

概要

音声合成の手法として近年注目されている波形接続型音声合成は，大量の録音音声から音素や音節を単位とした波形素片を取り出し，接続することによって合成音声を作成する．信号処理せず接続することで，話者性と高い自然性を保てる特徴がある．また，前後音素環境やモーラ情報などの付加的な情報を用いることで，品質が向上することが知られている [1] ．

例えば，音節を単位とした音節波形接続方式 [1] では，地名を対象に実験した結果，実用的な品質が得られたことが報告されている．また，同様の手法を普通名詞に適用した場合も，明瞭性の高い合成音声を作成できたことが示されているが，アクセント型のばらつきによる自然性劣化も指摘されている [2] ．

また，波形接続型音声合成では基本的に信号処理を行わないため，素片単位や接続位置，最適な素片を選び出す方法なども非常に重要となる．

そこで本研究では，普通名詞を対象として，音素，または音節を素片単位とした場合に，アクセントを考慮することで合成音声の品質をどの程度改善できるかについて調査した．その結果，聴覚実験における合成音声の単語理解度は，音素単位の場合で 99.3 % ，音節単位の場合で 99.6 % が得られた．また，オピニオンスコアはそれぞれ 4.1 ， 4.3 であった．アクセントを考慮することで，オピニオンスコアはそれぞれ 0.4 ， 0.3 の向上となり，自然性の向上のためにアクセントが有効であることが分かった．

一方，自然音声の単語理解度は 99.8 % ，オピニオンスコアは 4.9 であった．アクセントの考慮によってオピニオンスコアは上昇したが，自然音声には及ばなかった．

目次

1	はじめに	5
2	波形接続型単語音声合成	7
2.1	モーラ情報, アクセントと F_0 周波数	7
2.2	波形接続型単語音声合成の概説	10
2.3	合成音声の例	11
3	評価実験	14
3.1	実験環境	14
3.2	波形接続に関する補則	14
3.3	評価方法	15
4	実験結果	16
5	考察	17
5.1	自然音声と合成音声の差	17
5.2	了解度試験の解析	19
5.3	オピニオン評価の解析	20
5.4	素片単位について	23
5.5	素片接続位置について	23
6	まとめ	24

目 次

1	話者 MAU における 4 モーラ語名詞 (約 1,600 件) の F_0 周波数分布	8
2	話者 MAU における 4 モーラ語名詞の F_0 周波数分布 (アクセント別)	8
3	話者 FTK における 4 モーラ語名詞 (約 1,600 件) の F_0 周波数分布	9
4	話者 FTK における 4 モーラ語名詞の F_0 周波数分布 (アクセント別)	9
5	「伝説」(自然音声)	12
6	「伝説」(音素単位)	12
7	「伝説」(音節単位)	12
8	「身分」(自然音声)	13
9	「身分」(音素単位)	13
10	「身分」(音節単位)	13
11	「瓦」(音節単位 , アクセント考慮 , 音素境界接続)	18
12	「瓦」(音素単位 , アクセント考慮 , 音素境界接続)	18
13	「瓦」(音素単位 , アクセント考慮 , 音素中心接続)	18
14	「会話」(自然音声)	21
15	「会話」(合成音声)	21
16	「結局」(自然音声)	22
17	「結局」(合成音声)	22

表目次

1	実験結果 (1): 接続位置 = 音素境界	16
2	実験結果 (2): 接続位置 = 音素中心	16
3	了解度試験における間違いの例	19
4	了解度試験における間違いの集計結果 (一部)	19

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　　でよろしいですか」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が入挿される．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」などの駅名や建物名などの単語音声が可変部「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，可変部に挿入する単語が増大した場合，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，固定部と可変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．例えば，音素や音節，CV，VCV を単位とした規則音声合成がある．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 [3][4]．しかし，いずれの場合においても，直接人の声を録音した音声のように，高い品質を安定して得ることが難しい点が問題である [5]．

一方，録音した音声波形の一部（以下「音声素片」）を用いて別の音声を合成する方法があり，一般に，波形接続型音声合成と呼ばれる．

波形接続型音声合成は，音声素片を取り出し，接続することによって合成音声を作成する．接続単位については，音素，CV，VCV，CVC など，様々な単位が提案されている [6]．

いずれの場合においても，基本的に信号処理を行わず，取り出した波形をそのまま用いるため，話者の声の特徴（以下，「話者性」）や高い自然性を保つことが可能である．また，複数の接続単位を用いたり，前後音素環境やモーラ数，モーラ位置（以下，モーラ数とモーラ位置を合わせて「モーラ情報」）を考慮することによって品質が向上することが確認されている [1][7][8]．最近では素片選択の際に HMM を用いる方法もある [4]．

しかし，波形接続型音声合成においては韻律の扱いが問題となる．尤も，波形接続型音声合成に限らず，一般に音声合成において，韻律制御は重要な課題であるが [9]，音声合成の対象として小さな単位である単語を合成する場合においては，地名などの固有名詞では 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することが可能である [1]．しかし，より一般的な普通名詞では，例えば「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声素片が選択される場合がある．過去の研究において，アクセントを未考慮であったために不自然な音声を作成される場合があったことが報告されている [2]．

以上より本研究では，普通名詞を対象として，音素または音節を素片単位とした場合に，アクセントを考慮することで合成音声の品質をどの程度改善できるかについて調査した．そして，波形接続型音声合成を用いた単語音声生成におけるアクセントの有効性を確認した．

以降，2章で波形接続型音声合成を用いた単語音声生成（波形接続型単語音声合成）について説明する．そして，3章で評価実験に関する説明を行い，4章で実験結果を報告する．実験結果については5章で考察する．最後に6章で今後の課題を述べる．

2 波形接続型単語音声合成

2.1 モーラ情報，アクセントと F_0 周波数

波形接続型音声合成を含め，一般に音声合成においては，韻律の扱いが問題となる．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．

一方，音声合成の対象として小さな単位である単語を合成する場合においては，地名などの固有名詞では 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することが可能である [1]．しかし，より一般的な普通名詞では，例えば「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声素片が選択される場合がある．過去の研究において，アクセントを未考慮であったために不自然な音声を作成される場合があったことが報告されている [2]．

