概要

録音編集方式では、固定部と可変部をつなげた際の違和感を軽減するために、一般に同一話者による音声を大量に必要とする。しかし、実際に大量の音声を同一話者から録音するのは困難である。そこで解決のために「音節波形接続」方式が提案されている。音節波形接続方式により作成された合成音声の品質は、十分実用的であることが過去の論文により報告されている [1][2]。

しかし、従来提案されている音節波形接続方式では、固有名詞や普通名詞を対象にしており、文節単位での音声に関しては有効性が示されていない。

そこで本研究では、音節波形接続方式を用いて文節単位の合成音声を音節の選択条件を変えて作成し、文節に対する有効性を調べた。その結果、了解度はどの合成音声もほとんど変わらず高い値となったが、自然性では自然音声には及ばないものの十分実用可能な品質の音声が作成された。

目 次

図 目 次

表 目 次

1　はじめに

近年、音声ガイダンスの文章のように、ガイダンスを利用するユーザーに依存しない部分（固定部）と、ユーザーに依存する部分（可変部）が存在する文章を合成する方法として、録音編集方式が広く使われている。これは、固定部と可変部をそれぞれ別々に録音しておき、出力内容に合わせて、可変部を固定部に挿入することによって、出力音声を作成する方法である。

例えば、「次の交差点を○○です。」というガイダンス文を出力する場合、「次の交差点を」と「です。」は固定部であり、「○○」の部分は、「直進」「右折」「左折」のような言葉を可変部として準備し、状況に応じて固定部に挿入することになる。

録音編集方式では、固定部と可変部をつなげた際の違和感を軽減するために、一般に同一話者による音声を大量に必要とする。しかし、実際に大量の音声を同一話者から録音するのは困難である。そこで解決のために「音節波形接続」方式による音声合成が提案されている。音節波形接続方式により作成された合成音声の品質は、十分実用的であることが過去の論文により報告されている [1][2]。

しかし、従来提案されている音節波形接続方式では固有名詞や普通名詞などの名詞を対象としており、文節単位での音声に対しては有効性が示されていない。

そこで、本研究では、音節波形接続方式を用いて文節単位の合成音声を作成した。そして、従来の音節による方法、アクセント位置を考慮する方法、FFT を音節選択に利用する方法の 3 つの音節選択方法で音声を作成し、それぞれの方法において文節に対する有効性と問題点を調べた。その結果、了解度はどの音節選択方法を用いた場合も自然音声とほとんど変わらなかったが、自然性は自然音声には及ばないものの、十分実用的であることが分かった。

以下、本稿では、まず 2 章で音節波形接続方式について説明する。そして 3 章で従来の方法での音声合成、4 章でアクセント位置を音節選択に使用した音声合成、5 章で FFT を用いた音節選択による音声合成の実験と結果について述べる。実験結果に対する考察については 6 章で述べる。

2　音声波形接続による音声合成

2.1　音節波形接続方式

音節波形接続方式は、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取り出し、接続することによって合成音声を作成する手法である。この方式では話者の音声波形をそのまま使用することができる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。

この音節波形接続方式により作成された合成音声の品質は、過去の論文から水澤氏により固有名詞 (地名) に対して有効性が示されている [1]。また、石田氏により普通名詞に関して十分実用的であることが報告されている [2]。

2.2　従来の音節波形接続方式による音声合成

従来の音節波形接続方式は合成対象語から以下の情報により音節部品ラベルを選択する。

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：文節中のモーラ位置
M：文節のモーラ数

以降、本稿では音節部品ラベルを「$Sy(P,N)_{m,M}$」の形式で表現する。例えば「高いに」(ta/ka/i/ni) という音声に対する音声部品ラベル列は、「$ta(,k)_{1,4}$」「$ka(a,i)_{2,4}$」「$i(a,n)_{3,4}$」「$ni(i,)_{4,4}$」となる。

各音節部品ラベルに一致するような音節部品を取り出す。そして、音声の開始時間と終了時間を元に波形データを切り出し、単純に接続して合成する。

3　評価実験

3.1　実験環境

音声データベースに ATR の単語発話データベースに Aset1 の中で、文が文節単位で区切られて発話されている DSB(115 文) で実験を行なう。本研究は文節についての有効性を調べるため、データベース DSB から取り出せる文節の中から単語のみの文節を除いた 639 文節を音声作成に使用する。そして、FTK,FYN の 2 話者で作成したそれぞれ 10 文節について、自然音声、本研究で作成した合成音声、市販の合成機の合成音声 (市販の音声) を準備する。

