成することで文章を作成する。そのため、合成時の違和感を軽減するためには、固定部と可変部を同一話者から録音する必要がある。しかし、通常可変部に必要となる音声は膨大な量となる。例えば前述のカーナビゲーションシステムの場合、地名、駅名、そして建物名を可変部として準備する必要がある。よって、現実的な時間で録音しようした場合は、複数の話者により録音する場合がある。結果、文章中に複数話者の音声が混在する結果となり、ユーザーに違和感を与える原因となる。仮に同一話者から録音できたとしても、音の高さや速さが均一でなくなる可能性があり、安定した品質の音声が得られるとは考えにくい。

固定部と可変部で同一話者による音声を用いるためには、固定部、可変部ともに音声合成によって必要な音声を作成する方法も考えられる。この方法としては、音素、音節、CV（子音－母音）、VCV（母音－子音－母音）を単位とした規則音声合成がある。規則音声合成では、音声の特徴を特徴パラメータとして持ち、それらを変形することによって音声を作成する。しかし、現状では人間の自然音声を録音した場合のように高い品質を得ることが困難である。

1.3 波形接続型音声合成とその問題点

そこで、高い品質の合成音声を得るための方法として、合成したい話者の音声をあらかじめ録音して収集し、そこから部分的に取り出した音声波形を信号処理せずに接続して任意の音声を合成する方法が提案された [1]。また、[1] と似た方法として、可変部に、録音した音声の音節波形を切り出して接続することにより合成する方法が提案されている [2]。

参考文献 [1], [2] による方法は、話者性と高い自然性を保てる特徴がある。参考文献 [2] の音節波形接続方式では、モーラ情報と前後の音素環境を元に音節波形を選択している。そして、可変部として地名を対象に実験した結果、十分実用的な品質が得られたことが報告されている。

しかし、従来提案されている音節波形接続方式 [2] では、ピッチ周波数のばらつきが比較的小さい固有名詞を対象とすることでモーラ情報を効果的に利用していると考えられ、より一般的な普通名詞に対しての有効性は示されていない。音節の選択候補が複数ある場合の選択方法も問題となる。また、音節波形を切り出す際には、音素境界位置をラベリングしたデータが必要となるが、ラベリングデータは通常人手によって作成されるため、コストがかかる。そこで、ラベリングの負担の軽減のために、自動的にラベリングを行うシステムが提案されている。しかし、その精度は人手によるものに比べるとまだ不十分である。その結果、自動音素ラベリングによるラベリングデータを用いて音節波形を切り出した場合に、品質にどの程度の影響があるか不明である。

1.4 本研究の目的

以上を踏まえ、本研究では、複数候補からの音節波形の絞り込み手法の検討、普通名詞への有効性の調査、および自動音素ラベリングデータの有効性を調査する。具体的には、まず、複数候補の絞り込みのための手法を提案し、ランダムに音節を選択した場合との合成音声の品質を比較する。次に、音節波形接続方式を用いて普通名詞の合成音声を作成し、聴覚実験によって品質を調査する。それにより、普通名詞に対する有効性を調べる。また、自動音素ラベリングによって作成したラベリングデータを使用して合成音声を作成し聴覚実験を行う。そして、合成音声の品質にどの程度影響するのかを調査することで、音節波形接続方式における自動音素ラベリングの影響を調べる。

2. 音節波形接続方式

2.1 概　　要

本研究では、単語音声合成の方法として音節波形接続方式 [2] を使用する。合成の対象となる音声を作成するために、録音されている音声の波形を部分的にデータベースから選択し、切り出す。選択する際には、前後の音素環境と単語のモーラ数、モーラ位置を用いる。そして、切り出した音節波形を接続することによって合成音声を作成する。元の波形に対して信号処理を行わないため、自然性が高い合成音声を作成することが可能である。

2.2 モーラ情報とピッチ情報

音節波形接続方式では、韻律的な情報として、モーラ情報を使用する。本論文で言うモーラ情報とは、単語のモーラ数とモーラ位置のことである。

特定話者の単語発話において、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]。図 1 に示すのは、参考文献 [2] から引用した図であり、5 モーラ語の地名 2,800 件のピッチ周波数の平均

と分散を示している。図から、ピッチ周波数の分散は小さく、モーラ数とモーラ位置が決まればピッチ周波数がほぼ決定できることが分かる。また、4 モーラ、6 モーラの単語についても同様の傾向にあったと報告されている。よって、モーラ情報と韻律情報には依存関係があると考えられるので、韻律情報としてモーラ情報が使用できると考えられる。