図1，図2および図3，図4はそれぞれ，ATR 単語発話データベース Aset に含まれる男性話者 MAU，女性話者 FTK が発話した4モーラ語名詞(約1,600件)における，モーラ位置と F_0 周波数の依存関係を示した図である．横軸は時間(モーラ位置で正規化)，縦軸は F_0 周波数の平均と分散を示している．なお， F_0 周波数の推定には ESPS/waves+[11] を使用した．また，図2および図4における単語のアクセント型分類には，NHK 日本語発音アクセント辞典 [12] を利用した．

図から分かるように，アクセントを考慮しない場合，ほぼ0型と同じ傾向を示しているが，アクセント別に見た場合，特に女性話者 FTK において，1型，2型，3型では0型と異なった傾向を示している．よって，アクセントを考慮することにより， F_0 周波数との依存関係をより有効に利用でき，自然性の向上が期待できる．

そこで本研究では，素片選択においてモーラ情報および単語のアクセントを考慮することで，合成音声の品質をどの程度改善できるかについて調査する．

なお，過去の研究において，モーラ情報は，音素ラベリング [10] や音声認識 [13][14] などの分野において効果があることが報告されている．

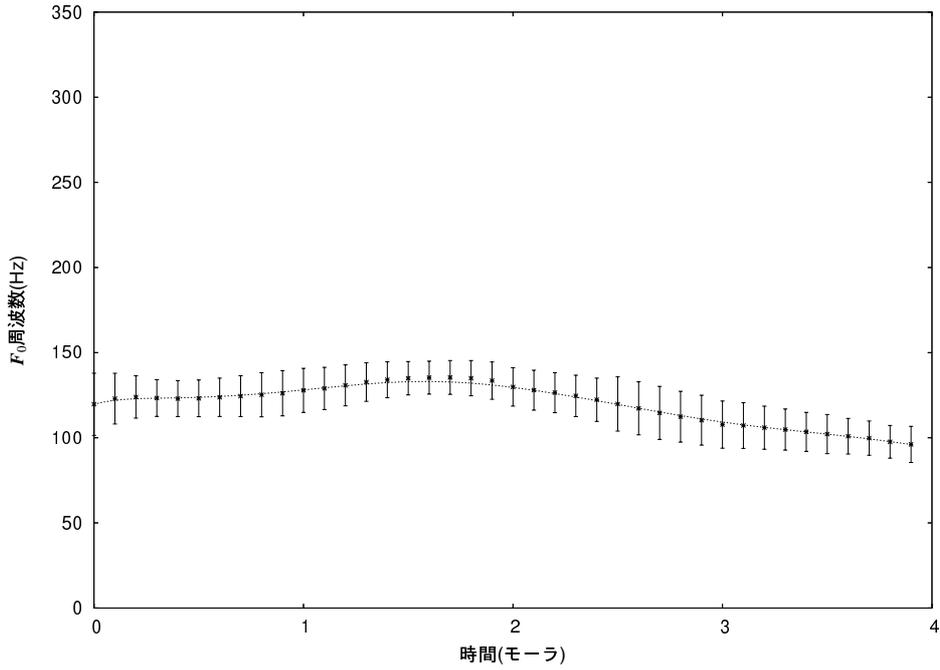


図 1: 話者 MAU における 4 モーラ語名詞 (約 1,600 件) の F_0 周波数分布

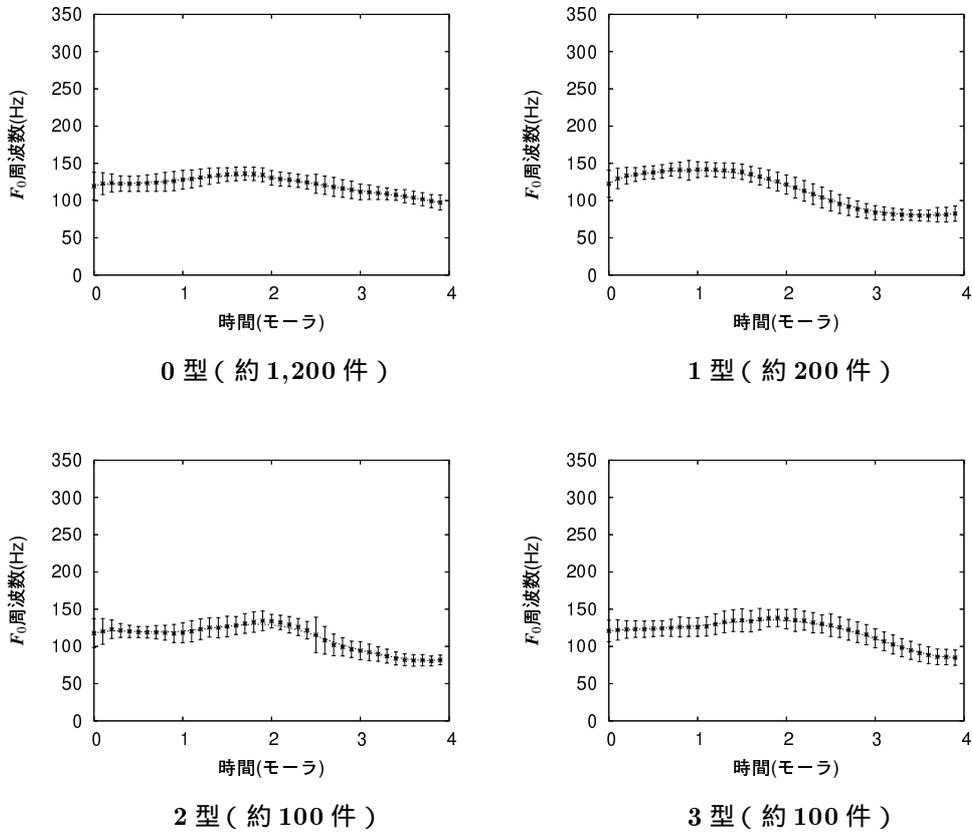


図 2: 話者 MAU における 4 モーラ語名詞の F_0 周波数分布 (アクセント別)