3.1.1　市販の合成機による合成音声の作成

本研究では、市販の音声として、富士通株式会社の「Linux 版日本語音声合成ライブラリー」[3] によって作成したものを使用する。作成では、まず、評価対象の単語の漢字表記をライブラリのプログラムで表音文字に変換する。そして、変換された文字列に話者の声質情報を付加し、ライブラリの合成プログラムによって音声を作成する。

例として「高いに」という音声を作成する場合について説明する。
まず、漢字表記から表音表記に変換する。

「高いに」→「タカ＊イニ,」

そして、話者の声質情報を付加する。声質情報には性別、声の高低、発声速度、音量、広域強調、抑揚の情報がある。声の高さは 1〜5 で設定でき、値が大きくなるほど声が高くなる。また、発声速度は 0〜9 で設定でき、値が大きくなるほど速度が早くなる。本研究では、話者に合わせ、FTK は声の高さも早さも普通なので「女声３」「速さ５」とし、FYN は FTK より発声速度が遅いため「女声３」「速さ２」という情報を付加した。

「＜女声３＞＜速さ５＞タカ＊イニ,」

最後に、合成プログラムに表音文字列を与え、合成音声を作成する。

3.2 評価音声

評価に用いる音声は準備した自然音声、本研究で作成した合成音声、市販の合成機の合成音声のそれぞれ 10 文節を固定部に挿入し、文の形にした各 12 文とする。下に評価に用いる文章を示す。太文字になっているのが評価対象の文節である。

- 原稿の 提出は 締め切りに 遅れない **ように** して ください。
- 参加 登録の 申し込み 用紙は どの **ように** して 手に 入れれば よろしいのでしょうか
- 英語の 発表は 日本語に 自動 翻訳される **ように** なって います。また、日本語の 発表は 英語に 翻訳されます。
- 会場は **京都に** ございます 国際 会議場を 全館 貸切って 行ないます。
- ですが、少々 高めに なって おりますので、**京都の** 民宿 協会の 資料も お送りする ようには して います。
- これも **資料の** 中に 説明が ありますが、一応 その 時の 換算を 調べて戴く ことに なると思います。
- **高いに** 倉庫の 倉の 字です。
- 発表して いただくと いう 事に なりました **場合に** 資料を 送付 いたしますが、この 中に 詳しい 説明図の ような ものも 含まれる 予定です。
- 聴講のみの **場合は** 当日の 受付けも 可能で、予稿集代を 含めた 費用は 3 万 5 千円 かかります。
- 第一回 通訳 電話 国際 会議に 参加の 登録を ご希望される 方は、所定の**用紙に** 住所 氏名と 発表 聴講の 別を 明記して、国際 会議 事務局までお申し込み 下さい。
- そうですか。わかりました。では、申し込み **用紙の** 件、よろしく お願い 申し 上げます。
- 最終的に **論文の** 受諾 通知は いつに なるんですか。

3.3 評価実験

合成音声の評価のために、音声研究に関わった経験のない８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は以下に示す。

(1) 了解度試験

文節単位の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、本研究で作成する合成音声、市販の合成機で作成した合成音声の評価用の文をランダムにヘッドフォンから被験者に聞かせ、評価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

文節単位の音声の自然性を調べるために、オピニオン評価を行なう。オピニオン評価では、自然音声、本研究で作成する合成音声、市販の合成機で作成した合成音声の評価用の文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。

そして、評価実験では評価対象となる文節の部分の空欄に了解度試験では聞こえた音を、オピニオン評価では自然性の 1〜5 の数字を書き込んでもらう。その例を下に示す。

例： 原稿の 提出は 締め切りに 遅れない (　) して ください。

4　従来の音節選択での文節に対する有効性の確認

4.1　合成音声の作成

まず、合成対象となる文節から音声部品ラベル列を作成する。そして、各音節部品ラベルに一致するような条件で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。

切り出す波形データを選択する方法として、音節部品データベースを利用する。音節部品データベースには、異なる音節部品ラベルごとに作成したデータベースファイルを格納する。データベースファイル名は「$M/m.Sy.P.N$」という名前にする。M,m,Sy.P.Nはそれぞれ以下のとおりである。

M：文節のモーラ数
m：文節中のモーラ位置
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)

各データベースファイルには、以下の情報を格納する。
Fn：録音ファイル番号
St：音節開始時間 (ms)
Ed：音節終了時間 (ms)
Cs：文節内の音節開始時間 (ms)

音節波形接続方式では、取り出す波形データ候補が複数出る場合があり、選択する方法が問題となる。本実験では、第１音節についてはデータベース内の先頭要素を選択した。また、それ以外の音節については、取り出す音節の文節内の音節開始時間が直前の音節までの音声の継続時間の和により近いものを選択した。それでも候補が出た場合は、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した。