過去の研究では、モーラ情報は音素ラベリングや音声認識などの分野において効果があることが報告されている [6] [7]。

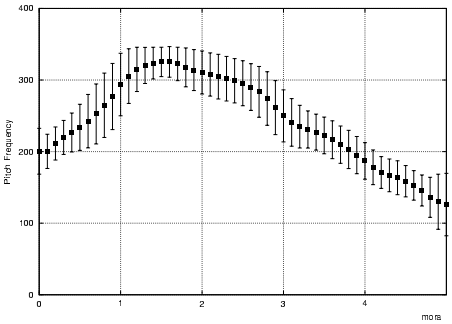

図 1　モーラ情報とピッチ周波数の関係

2.3　音節波形接続方式による音声合成

音節波形接続方式における音声合成では、まず以下の情報を含む音節部品を選択する。

Sy：音節
P：直前の音素 (前音素環境)
N：直後の音素 (後音素環境)
m：単語中のモーラ位置
M：単語のモーラ数

例えば、音節部品を「$Sy(P,N)_{m,M}$」の形式で表現すると、「反対」(ha/ng/ta/i) という音声に対する音節部品は次のように表現できる。

まず、単語のモーラ数は 4 なので、$M = 4$ である。

第 1 モーラの音節は「ha」であり、直前の音素がなく、直後の音素が「ng」であるので「ha(,ng)$_{1,4}$」となる。

第 2 モーラの音節は「ng」であり、直前の音素が「a」、直後の音素が「t」であるので「ng(a,t)$_{2,4}$」となる。

第 3 モーラの音節は「ta」であり、直前の音素が「ng」、直後の音素が「i」であるので「ta(ng,i)$_{3,4}$」となる。

第 4 モーラの音節は「i」であり、直前の音素が「i」、直後の音素がないので「i(a,)$_{4,4}$」となる。

音節部品を取り出したら、各音節の開始時間と終了時間を元に波形データを切り出し、それらを単純に接続して合成する。

音声合成の概略について図 2 に示す。

3.　自動音素ラベリング

本研究では、自動音素ラベリングの方法として、基本的な HMM によるセグメンテーションの方法を用いる。この方法では、Baum-Welch アルゴリズムによって音素 HMM を学習し、

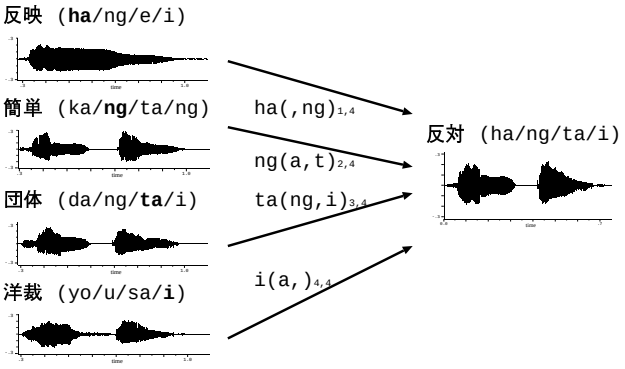

図 2　音声合成の概略図

Viterbi アルゴリズムによって音素境界位置を計算する。

自動セグメンテーションに関する研究としてはこの他に、HMM とベイズ確率を用いた方法 [3]、ルールベースによる方法 [4]、知識処理を用いる方法 [5] などがある。また、HMM を用いたセグメンテーションに関しては、さらに高い精度を得るための研究も行われている [6]。なお、これらの手法と本研究で用いた標準的な手法では、精度に大きな差がないと考えている。

4.　評 価 実 験

4.1　音声データベース

音声データベースとして、ATR の単語発話データベース Aset(5,240 件) を使用する。話者には、ピッチ周波数が比較的低く、収録された音声にエコーが少ないため、FTK と FYN の 2 話者を選ぶ。

4.2　自動音素ラベリング

自動音素ラベリングを行うためのツールとして、HTK [9] を使用する。使用するパラメータについては表 1 に示す。

表 1　使用するパラメータ

標本周波数	16kHz
特徴パラメータ	16 次 MFCC
音響モデル	4 状態 3 ループ Diagonal
mixture	3

自動音素ラベリングのラベルデータの作成は、Aset のデータ 5,240 単語を奇数番と偶数番に分け、奇数番の単語で音素 HMM を学習して偶数番の単語をラベリングする。また、偶数番の単語で音素 HMM を学習して奇数番の単語をラベリングすることで、5,240 件全てについてラベルデータを作成する。