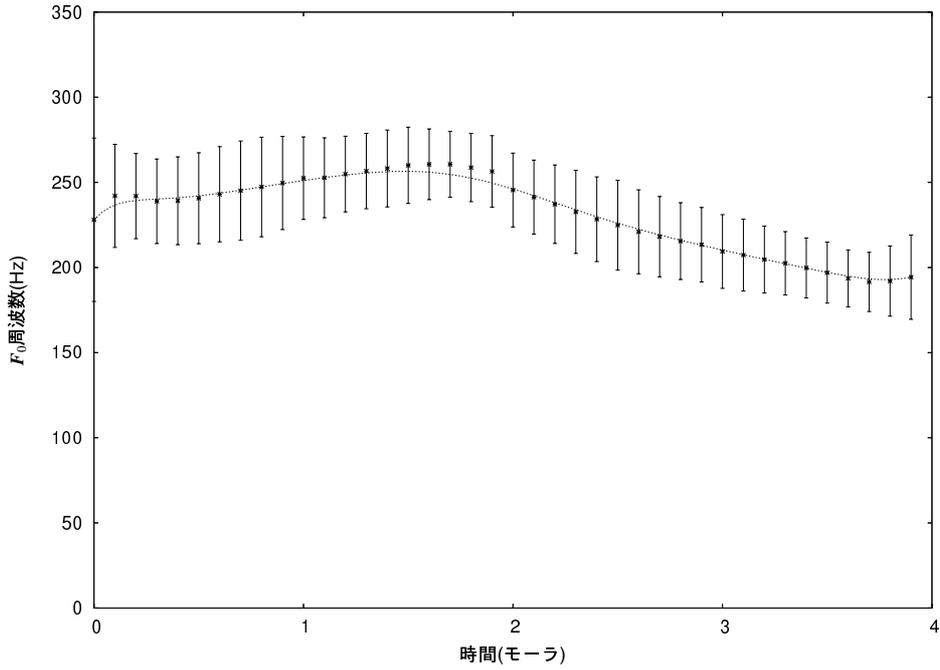


図 3: 話者 FTK における 4 モーラ語名詞 (約 1,600 件) の F_0 周波数分布

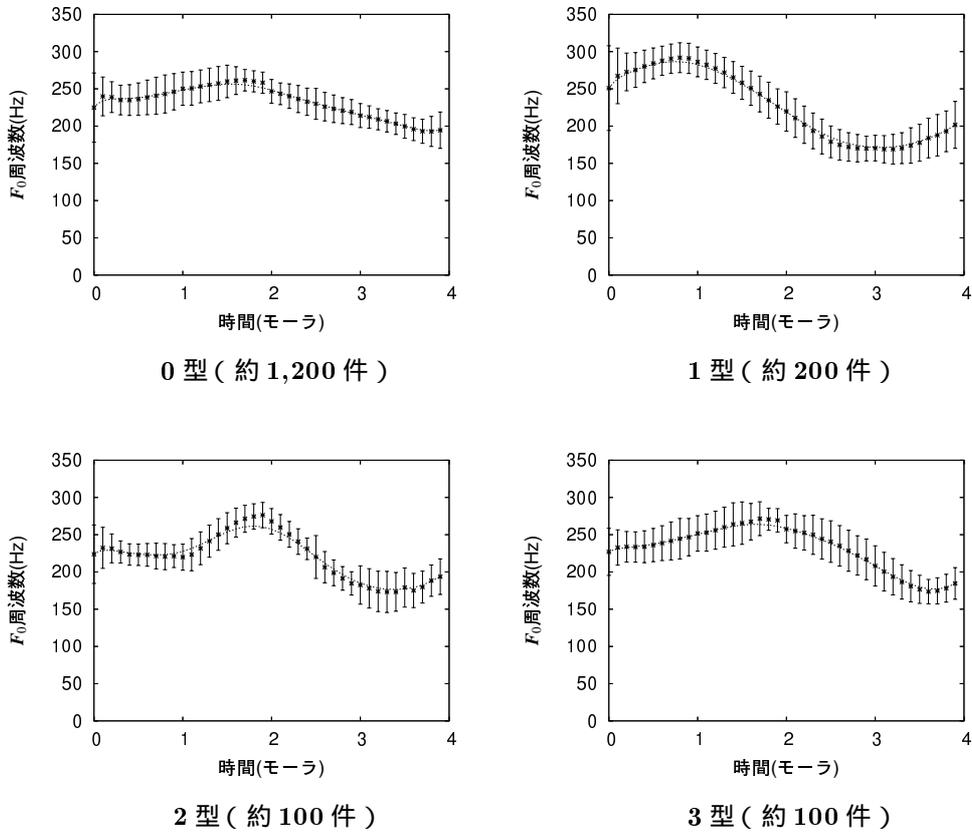


図 4: 話者 FTK における 4 モーラ語名詞の F_0 周波数分布 (アクセント別)

2.2 波形接続型単語音声合成の概説

本研究で用いる波形接続型音声合成では，まず，以下の情報が一致する素片を選択する．単語のアクセントについては，NHK 日本語発音アクセント辞典 [12] を参考にラベルデータに対してアクセントを付加する．

- ・ 音素または音節
- ・ 直前の音素 (前音素環境)
- ・ 直後の音素 (後音素環境)
- ・ 単語中のモーラ位置
- ・ 単語のモーラ数
- ・ 単語のアクセント

そして，素片の開始時間と終了時間を元に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 合成音声の例

本研究で作成した合成音声の例として「伝説」(/de/N/se/tsu/)および「身分」(/mi/bu/N/)について、音素単位、音節単位で合成した場合の例を以下に示す。なお「_」は音の強弱(アクセント)を表している。()内強調部は、実際に選択される部分を示している。また、波形データおよびスペクトログラムについては、図5~図10に示す。波形データおよびスペクトログラムの出力にはWavesurfer[15]を使用した。