この音節の選択方法を簡単に示したのが下の図１である。

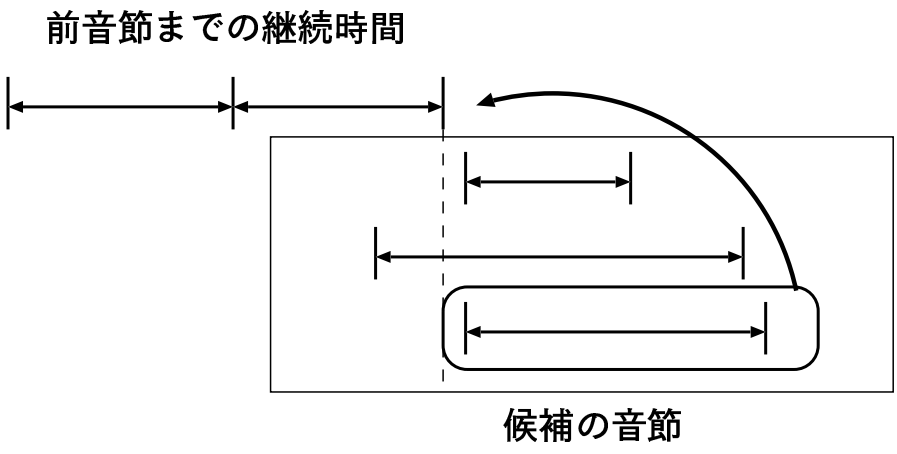

図 1: 音節の選択方法

例として、「高いに」(ta/ka/i/ni) という音声を合成する場合を説明する。概略を図 2 に示す。

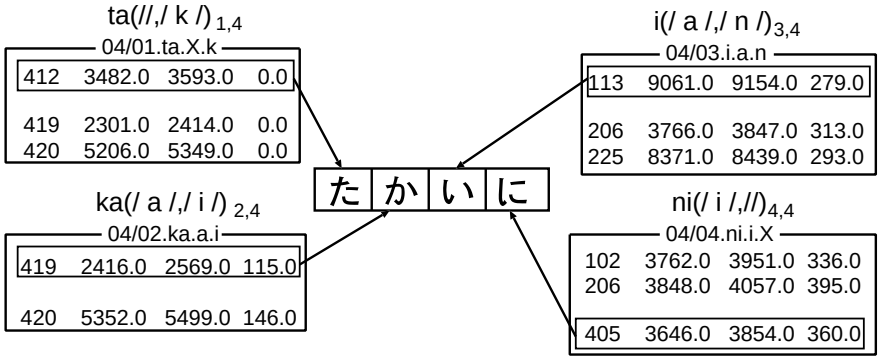

図 2: 「高いに (ta/ka/i/ni)」の作成 (従来の音節選択)

まず、音声部品ラベルを作成する。例の場合「ta/ka/i/ni」であるので、「$ta(,k)_{1,4}$」「$ka(a,i)_{2,4}$」「$i(a,n)_{3,4}$」「$ni(i,)_{4,4}$」という音節部品ラベル列が得られる。

次に音節部品データベースファイルから、各音節ごとに切り出せるデータを選択する。

第 1 音節については、データベースファイルの先頭要素を選ぶことにする。
例では、「Fn=412,St=3482.0,Ed=3593.0,Cs=0」を選ぶ。

第 2 音節以降の音節については、

(1) 音節開始時間が直前の音節までの音声の継続時間の和に最も近いもの

(2) 音節継続時間がより長いもの

を優先的に選ぶ。

まず、第 2 音節は、第 1 音節の音声継続時間が 111.0ms なので、開始時間が最も近いものを探し、「Fn=419,St=2416.0,Ed=2569.0,Cs=115」を選択する。

次に、第 3 音節は、第 2 音節までの音声継続時間の和が 264.0ms なので、開始時間が最も近いものを探し、「Fn=113,St=9061.0,Ed=9154.0,Cs=279」を選択する。

最後に、第 4 音節は、第 3 音節までの音声継続時間の和が 357.0ms なので、開始時間が最も近いものを探し、「Fn=405,St=3646.0,Ed=3854.0,Cs=360」を選択する。

切り出すデータが決まったら、ファイル番号をもとに音声データを読み込み、該当区間の音節波形を切り出す。そして、それらを接続し、合成する。

4.2 従来の音節選択で作成した音声一覧

従来の音節選択で作成した音声を表1、2で示す。

表 1: 作成した音声 (FTK, 従来の音節選択)