4.3　音 声 合 成

4.3.1　評 価 対 象

4 モーラ単語 50 個について、自然音声、手動ラベルデータを用いた合成音声（手動ラベル）、自動音素ラベリングのラベルデータを用いた合成音声（自動ラベル）、そして市販の合成機の合成音声（市販の合成音）で評価する。

なお本研究では、音節波形接続方式による合成音声について

は、厳しい条件で評価するために、同一の録音データからは1音節しか切り出せないこととする。評価対象は 4 モーラ語であるので、4 種類の録音データからの音節波形を接続して作成する。Aset の 4 モーラ単語を作成した場合、約 200 個の合成音声を作成することができるが、データベースに偏りがあり撥音を含む音声が多く作られるため、音節のバランスを考えて 50 単語で実験を行うこととする。

本研究で作成した音声の一部を表 2 に示す。例えば「反対」(ha/ng/ta/i) という合成音声は、「反映」の第 1 音節「ha」、「安定」の第 2 音節「ng」、「軍隊」の第 3 音節「ta」、そして「展開」の第 4 音節「i」の音節波形を接続して作成されている。

表 2　作成した音声（一部）

音声	音節 1	音節 2	音節 3	音節 4
反対	反映 (/**ha**/ng)	安定 (a/**ng**/t)	軍隊 (ng/**ta**/i)	展開 (a/**i**/)
雑談	雑音 (/**za**/ts)	活動 (a/**tsu**/d)	切断 (u/**da**/ng)	朝刊 (a/**ng**/)
妊娠	人形 (/**ni**/ng)	進出 (i/**ng**/sh)	安心 (ng/**shi**/ng)	雑巾 (i/**ng**/)
発見	醗酵 (/**ha**/q)	発行 (a/**q**/k)	石鹸 (q/**ke**/ng)	実験 (e/**ng**/)
分配	分解 (/**bu**/ng)	運搬 (u/**ng**/p)	心配 (ng/**pa**/i)	正体 (a/**i**/)

4.3.2　音節波形接続方式による合成音声の作成

まず、合成対象となる単語から音節部品ラベル列を作成する。音節部品ラベルとは、 2.3 節で示した情報を列挙したラベルのことである。そして、各音節部品ラベルに一致するようなラベルデータを音節部品データベースから取り出し、音節の波形位置を元に波形データを切り出し、単純につなぎ合わせて合成する。

作成した合成音声の波形データの例として、「反対」(ha/ng/ta/i) の波形とスペクトログラムを図 3 に示す。波形とスペクトログラムの表示には、Praat [11] を使用した。

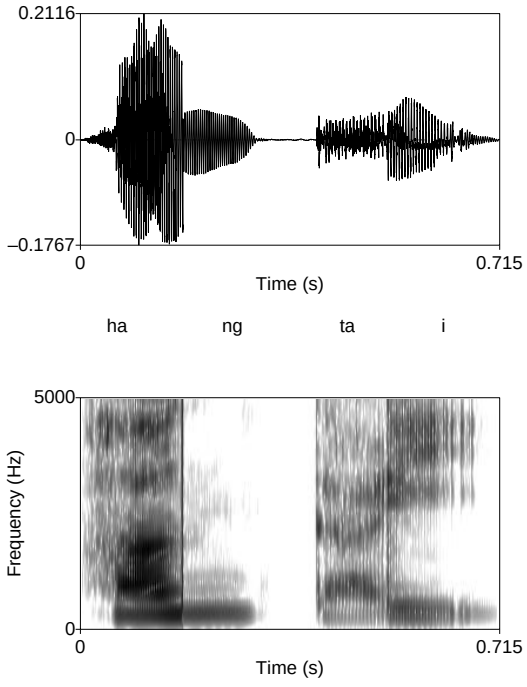

図 3　合成音声「反対」(ha/ng/ta/i) の波形とスペクトログラム

4.3.3　市販の合成機による合成音声の作成

本研究では比較のために、市販の合成音として富士通株式会社の「Linux 版日本語音声合成ライブラリー」[10] によって作成したものを使用する。このライブラリでは、PSOLA 方式を用いて合成音声を作成していると思われる。作成において、話者の声質情報として、女声で FTK,FYN に近いと思われる声の高さを付加する。