伝説 (/d.e/N/s,e/ts,u/) = 伝票 (/d.e/N/py,o/u/)
+ 電報 (/d.e/N/p,o/u/)
+ 幻想 (/g.e/N/s,o/u/)
+ 間接 (/k.a/N/s,e/ts,u/)
+ 密接 (/m.i/q/s,e/ts,u/)
+ 情熱 (/j.o/u/n,e/ts,u/)
+ 忠実 (/ch.u/u/j,i/ts,u/)

伝説 (/de/N/se/tsu/) = 伝票 (/de/N/pyo/u/)
+ 伝染 (/de/N/se/N/)
+ 建設 (/ke/N/se/tsu/)
+ 鋼鉄 (/ko/u/te/tsu/)

身分 (/m,i/b,u/N/) = 身なり (/m,i/n,a/r,i/)
+ 身振り (/m,i/b,u/r,i/)
+ 気分 (/k,i/b,u/N/)
+ 部分 (/b,u/b,u/N/)
+ 処分 (/sh,o/b,u/N/)

身分 (/mi/bu/N/) = 身振り (/mi/bu/ri/)
+ 気分 (/ki/bu/N/)
+ 処分 (/sho/bu/N/)

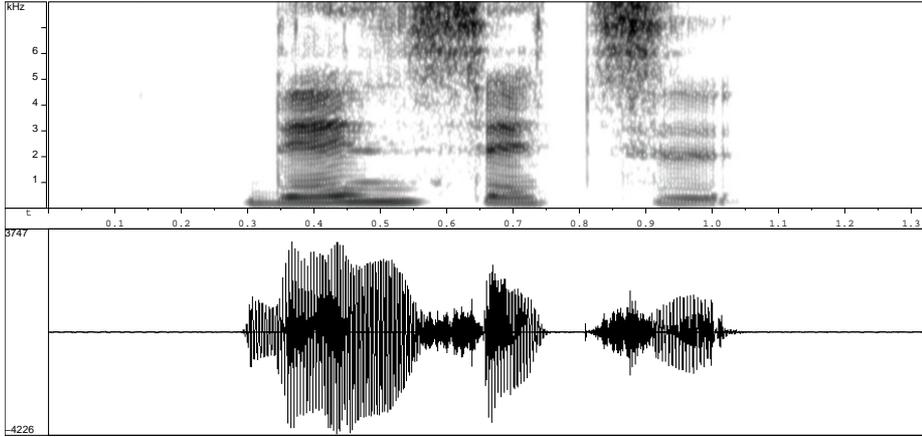


図 5: 「伝説」(自然音声)

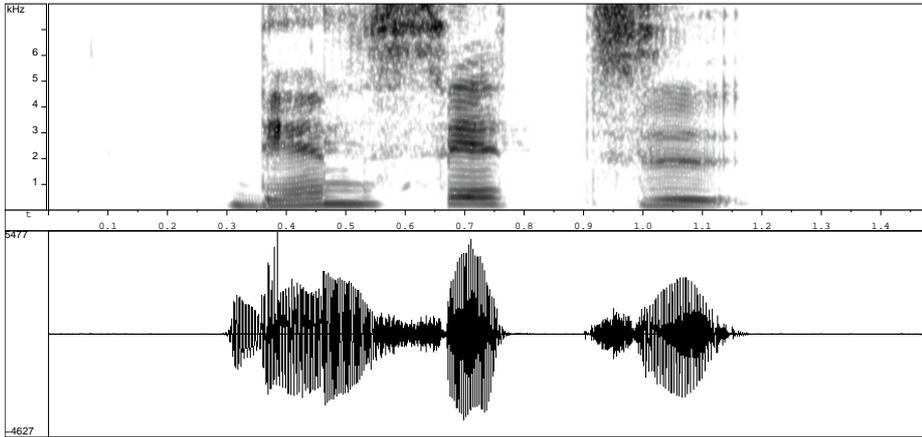


図 6: 「伝説」(音素単位)

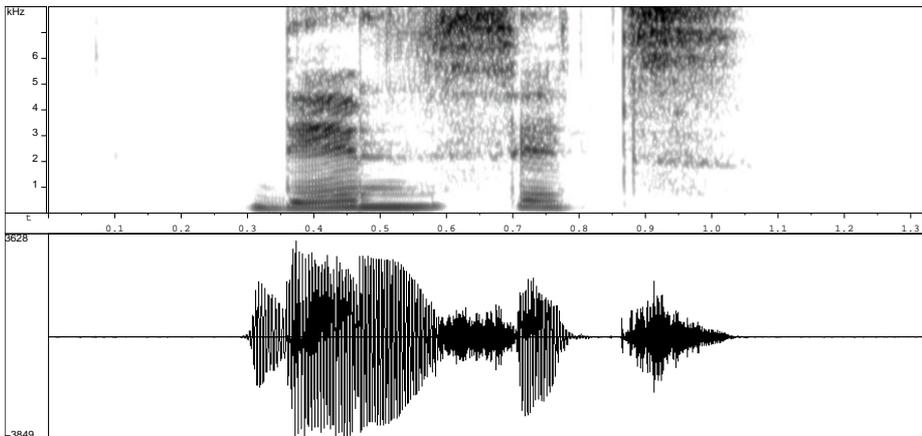


図 7: 「伝説」(音節単位)

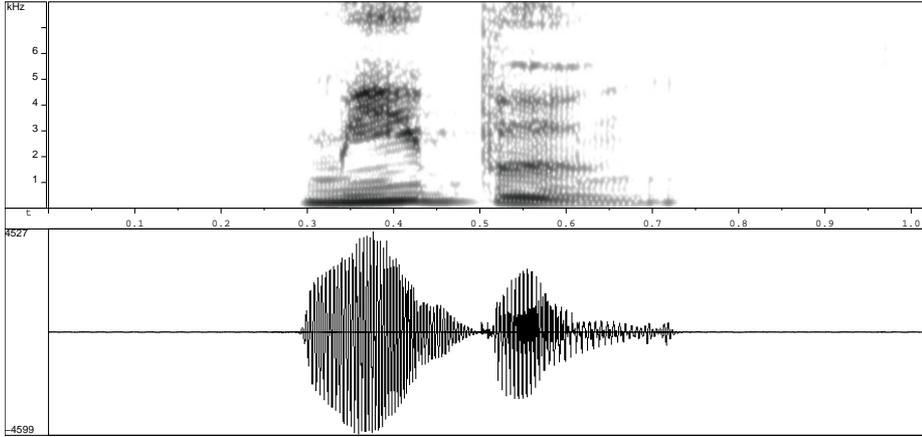


图 8: 「身分」(自然音声)