音声	音節 1	音節 2	音節 3	音節 4	音節 5
ように	ような (**yo-u**/na)		特に (to/ku/**ni**)		
京都に	京都は (**kyo-u**/to/ha)		京都の (kyo-u/**to**/no)	英語に (e-i/go/**ni**)	
京都の	京都は (**kyo-u**/to/ha)		京都に (kyo-u/**to**/ni)	英語の (e-i/go/**no**)	
資料の	資料を (**shi**/ryo-u/wo)	資料に (shi/**ryo-u**/ni)		太郎の (ta/ro-u/**no**)	
高いに	高めに (**ta**/ka/me/ni)	高いと (ta/**ka**/i/to)	場合に (ba-a/**i**/ni)	会議に (ka/i/gi/**ni**)	
場合に	場合は (**ba-a**/i/ha)		市内の (shi/na/**i**/no)	会議に (ka/i/gi/**ni**)	
場合は	場合に (**ba-a**/i/ha)		議題は (gi/da/**i**/ha)	会議は (ka/i/gi/**ha**)	
用紙に	用紙を (**yo-u**/shi/wo)		用紙の (yo-u/**shi**/no)	形に (ka/ta/ti/**ni**)	
用紙の	用紙を (**yo-u**/shi/wo)		用紙に (yo-u/**shi**/ni)	通知の (tsu-u/chi/**no**)	
論文の	論文を (**ro**/N-bu/N/wo)	論文は (ro/**N-bu**/N-ha)		英文の (e-i/bu/**N-no**)	

表 2: 作成した音声 (FYN, 従来の音節選択)

音声	音節 1	音節 2	音節 3	音節 4	音節 5
ように	ような (**yo-u**/na)		特に (to/ku/**ni**)		
京都に	京都は (**kyo-u**/to/ha)		京都の (kyo-u/**to**/no)	英語に (e-i/go/**ni**)	
京都の	京都は (**kyo-u**/to/ha)		京都に (kyo-u/**to**/ni)	英語の (e-i/go/**no**)	
資料の	資料を (**shi**/ryo-u/wo)	資料に (shi/**ryo-u**/ni)		太郎の (ta/ro-u/**no**)	
高いに	タクシーを (**ta**/ku/shi/wo)	高いと (ta/**ka**/i/to)	機械に (ki/ka/**i**/ni)	会議に (ka/i/gi/**ni**)	
場合に	場合は (**ba-a**/i/ha)		高いに (ta/ka/**i**/ni)	会議に (ka/i/gi/**ni**)	
場合は	場合に (**ba-a**/i/ha)		議題は (gi/da/**i**/ha)	私は (wa/ta/shi/**ha**)	
用紙に	用紙を (**yo-u**/shi/wo)		用紙の (yo-u/**shi**/no)	会議に (ka/i/gi/**ni**)	
用紙の	用紙を (**yo-u**/shi/wo)		用紙に (yo-u/**shi**/ni)	市内の (shi/na/i/**no**)	
論文の	論文を (**ro**/N-bu/N/wo)	論文に (ro/**N-bu**/N-ni)		英文の (e-i/bu/**N-no**)	

4.3 実験結果

従来の音節選択による文節に対する有効性を確認した。実験結果を表 3 に示す。

表 3: 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	96	99	98	4.5	4.7	4.6
合成音声	100	98	99	3.0	3.0	3.0
市販の音声	77	76	77	1.6	1.9	1.8

表 3 から、文節を対象にした場合でも、了解度は 99 %となり自然音声とほぼ同じ高い値になっており、オピニオンスコアは自然音声には及ばないが 3.0 となり、実用性のある合成音声が作られたことが確認された。

オピニオン評価で、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。

参考までに、固有名詞 (地名) に対する評価実験の結果を [1] より引用し、表 4 に示す。なお、[1] の実験本実験と同様に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なっている。

表 4: 固有名詞に対する実験結果

	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
合成音声	97.9	99.1	98.5	4.13	4.03	4.08
市販の音声	90.9	94.3	92.6	1.76	1.72	1.74

表 5 から分かるように固有名詞に対する結果の方が文節を対象にした本実験よりも自然音声に近い良い結果が出ているが、文節に対する結果も似ており、音節波形方式が、文節にも十分に有効であると言える。

5　アクセント位置情報を含んだ音節選択

従来の研究では、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良い [1]、と仮定されており、４の実験もしたがった。しかし、本研究では文節を対象としているためアクセントを考慮しなかったために不自然な音声が作成されてしまう場合があった。

例えば「京都に」(kyo-u/to/ni) アクセントは「kyo」に置かれるべきである。音声の強弱を表すと、「k͡yo͡|u to ni」となる。一方４の実験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、下の音声を選択し、音節波形を取り出した。

「京都は」(kyo-u/to/ha)
「京都の」(kyo-u/to/no)
「英語に」(e-i/go/ni)