4.4　評価方法

合成音の評価のために、音声研究に関わった経験のない人 5 名を対象に、了解度試験とオピニオン評価を行う。評価の具体的な方法について以下に示す。

(1) 了解度試験

単語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、自然音声と 3 種類の合成音声をランダムにヘッドフォンから被験者に聞かせ、どのように聞こえたかを仮名で書き取らせる。自分の知識などは用いず、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

単語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、自然音声と 3 種類の合成音声をランダムにヘッドフォンから被験者に聞かせ、どの程度自然に聞こえたかを 5 段階（1 が最も不自然、5 が最も自然）で評価するように指示する。

5.　実験結果

5.1　音節の選択方法の違いによる品質の比較

まず、音節の選択方法に関して実験を行う。

音節波形接続方式では、取り出す波形データの候補が複数出る場合があり、選択する方法が問題となる。そこで本研究では、音節の選択方法について、以下のように規則を定める。概要については、図 4 に示す。

（1） 第 1 音節についてはデータベース内の先頭要素を選択する。

（2） 第 2 音節以降については、多くの候補数を削減するために、取り出す音節の開始時間が直前の音節の終了時間に一番近いものを選択する。

（3） それでも候補が複数出た場合は、音節の継続時間が長い方が聞き取りやすくなると考え、一番長いものを選択する。

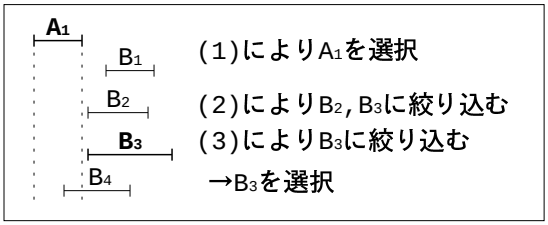

図 4　音節の選択方法の概要

以上の規則によって音節を選択して音声を作成した場合と、完全にランダムに音節を選択して音声を作成した場合の品質を比較した。実験結果を表 3 に示す。

表 3　選択方法の違いによる結果

	了解度　正解率 (%)			オピニオンスコア		
	FTK	FYN	平均	FTK	FYN	平均
規則あり	99.97	99.93	99.95	3.43	3.73	3.58
ランダム	99.92	99.98	99.95	3.14	3.74	3.44

表 3 から、了解度では違いが見られなかったが、オピニオンスコアは、特に FTK では上で述べた規則を用いた方が良い結果が出ていた。よって、規則を用いて音節を選択した方がよい品質が得られる可能性があることが分かった。

5.2　普通名詞に対する有効性の確認

普通名詞に対する音節波形接続方式の有効性を調べた。実験結果を表 4 に示す。

表 4　実験結果 (1)

	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	99.5	100	99.75	4.84	4.93	4.89
手動ラベル	98.7	99.0	98.85	3.56	3.94	3.75
市販の合成音	97.5	97.8	97.65	3.12	3.26	3.19

表 4 から、了解度、オピニオンスコアとも高い値となっており、普通名詞に対して十分に実用性のある合成音声が作られたことが分かった。

実験結果を詳しく見てみると、了解度試験ではいくつかの似た音節を間違えている被験者が多かった。被験者が間違えた単語の例を表 5 に示す。なお、() 内の数字は間違えた被験者の数である。

表 5　間違いの例 (1)

	正解	間違いの例
f1	ぐんかん (軍艦)	ぶんかん (2)
f2	こくりつ (国立)	ほくりつ (2)
f3	こっかい (国会)	ほっかい (1), こったい (1)
f4	さいがい (災害)	さいない (1)
f5	たいがい (大概)	たいない (1)
f6	さくじつ (昨日)	かくじつ (1)
f7	かいがい (海外)	たいがい (1)
f8	はんたい (反対)	たんかい (1)

表 5 から分かるように、「ぐ (/gu/)」と「ぶ (/bu/)」、「こ (/ko/)」と「ほ (/ho/)」、「が (/ga/)」と「な (/na/)」、「か (/ka/)」と「さ (/sa/)」と「た (/ta/)」を間違えた被験者が多かった。

オピニオン評価では、アクセント位置のおかしいもの、音節の接続部分が不自然なものに対して評価が低かった。なお、市販の合成機による音声も品質そのものは悪くはなく、話者性を考慮しなければ十分な品質が得られた。