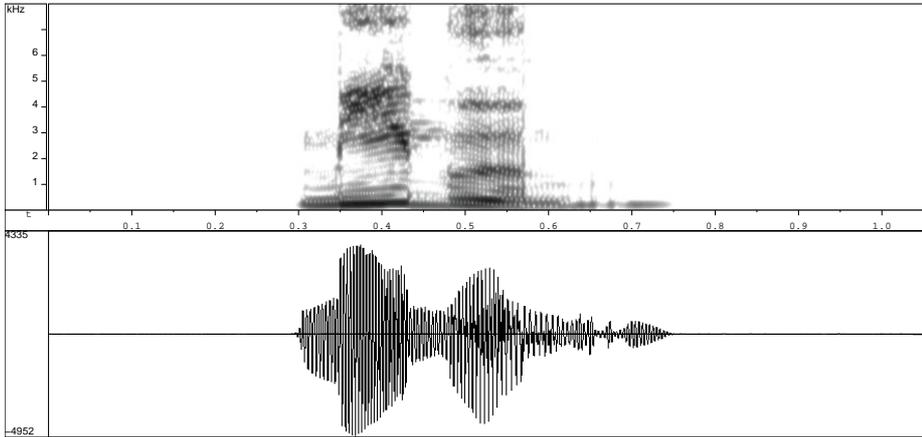


图 9: 「身分」(音素单位)

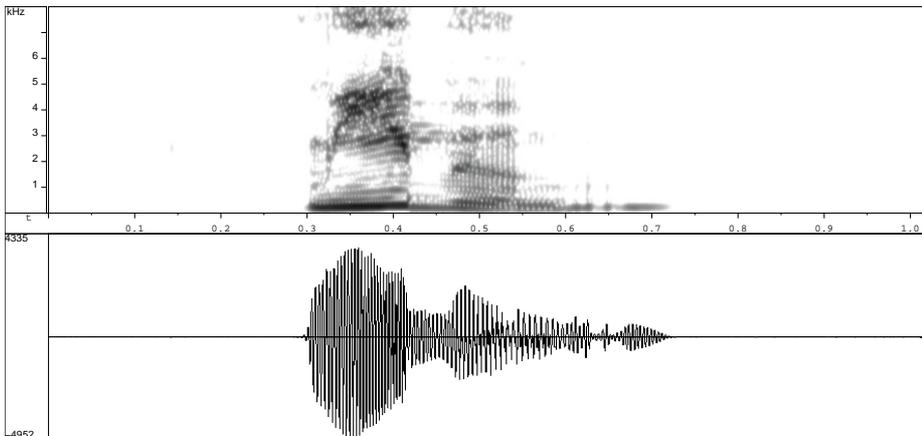


图 10: 「身分」(音节单位)

3 評価実験

3.1 実験環境

アクセントの有効性および最適な素片単位，素片接続位置について調査するために聴覚実験を行う．

まず，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．そして，Asetに含まれる3,4モーラ語名詞について，以下の条件で各100音声(3,4モーラ語各50音声)を準備する．

- ・ 自然音声
- ・ 音節を単位とした波形接続型合成音声 (syl)
- ・ 音素を単位とした波形接続型合成音声 (tri)
- ・ 市販の合成機の合成音声 (市販の音声)

また，波形接続型音声合成による合成音声では，アクセントを未考慮のもの (na) と，考慮したもの (ac) を準備する．市販の合成機による音声では，常にアクセントを考慮する．

話者には， F_0 周波数のばらつきが比較的小さく，収録された音声にエコーの少ない，FTK と FYN の2話者を選ぶ．

3.2 波形接続に関する補則

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，波形の接続位置を音素境界または音素中心とし，さらに，接続部における2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素片開始時間と素片終了時間をもとに，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．音素境界と音素中心のどちらがより適しているかについて，評価実験により明らかにする．

3.3 評価方法

合成音声の評価のために，20代男性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，了解度試験とオピニオン評価を行う．評価は，作成した単語を文に埋め込んで行うのではなく，単語音声のみで行う．

(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，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

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．

本研究では，5段階の評価基準については，以下のように被験者に指示する．

- 5: 自然音声
- 4: 音量正常，アクセント正常，接続部の違和感小
- 3: 音量正常，アクセント違和感小，接続部の違和感小
- 2: 音量違和感小，アクセント違和感中，接続部の違和感中
- 1: 音量違和感中～大，アクセント違和感大，接続部の違和感大

4 実験結果

実験結果を表1, 表2に示す.

表 1: 実験結果 (1): 接続位置 = 音素境界

	了解度 正解率 (%)			オピニオンスコア		
	評価音節数: 1,750			評価単語数: 100		
	FTK	FYN	平均	FTK	FYN	平均
自然音声	99.7	99.9	99.8	4.8	4.9	4.9
syl(na)	99.4	99.7	99.6	3.8	4.1	4.0
syl(ac)	99.4	99.7	99.6	4.2	4.4	4.3
tri(na)	98.2	99.3	98.8	3.4	4.0	3.7
tri(ac)	99.1	99.5	99.3	4.0	4.2	4.1
市販の音声	96.7	97.6	97.2	2.4	2.4	2.4