これらの音声の切り出す部分の音声の強弱 SW(x) を見てみると以下の通りになる。

「京都は」: SW(/kyo/) ＝ 強,SW(/u/) ＝ 弱
「京都の」: SW(/to/) ＝ 弱
「英語に」: SW(/ni/) ＝ 強

よって、「k͡yo͡|u to|n͡i」となり、アクセントが異なった、不自然な音声が作成されてしまった。

これは、アクセント位置を考慮することにより、より自然な音声の作成が可能であると考えられる。そこで今回はアクセント位置を音節選択に利用した音声について実験を行なう。

5.1 アクセント位置情報の有効性の確認

3章の実験と同様に切り出す波形データを選択する方法として、音節部品データベースを利用する。音節部品データベースには、異なる音節部品ラベルごとに作成したデータベースファイルを格納する。データベースファイル名は「$M/m.Sy.P.N.a$」という名前にする。M,m,Sy,P,N,aはそれぞれ以下のようになる。

M：文節のモーラ数
m：文節中のモーラ位置
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)
a：アクセントがある音素のモーラ番号

各データベースファイルには3章の実験と同様に、以下の情報を格納する。
Fn：録音ファイル番号
St：音節開始時間 (ms)
Ed：音節終了時間 (ms)
Cs：文節内の音節開始時間 (ms)

音声作成に使用する音節の選択方法は3章の実験と同様に、第1音節についてはデータベース内の先頭要素を選択した。また、それ以外の音節については、取り出す音節の文節内の音節開始時間が直前の音節までの音声の継続時間の和により近いものを選択した。それでも候補が出た場合は、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した。

例として、「高いに」(ta/ka/i/ni) という音声を合成する場合を説明する。概略を図 3 に示す。

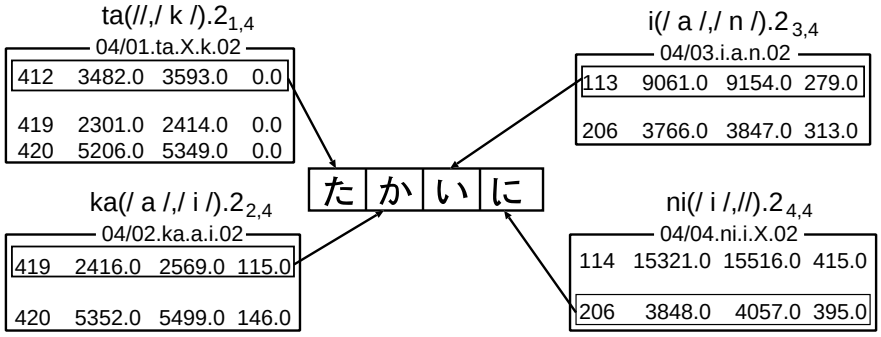

図 3: 「高いに (ta/ka/i/ni)」の作成 (アクセント位置)

まず、音声部品ラベルを作成する。例の場合「ta/ka/i/ni」であるので、「$ta(,k).2_{1,4}$」「$ka(a,i).2_{2,4}$」「$i(a,n).2_{3,4}$」「$ni(i,).2_{4,4}$」という音節部品ラベル列が得られる。

次に音節部品データベースファイルから、各音節ごとに切り出せるデータを 3 章の実験と同様の方法で選択する。

切り出すデータが決まったら、ファイル番号をもとに音声データを読み込み、該当区間の音節波形を切り出す。そして、それらを接続し、合成する。

5.2 アクセント位置を考慮した音節選択で作成した音声一覧

アクセント位置を考慮して作成した音声を表5、6に示す。

表 5: 作成した音声 (FTK, アクセント位置を考慮した音節選択)

音声	音節 1	音節 2	音節 3	音節 4	音節 5
ように	ような (**yo-u**/na)		特に (to/ku/**ni**)		
京都に	京都は (**kyo-u**/to/ha)		京都の (kyo-u/**to**/no)	言語に (ge/N/go/**ni**)	
京都の	京都は (**kyo-u**/to/ha)		京都に (kyo-u/**to**/ni)	倉庫の (so-u/ko/**no**)	
資料の	資料を (**shi**/ryo-u/wo)	資料に (shi/**ryo-u**/ni)		太郎の (ta/ro-u/**no**)	
高いに	高めに (**ta**/ka/me/ni)	高いと (ta/**ka**/i/to)	場合に (ba-a/**i**/ni)	機械に (ki/ka/i/**ni**)	
場合に	場合は (**ba-a**/i/ha)		高いに (ta/ka/**i**/ni)	形に (ka/ta/chi/**ni**)	
場合は	場合に (**ba-a**/i/ha)		議題は (gi/da/**i**/ha)	行くには (i/ku/ni/**ha**)	
用紙に	用紙を (**yo-u**/shi/wo)		用紙の (yo-u/**shi**/no)	会議に (ka/i/gi/**ni**)	
用紙の	用紙を (**yo-u**/shi/wo)		用紙に (yo-u/**shi**/ni)	通知の (tsu/u/chi/**no**)	
論文の	論文を (**ro**/N-bu/N/wo)	論文は (ro/**N-bu**/N-ha)		英文の (e-i/bu/**N-no**)	