5.3　自動ラベリングの有効性の調査

ラベリング方式の違いによる合成音声の品質への影響を調べた。実験結果を表 6 に示す。

表 6　実験結果 (2)

	了解度　正解率 (%)			オピニオンスコア		
	FTK	FYN	平均	FTK	FYN	平均
手動ラベル	98.7	99.0	98.85	3.56	3.94	3.75
自動ラベル	98.4	99.4	98.90	3.53	3.87	3.70

表 6 から、ラベリング方式の違いによる合成音声の品質に大きな違いはなく、自動ラベリングの有効性が確認できた。

了解度試験では、手動ラベル、自動ラベルともに似たような音節で間違えていた。自動ラベルを用いた合成音声で間違えていた例を表 7 に示す。

表 7　間違いの例 (2)

	正解	間違いの例
f1	ぐんかん (軍艦)	ぶんたん (1)
f3	こっかい (国会)	こったい (1)
f7	かいがい (海外)	かいない (1)
f8	はいせき (排斥)	たいせき (1), かいせき (1)
f9	たいかく (体格)	たいかつ (1)
f10	ばいかい (媒介)	ばいない (1)
f11	でんせつ (伝説)	ぜんせつ (1)

オピニオン評価の結果も、ラベリング方式の違いによる特徴はなかった。傾向としては前節で述べた通り、アクセント位置のおかしいもの、音節の接続部分の不自然なものに対して評価が低かった。

6.　考　　察

6.1　音節の選択方法に関する考察

音節の選択方法に関して、ランダムに選択した場合に評価が悪くなっていた単語を見ると、特に 4 モーラ目に音節継続時間が短い音節を選択した単語で評価が悪くなっていた。例えば「軍艦」(gu/ng/ka/ng) では、ランダムに選択した場合は 4 モーラ目の撥音の音節継続時間が極端に短かったため、接続した場合に 4 モーラ語として正しく聞こえる音声が作成されなかった。その結果、了解度試験で間違えた話者が多く、オピニオンスコアも悪い値となった。

なお FYN については、FTK に比べて録音データのアクセントがあまり明確でなく、その結果、作成された音声へのアクセント型の影響が少なかったため、規則の有無に関わらずオピニオンスコアが高くなったと考えられる。

6.2　固有名詞と普通名詞の結果の比較

固有名詞と普通名詞の結果を比較するために、固有名詞（地名）に対する評価実験の結果を表 8 に示す。この表は参考文献 [2] から得ている。なお参考文献 [2] の実験では、評価ガイダンス文に単語音声を埋め込んだものに対して評価を行っており、単語音声単独での評価とは異なる。

表 8　固有名詞（地名）に対する実験結果

	了解度 正解率 (%)			オピニオンスコア		
	話者 A	話者 B	平均	話者 A	話者 B	平均
自然音声	97.9	99.6	98.8	4.86	4.91	4.89
手動ラベル	97.9	99.1	98.5	4.13	4.03	4.08
市販の合成音	90.9	94.3	92.6	1.76	1.72	1.74

表 4 および表 8 から、固有名詞に対する結果と普通名詞に対する結果は似ており、音節波形接続方式が、固有名詞だけではなく、普通名詞に対しても十分有効であることが分かった。ただし、普通名詞では推論が利くため、音節の了解度では、特に市販の合成音で本研究の結果が参考文献 [2] の結果に比べて高くなっている。

6.3　アクセント型の不自然な音声

過去の研究では、地名ではモーラ数が同じ場合はピッチ周波数の分散が小さいため、アクセント型を意識しなくても良い [2]、と仮定されており、本研究もこれに従った。しかし、本研究では地名ではなく普通名詞を対象としたため、普通名詞では特に重要となるアクセント型を未考慮のままでは極端に不自然な音声が作成されてしまう場合があった。このような音声として、「海外」(ka/i/ga/i) →「ka¯i g¯a i」、「大会」(ta/i/ka/i) →「t¯a|i ka i」、「大概」(ta/i/ga/i) →「t¯a|i ga i」などがあった。

例えば、「海外」(ka/i/ga/i) という音声の場合、音の強弱を表すと、「k¯a|i ga i」となり、アクセントは「ka」の位置にある。一方、本研究で作成した音声合成プログラムでは、この音声を作成するために、