表 2: 実験結果 (2): 接続位置 = 音素中心

	了解度 正解率 (%)			オピニオンスコア		
	評価音節数: 1,750			評価単語数: 100		
	FTK	FYN	平均	FTK	FYN	平均
自然音声	99.7	99.9	99.8	4.8	4.9	4.9
syl(na)	98.4	99.6	99.0	3.4	3.8	3.6
syl(ac)	99.0	99.7	99.4	3.6	4.0	3.8
tri(na)	98.3	98.6	98.5	3.1	3.6	3.4
tri(ac)	98.2	99.1	98.7	3.4	3.8	3.6
市販の音声	96.7	97.6	97.2	2.4	2.4	2.4

表1, 表2から, 了解度はアクセントの考慮, 未考慮に関わらず, 非常に高い正解率を得た. 一方, オピニオンスコアは, アクセントを考慮することにより, 音節単位の場合で4.0から4.3へ, 音素単位の場合で3.7から4.1へ, それぞれ改善が見られた. また, 素片単位に注目した場合, 音素を単位とした場合より音節を単位とした方が良くなった. 素片接続単位に注目した場合, 音素中心で接続するより音素境界で接続した方が良い結果が得られた. しかし, いずれの条件下でも, 了解度は自然音声と同程度であったが, オピニオンスコアは及ばなかった.

なお, 市販の合成機による合成音声のオピニオンスコアについては, 話者性を考慮する被験者と考慮しない被験者で評価に大きなばらつきがあった.

5 考察

5.1 自然音声と合成音声の差

実験結果から分かる通り，波形接続型音声合成による合成音声は，了解度は自然音声と同等の正解率が得られたが，自然性は依然として低いままであった．アクセント考慮時のオピニオンスコアは音節単位で 4.3，音素単位で 4.1 であり，3.3 節で述べたオピニオン評価の基準によると，音量やアクセントは正常だが，接続部の違和感が残った状態であると言える．

よって，被験者が自然音声と合成音声との差を感じる最も大きな原因は，接続部における違和感であると言える．また，スコアが音素単位より音節単位の方が高いことから，接続部が少ない方が違和感が発生する可能性が低くなり，自然性が向上すると考えられる．

例えば「瓦」という音声では「音節単位，アクセント考慮，音素境界接続」(図 11)の条件下の音声では，被験者 5 人全員が 5 点をつけており，非常に品質の良い合成音声を作成できていた．しかし「音素単位，アクセント考慮，音素境界接続」(図 12)では平均 4.4 点「音節単位，アクセント考慮，音素中心接続」(図 13)では平均 4 点などとなっており，実際の音を聞いてみたところ，接続部にやや違和感の残る音声となっていた．

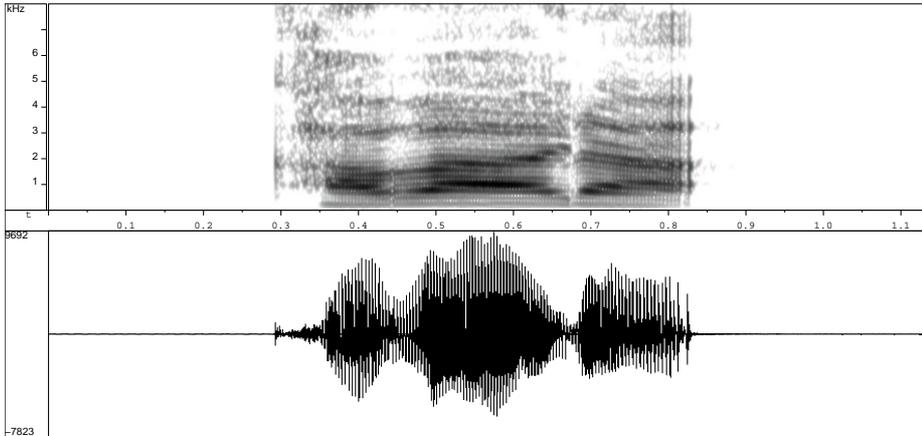


図 11: 「瓦」(音節単位, アクセント考慮, 音素境界接続)

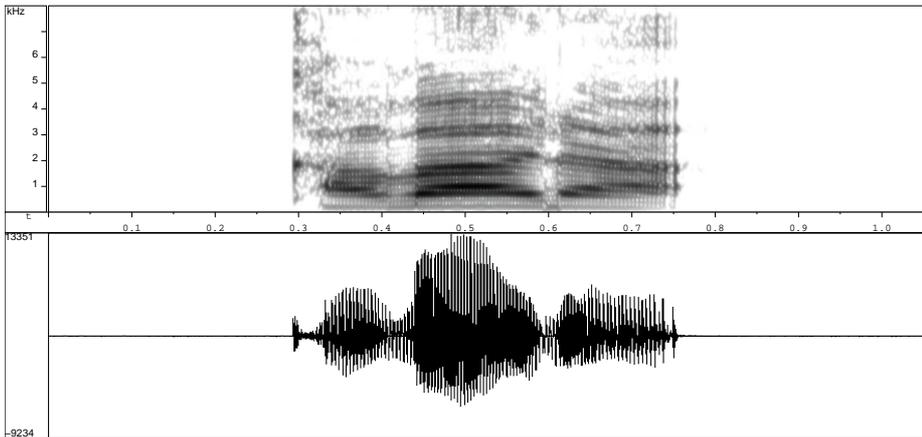


図 12: 「瓦」(音素単位, アクセント考慮, 音素境界接続)

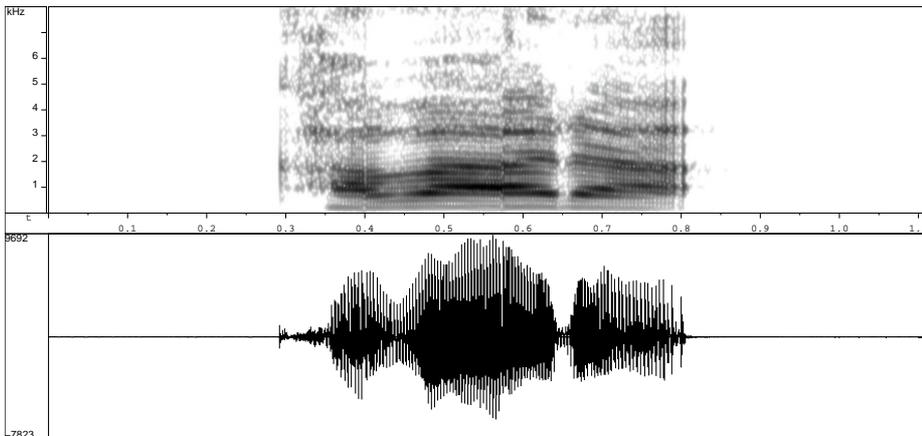


図 13: 「瓦」(音素単位, アクセント考慮, 音素中心接続)