表 6: 作成した音声 (FYN, アクセント位置を考慮した音節選択)

音声	音節 1	音節 2	音節 3	音節 4	音節 5
ように	ような (**yo-u**/na)		特に (to/ku/**ni**)		
京都に	京都は (**kyo-u**/to/ha)		京都の (kyo-u/**to**/no)	言語に (ge/N/go/**ni**)	
京都の	京都は (**kyo-u**/to/ha)		京都に (kyo-u/**to**/ni)	倉庫の (so-u/ko/**no**)	
資料の	資料を (**shi**/ryo-u/wo)	資料に (shi/**ryo-u**/ni)		太郎の (ta/ro-u/**no**)	
高いに	高めに (**ta**/ka/me/ni)	高いと (ta/**ka**/i/to)	機械に (ki/ka/**i**/ni)	形に (ka/ta/chi/**ni**)	
場合に	場合は (**ba-a**/i/ha)		高いに (ta/ka/**i**/ni)	機械に (ki/ka/i/**ni**)	
場合は	場合に (**ba-a**/i/ha)		議題は (gi/da/**i**/ha)	私は (wa/ta/shi/**ha**)	
用紙に	用紙を (**yo-u**/shi/wo)		用紙の (yo-u/**shi**/no)	会議に (ka/i/gi/**ni**)	
用紙の	用紙を (**yo-u**/shi/wo)		用紙に (yo-u/**shi**/ni)	会議の (ka/i//gi/**no**)	
論文の	論文を (**ro**/N-bu/N/wo)	論文は (ro/**N-bu**/N-ha)		英文の (e-i/bu/**N-no**)	

5.3 実験結果

アクセント情報を考慮した音節選択による合成音声の。実験結果を表7に示す。

表 7: 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
自然音声	96	99	98	4.5	4.7	4.6
アクセント情報なし	100	98	99	3.0	3.0	3.0
アクセント情報あり	96	98	97	3.2	3.0	3.1

表7から、了解度は自然音声とほぼ同程度の高い値となっているが、オピニオンスコアはアクセント情報を用いた音声の方が若干高い値となったが、値としてほとんど違いがみられなかった。

これは今回はアクセント位置を考慮した音節選択による合成音声を作成したが、アクセント位置が同じでもアクセント型が違うものがあるため、不適切な音声を選択してしまった場合があった。そして結果としてアクセントが不自然な音声が作成されてしまいオピニオンスコアに差があまり見られなかった。

6　FFTを用いた音節選択

音節を選ぶ条件として、４章の実験では「文節のモーラ数」、「単語中のモーラ位置」、「音節」、「直前の音素」「直後の音素」を用いており、５章の実験ではさらに「アクセント位置」を考慮するため、音節選択の条件が厳しくなり作成可能な音声が少なくなってしまう。そこで、本実験では作成できる音声を増やすために「文節中のモーラ数」と「アクセント位置」は考慮せずに、音節の候補を選ぶ。また、今回はあらかじめ作成する音声の音声波形が分かっているという前提でFFTを使用して音節を選択し音声を作成する。そうすることで、作成される音声がどのくらい増加するか、またFFTを音節選択に利用し似ている音声波形を選択することでどの程度の品質が得られるかを調べる。

6.1 FFTを用いた合成音声の作成

３、４章の実験と同様に切り出す波形データを選択する方法として、音節部品データベースを利用する。音節部品データベースには、異なる音節部品ラベルごとに作成したデータベースファイルを格納する。データベースファイル名は「$m.Sy.P.N$」という名前にする。m,Sy,P,Nはそれぞれ以下のとおりである。

m：文節中のモーラ位置
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)

各データベースファイルには以下の情報を格納する。
Fn：録音ファイル番号
St：音節開始時間 (ms)
Ed：音節終了時間 (ms)

音声作成に使用する音節の選択方法は、作成する元の音と候補の音にそれぞれ FFT を掛け、その差が最も小さいものを選択する。

例として、「高いに」(ta/ka/i/ni) という音声を合成する場合を説明する。

まず、音声部品ラベルを作成する。例の場合「ta/ka/i/ni」であるので、「$ta(,k).2_1$」「$ka(a,i).2_2$」「$i(a,n).2_3$」「$ni(i,).2_4$」という音節部品ラベル列が得られる。