「会員」(ka/i/i/N)
「大学」(da/i/ga/ku)
「大概」(ta/i/ga/i)
「後悔」(ko/u/ka/i)

という音声を選択し、音節波形を切り出した。音声の切り出す部分の音の強弱 $SW(x)$ を見てみると、以下の通りとなる。

「会員」: $SW(/ka/) =$ 弱
「大学」: $SW(/i/) =$ 強
「大概」: $SW(/ga/) =$ 強
「後悔」: $SW(/i/) =$ 弱

よって、「k¯a|i g¯a i」となり、アクセント型の異なった、不自然な音声が作成されてしまった。

これは、ラベルデータにアクセント情報を付加し、合成時にアクセント型を考慮することにより、より自然な合成音声の作成が可能であると考えられる。しかし、全ての音節部品に対するデータベースに対し、$SW =$ 強, $SW =$ 弱 の音節部品がそれぞれ最低１つ以上含まれる必要があり、そのためには、録音件数を増やしてデータベースをさらに拡充することが必要であると考えられる。

6.4　ラベリング精度の影響

実験に使用したデータでは数が少なかったものの、自動音素ラベリングのラベリングデータを用いて音節波形を選択した場合に、音節波形に前後の音節の不要な音が入ったり、逆に必要な部分がうまく入っていないことがあった。その結果、接続した部分が不自然に聞こえる単語があった。特に「軍艦」(gu/ng/ka/ng)、「伝染」(de/ng/se/ng) など、撥音を含む音声に多く見られた。

これは、ラベリング精度の問題であるので、自動音素ラベリングの精度を向上させることによってある程度解決することが可能であると考えられる。

7.　ま　と　め

本研究では、従来の音節波形接続方式による音声合成の普通名詞への有効性について調査した。また、自動ラベリングデータを用いた場合の合成音声への影響を確認した。

音節波形の選択方法に関する実験では、本研究で提案した選択方法を用いた場合、完全にランダムに選択した場合より高いオピニオンスコアが得られることが分かった。

普通名詞に対する有効性の確認のための実験では、合成音声の単語了解度は 98.85 %、オピニオンスコアは 3.75 が得られた。一方、自然音声の単語了解度は 99.75 %、オピニオンスコアは 4.89 であった。自然音声には及ばないものの、了解度の非常に高い合成音声が得られることが分かった。

また、自動ラベリングの有効性の確認に対する実験では、了解度の差は 0.05 %、オピニオンスコアの差は 0.05 で、その差は非常に小さかった。よって、音節波形接続方式に自動ラベリングを用いても問題がないことが分かった。

今後は、考察で述べたアクセント型を考慮した音声合成や波形データの候補が複数出た場合の絞り込みの手法の検討などを行い、さらに自然性の高い合成音声の作成を目指したい。

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," 信学技報, SP96-7, pp45-52 (1996.5).
[2] 水澤 紀子, 村上 仁一, 東田 正信, "音節波形接続による単語音声合成," 信学技報, SP99-2, pp.9-16 (1999.5).
[3] 中川 聖一, 大黒 慶久, "連続音声認識における音韻認識率と文認識率との関係," 信学論, J72-D-II, No.2, pp.207-217 (1989.2).
[4] 古市 千枝子, 相澤 桂, 井上 和彦, 今井 聖, "音声認識におけるルールベース法による話者独立音素セグメンテーション," 音響学会誌 55, pp.707-716 (1999.10).
[5] 鬼山 康人, 野村 康雄, 荒井 和博, 山下 洋一, 北橋 忠宏, 溝口 理一郎, "知識処理に基づく音声自動ラベリングシステム—処理方式の改良と性能評価—," 信学技報, SP90-84, pp.53-60, (1991.1).
[6] 前田 智広, 村上 仁一, 池原 悟, "モーラ情報を用いた音素ラベリング方式の検討," 信学技報, SP2001-53, pp.25-30 (2001.8).
[7] 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研究," 信学技報, SP2001-45, pp.1-5 (2001.8).
[8] 村上 仁一, 水澤 紀子, 東田 正信, "音節波形接続による単語音声合成", 電子情報通信学会論文誌 D-2 採録予定
[9] Hidden Markov Model Toolkit (HTK) http://htk.eng.cam.ac.uk/
[10] 富士通株式会社, "Linux 版音声合成ライブラリー," http://www.createsystem.co.jp/
[11] Praat: doing phonetics by computer http://fonsg3.let.uva.nl/praat/