5.2 了解度試験の解析

了解度試験において，波形接続型音声合成で作成した音声で，多くの被験者が間違えた音声について表3に示す．また，了解度試験における間違いを，子音部，母音部に分けて集計した結果(一部)を表4に示す．なお，すべての結果については付録に示す．

表3: 了解度試験における間違いの例

	正解	間違いの例
f1	便利(べんり)	でんり
f2	無断(むだん)	むらん
f3	売店(ばいてん)	ばいて__
f4	来年(らいねん)	らえねん
f5	黒板(こくばん)	ここばん

表4: 了解度試験における間違いの集計結果(一部)

原因	話者	音節単位				音素単位				合計	
		境界		中心		境界		中心			
		ac有	ac無	ac有	ac無	ac有	ac無	ac有	ac無		
最終モーラの撥音の欠落	FTK	1	5	1	3	1	11	1	7	30	
	FYN		1							1	
母音部の誤り	[i] [e]	FTK			6	4		1	8	3	22
		FYN		1	3	3			2	10	19
	[u] [o]	FTK			3	3			3	4	13
		FYN			1				1	4	6
子音部の誤り	[k] [t]	FTK	1			2	1		2		7
		FYN				1	1	1	3		6
	[d] [r]	FTK	3				1	1			5
		FYN					2	1	1	1	5
	[b] [d]	FTK	1	1	2	1	1	2	2	2	12
		FYN									0
	[m] [r]	FTK	1			1	2	3		1	8
		FYN							1		1
	[r] [d]	FTK									0
		FYN			1				3	1	5

表3, 表4から分かるとおり，子音部では例えば f1 および f2 のように，発声方法の似た音を間違える場合が多かった．また，特に FTK における実験では，f3 のように最終モーラの撥音を聴き逃す被験者が多かった．これは，FTK の発話速度が FYN に比べ速かったためであると思われる．母音では，音素中心で接続した場合，f4 および f5 に示すように「い」と「え」，「う」と「お」を間違える場合が多く見られた．

5.3 オピニオン評価の解析

FTK, FYN 共に評価の悪かった音声のうち、接続部の違和感やアクセントの不自然さ以外の原因としては、特に、音量に違和感のある音声挙げられる。これは、音声コーパスの録音音声の音量にばらつきがあることが影響していた。例えば「会話」という音声では、「対話」と「内輪」の音量差により、接続時に第2, 3モーラ間において音量が極端に変わり、違和感が発生した「会話」の波形とスペクトログラムを図14および図15に示す。しかしこの問題は、音声コーパスの作成時にあらかじめ音量を揃えておくことで解決可能であると考えられる。

また、各音素・音節の継続時間長の問題により、違和感を感じる場合があった。例えば「結局」という音声において、第2モーラの促音の継続時間が極端に短い音声を作成されてしまった。この問題は、3.2節で述べた、波形接続時の位相の調整による弊害であった。促音部では、話者の発声方法によって振幅値がゼロ（つまり無音）となる場合があり、計算機によって抽出位置を自動的に調整する際、負から正になる位置を見つけられず、継続時間長が極端に短く調整されてしまう場合があった。波形データとスペクトログラムについて、自然音声の場合を図16に、合成音声の場合を図17に示す。本研究では継続時間長の制御は行わなかったため、このような問題に対処することができなかった。今後、さらなる自然性の向上のためには、継続時間長の制御が重要であると思われる。また、位相の調整について、促音では処理を行わないなどの対処も必要である。

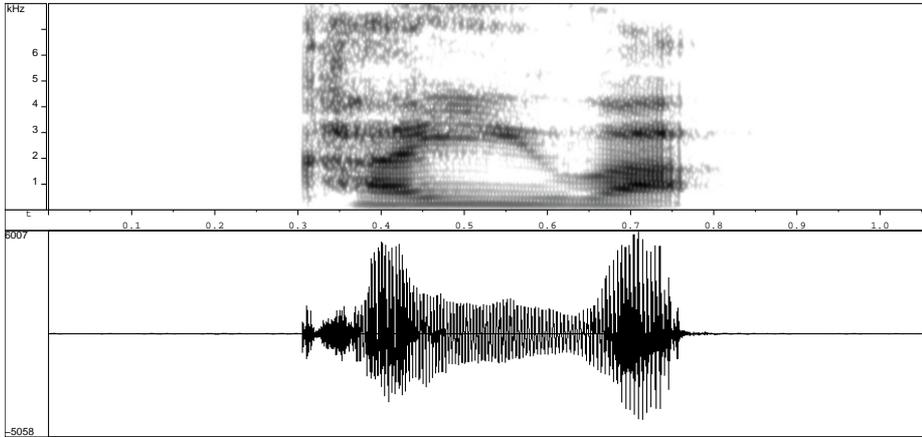


図 14: 「会話」(自然音声)

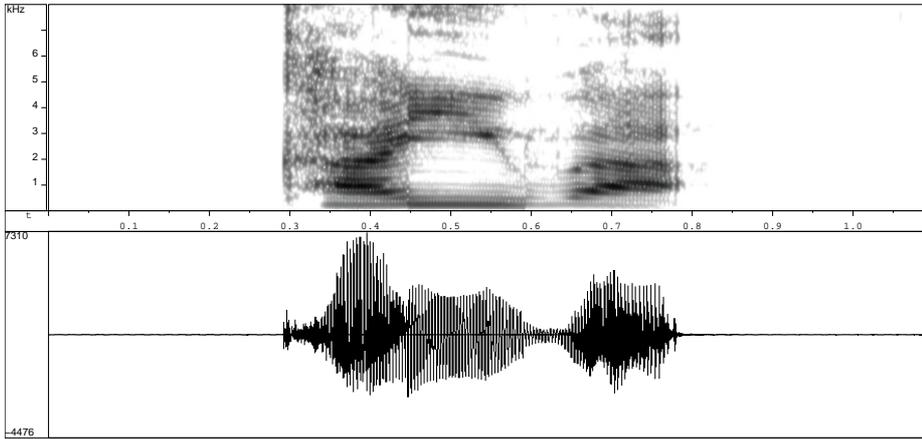


図 15: 「会話」(合成音声)