次に音節部品データベースファイルから、各音節ごとに元になる音と候補の音の FFT の差が最も小さいものを切り出すデータとして選択する。

切り出すデータが決まったら、ファイル番号をもとに音声データを読み込み、該当区間の音節波形を切り出す。そして、それらを接続し、合成する。

6.2 音節選択に FFT を用いて作成した音声一覧

音節選択に FFT を用いて作成した音声を表 8、9 に示す。

表 8: 作成した音声 (FTK,FFT による音節選択)

音声	音節 1	音節 2	音節 3	音節 4	音節 5
ように	ような (**yo-u**/na)		いつに (i/tsu/**ni**)		
京都に	京都駅から (**kyo-u**/to/e/ki/ka/ra)		京都の (kyo-u/**to**/no)	英語に (e-i/go/**ni**)	
京都の	京都駅から (**kyo-u**/to/e/ki/ka/ra)		京都に (kyo-u/**to**/ni)	倉庫の (so-u/ko/**no**)	
資料の	資料を (**shi**/ryo-u/wo)	資料に (shi/**ryo-u**/ni)		太郎の (ta/ro-u/**no**)	
高いに	高いですね (**ta**/ka/i/de/su/ne)	高いと (ta/**ka**/i/to)	市内の (shi/na/**i**/no)	形に (ka/ta/chi/**ni**)	
場合に	場合は (**ba-a**/i/ha)		市内の (shi/na/**i**/no)	形に (ka/ta/chi/**ni**)	
場合は	場合に (**ba-a**/i/ha)		議題は (gi/da/**i**/ha)	用紙は (yo-u/shi/**ha**)	
用紙に	用紙を (**yo-u**/shi/wo)		用紙の (yo-u/**shi**/no)	形に (ka/ta/chi/**ni**)	
用紙の	用紙を (**yo-u**/shi/wo)		用紙に (yo-u/**shi**/ni)	通知の (tsu/u/chi/**no**)	
論文の	論文は (**ro**/N-bu/N/ha)	論文が (ro/**N-bu**/N-ga)		英文の (e-i/bu/**N-no**)	

表 9: 作成した音声 (FYN,FFT による音節選択)

音声	音節 1	音節 2	音節 3	音節 4	音節 5
ように	ような (**yo-u**/na)		特に (to/ku/**ni**)		
京都に	京都での (**kyo-u**/to/de/no)		京都の (kyo-u/**to**/no)	言語に (ge/N/go/**ni**)	
京都の	京都駅から (**kyo-u**/to/e/ki/ka/ra)		京都に (kyo-u/**to**/ni)	倉庫の (so-u/ko/**no**)	
資料の	資料を (**shi**/ryo-u/wo)	資料に (shi/**ryo-u**/ni)		太郎の (ta/ro-u/**no**)	
高いに	高輪は (**ta**/ka/na/wa/ha)	高いと (ta/**ka**/i/to)	市内の (shi/na/**i**/no)	会議に (ka/i/gi/**ni**)	
場合に	場合は (**ba-a**/i/ha)		市内の (shi/na/**i**/no)	会議に (ka/i/gi/**ni**)	
場合は	場合には (**ba-a**/i/ni/ha)		議題は (gi/da/**i**/ha)	時期には (ji/ki/ni/**ha**)	
用紙に	用紙は (**yo-u**/shi/ha)		用紙の (yo-u/**shi**/no)	会議に (ka/i/gi/**ni**)	
用紙の	用紙を (**yo-u**/shi/wo)		用紙に (yo-u/**shi**/ni)	所定の (sho/te/i/**no**)	
論文の	論文が (**ro**/N-bu/N/ga)	論文に (ro/**N-bu**/N-ni)		英文の (e-i/bu/**N-no**)	

6.3 実験結果

FFT を用いた音節選択による合成音声の有効性を確認した。実験結果を表 10、表 11 に示す。

表 10: 作成可能文節数

	FTK	FYN
合成音声 (従来)	17	17
合成音声 (アクセント)	11	11
合成音声 (FFT)	87	82

表 11: 実験結果 (3)

	了解度　正解率 (%)			オピニオンスコア		
	FTK	FYN	平均	FTK	FYN	平均
合成音声 (従来)	100	98	99	3.0	3.0	3.0
合成音声 (アクセント)	96	98	97	3.2	3.0	3.1
合成音声 (FFT)	96	97	97	3.0	3.2	3.1

表 10 から、音節候補の選択条件からモーラ数の情報を除くことで作成可能な音声の数が大幅に増えることを確認した。また、FFT を音素選択に用いることで、音節選択の条件を緩めても、了解度、オピニオン評価ともにアクセント情報を音節選択に利用して作成した音声と同程度の値となり、FFT の音節選択における有効性を確認した。

7　考察

7.1　アクセント位置を考慮した合成音声

従来の音節選択ではアクセントが不自然な音声が作成されため、本研究では音節選択にアクセント位置を考慮した実験を行なった。しかし、アクセント位置を考慮したにもかかわらずアクセントが不自然に聞こえるものがあった。これは、アクセント位置が同じであってもアクセント型が違うものがあるため不適切な音節を選択したためだった。

音節「高いに」について図 4 で示す。

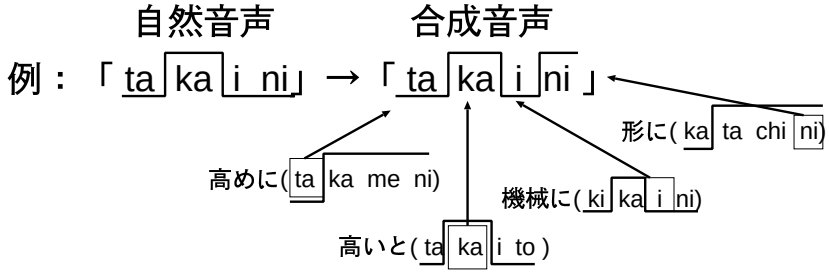

図 4: アクセント型のおかしな音

「高いに」という音節はアクセント位置が 2 モーラ目にあるため、アクセントを考慮した場合にはアクセント位置が 2 モーラ目にあるものを選択する。そしてこの場合に 4 モーラ目の「に」は「形に」の 4 モーラ目から作成される。しかし、「高いに」の「に」は低い音であるのに対して、「形に」の「に」は高い音を用いるためにアクセントがおかしくなり、不自然に聞こえてしまっていた。

これはさら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が作成可能だと考えられる。

本研究では文節の作成数が少なく評価対象の文が１２文となった。さらにアクセント型を考慮するとなると、音節選択の条件が増えてさらに作成可能な文節数が減少することになる。そのため、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。

7.2　FFT を音節選択に利用した合成音声

FFT を用いて波形が似たものを選択することで、音節選択の条件を緩めてもアクセント位置を考慮したものと同程度の品質の音声が得られた。しかし、作成された音声を聞いてみると、アクセントが不自然なものが含まれていた。FFT を使用してもアクセントは考慮できないので、やはりアクセント型を音節選択に使用することが必要だと考えられる。

8　まとめ

本研究では、名詞でしか有効性が示されていない従来の方法での音節波形接続を、音節に対して行ない有効性を調査した。そしてアクセント位置を考慮して作成した音声についても有効性を調べた。また、今回はデータベースに収録されている音節数が少なく作成できる音声の数が少ないため、従来の音節選択の条件からモーラ数の情報を除いき、作る音声の音声波形が分かっているという前提で FFT を用いて作成した音声について実験を行ない、作成音節数がどのくらい増え、どの程度の品質が得られるかを調べた。

従来の音節選択を用いた音声は了解度は自然音声と同程度の 99 %となり、自然性では自然音声には及ばないものの 3.0 となり、波形接続方式の文節での有効性が確認された。

アクセントを音節選択に用いた場合では、了解度は従来の方法と同様に高い値となった。しかし、差が出ると予想していたオピニオンスコアにおいてアクセント位置を考慮したことで 0.1 だけしか自然性が上昇せず、値として違いはほとんど見られなかった。

また、FFT を用いた場合は、音節選択条件からモーラ数の情報を除いたことで作成可能な音声数が増えた。また、自然性の面ではアクセントを考慮して作成した音声と同程度の品質の音声が作成できた。

本研究で作成した音声と自然音声とを比較すると、了解度に関してはほぼ同じ値となっている。しかし自然性に関しては、自然音声のオピニオンスコアは 4.6 であるのに対し、本研究で作成した音声はどれも 3 程度とまだ自然性に関しては自然音声には及ばなかった。

今後は、考察で述べたアクセント型を考慮した合成音声など、さらに自然性の高い合成音声の作成をしたい。また、今回はデータベースに収録されている音節数が少ないために、作成できる音節の数が少なくなったり、候補として選択される音節の数が少ないため作成した音声の品質にそれほど違いが見られなかったので、音声データベースの拡充も行ないたい。

謝辞

本論文作成に関して、ご指導を賜わりました池原悟教授、村上仁一助教授、徳久雅人助手に対して感謝します。また、聴覚実験にご協力いただいた計算機C研究室の方々感謝します。特に、音声合成に関していろいろと助言をくださった石田隆浩氏には大変お世話になりました。

参考文献

[1] 水澤, 村上, 東田：音節波形接続による単語音声合成, 信学技報, SP99-2(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,(2002-05)

[3] 富士通株式会社：Linux 版日本語音声合成ライブラリー　http://www.createsystem.co.jp/