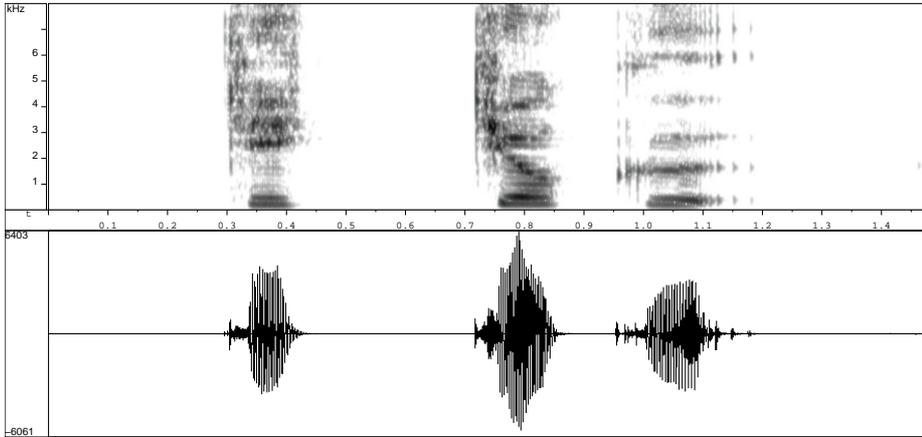


図 16: 「結局」(自然音声)

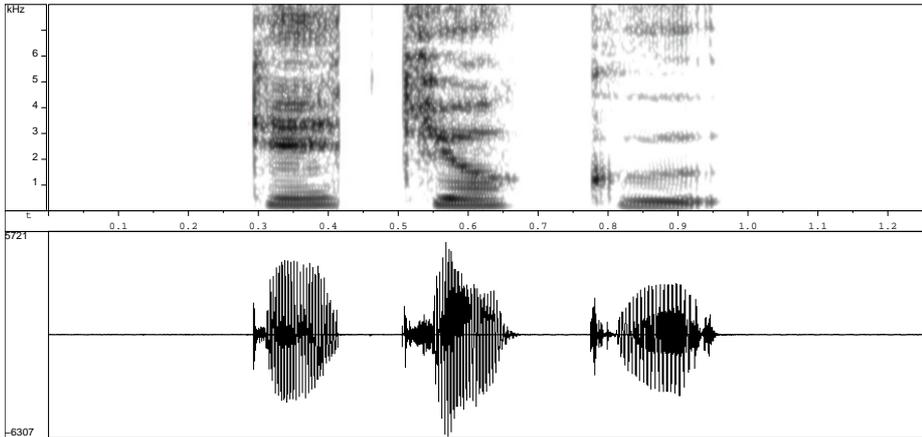


図 17: 「結局」(合成音声)

5.4 素片単位について

実験結果から，音素単位より音節単位の方が評価が高くなることが分かった．しかし，音節を単位とした場合，前後音素環境を考慮した場合の素片種類数は，一般に音素を基本単位とした場合より多くなる．そのため，音声コーパスの大きさが決定されている場合，作成できる音声は少なくなってしまう．今回の実験で作成できた音声は，音素単位で約 650 単語，音節単位で約 470 単語であり，約 180 個もの差があった．

この問題に対しては，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思われる．なお，環境などは異なるが，単語音声より短い刺激音声について，代替による自然性劣化に関する評価がなされている [16]．

5.5 素片接続位置について

素片の接続については，本研究では音素中心より音素境界で行った方が良い結果が得られた．しかし，特に母音連続の場合において，接続歪みの頻度分布は音素中心と音素境界ではほぼ同じであると報告されている例もある [7]．実験条件等は異なるが，今後さらなる検討が必要であると思われる．

6 まとめ

本研究では、波形接続型単語音声合成におけるアクセントの有効性を検証した。聴覚実験における合成音声の単語理解度は、音素単位の場合で 99.3 %，音節単位の場合で 99.6 % が得られた。また、オピニオンスコアはそれぞれ 4.1，4.3 であった。アクセントを考慮することで、オピニオンスコアはそれぞれ 0.4，0.3 の向上となり、自然性の向上のためにアクセントが有効であることが分かった。

一方、自然音声の単語理解度は 99.8 %，オピニオンスコアは 4.9 であった。アクセントの考慮によってオピニオンスコアは上昇したが、自然音声には及ばなかった。

今後の研究課題としては、さらなる自然性の向上を目指して、様々な制御を導入すると同時に、接続位置に関するさらなる調査、音声コーパスをより有効に利用できるように、特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と思われる。

参考文献

- [1] 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).
- [2] 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).
- [3] 益子貴史, 徳田恵一, 小林隆夫, 今井聖: 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).
- [4] 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).
- [5] 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).
- [6] 阿部匡伸: 音声合成のための合成単位の基礎, 電子情報通信学会技術研究報告, SP2000-73, pp. 35–42 (2000).
- [7] 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オ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).
- [8] 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).
- [9] 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).
- [10] 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).
- [11] ESPS/waves+ with Ensig, Entropic Research Laboratory, Inc.
- [12] NHK 出版: NHK 日本語発音アクセント辞典 新版, ISBN 4-14-011112-7 (1998).
- [13] 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).

- [14] 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).
- [15] Kåre Sjölander and Jonas Beskow: Wavesurfer,
<http://www.speech.kth.se/wavesurfer/>
- [16] 河井恒, 津崎実, 舩田剛, 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).