

アクセント型を考慮したモーラ単位選択の検討*

◎藤井 慶 (NAIST) ニック・キャンベル (NAIST/ATR/CREST)

1 はじめに

本研究では、波形接続型音声合成の単位選択においてアクセントの正しい音声単位列をより効率よく得る手法について検討する。

日本語アクセントは語意を理解する上で重要な要素であり、主に F_0 の相対的な高低変化により実現される。しかし従来の波形接続型音声合成システムではアクセントは F_0 値の系列の中に表されており、 F_0 の絶対的な高さや大局的な高低変化といった他の要素と分離されていない。そのため単位選択においてアクセント型の実現の重要度という観点からの重み付けが不可能であり、更に、特に十分に大きな音声コーパスを用いない場合、指定したアクセント型と一致する音声単位の組合せがあるにもかかわらず型の異なる音声単位列が最適な組合せとして選ばれる恐れがある。例えば図1では3型アクセントの語が合成ターゲットになっているが、2型アクセントの単位列が F_0 ターゲットコスト最小になる。

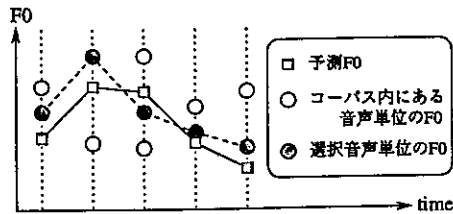


図1: 単位選択部でアクセント型が変化する例

そこで本研究ではアクセント型の判別 [1] で提案された F_{0ratio} の概念を基に、アクセント型を考慮した単位選択法を検討する。

2 F_0 の高低変化の表現

石井らはアクセント型判別の研究 [1] において、 i 番目のモーラと $i-1$ 番目のモーラの高低の変化を次式で定量化している。

$$F_{0ratio}(i) = 12 \log_2 \frac{F_{0mora}(i)}{F_{0mora}(i-1)} \quad (1)$$

モーラ間で F_0 が低 → 高と変化するとき F_{0ratio} は正の数を取り、高 → 低では負の数をとる。 $F_{0mora}(i)$ は i 番目モーラの F_0 を代表する値である。石井ら

*A study on accent-based mora unit-selection. By Kei Fujii(Nara Institute of Science and Technology(NAIST)) and Nick Campbell(NAIST/ATR/CREST)

は F_{0mora} として各モーラ内の F_0 の平均値や線形回帰直線により求めたターゲット値を検討している。

本研究ではこの F_{0ratio} の概念を単位選択部に取り入れる。ただし F_{0mora} には各モーラの F_0 の中央値を用いた。中央値を用いることで F_0 抽出失敗による倍ピッチの影響低減が期待される。

3 言語解析～韻律予測

従来テキスト音声合成では、まず言語解析部で入力テキストを解析し、読みやアクセント型、ポーズ位置などを決定する。その後 韻律予測部で F_0 値系列や各音声単位の時間長などを求める。言語解析部で得られる韻律特徴は質的に表現され、韻律予測部で得られる韻律特徴は量的に表現されており、韻律予測部において質的な特徴を量的な特徴にマッピングしていると考えられる。

本研究では韻律予測部においてまず音声単位毎に時間長を求め、次に言語処理部で得たアクセント型を基に各モーラ間の F_0 高低の関係を求める。高低の関係は“/”、“\”、“-”の3種で表す。“/”は低 → 高、“\”は高 → 低、“-”は高(低) → 高(低)を意味する。例えば3型アクセントである「あらゆる」の各モーラ間の高低関係は

$$a / ra - yu \setminus ru$$

と表される。以降、説明の便宜上 $i-1$ 番目と i 番目モーラ間の高低の関係を $T_{acc}(i)$ と表す。 $T_{acc}(i)$ はアクセント型のターゲットとして次項で述べるコスト計算に用いられる。

4 アクセントの選択コスト

本研究での単位選択におけるコスト計算の概要を図2に示す。

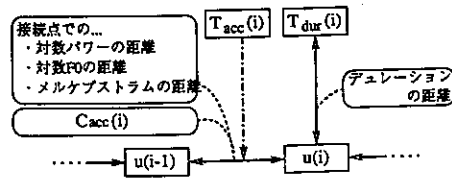


図2: コスト計算

図中の $u(i)$ は i 番目音声単位の候補、 $T_{dur}(i)$ は i 番目音声単位の予測時間長である。 $C_{acc}(i)$ はアクセントのコストであり、 $T_{acc}(i)$ と $F_{0ratio}(i)$ より以下のように求める。

$T_{acc}(i) = "/>$

$T_{acc}(i) = "\backslash" のとき \dots$
 $F_{0ratio}(i) < S_{fall}$ なら $C_{acc}(i) = 0$
 そうでなければ $C_{acc}(i) = F_{0ratio}(i) - S_{fall}$

$T_{acc}(i) = "-" のとき \dots$
 $C_{acc}(i) = |F_{0ratio}(i)|$

ここで S_{rise} , S_{fall} はしきい値であり、それぞれ $S_{rise} \geq 0$, $S_{fall} \leq 0$ を満たすよう設定する。図が示すように、提案手法でのアクセントのコスト計算は接続コスト計算と同様の形式になっている。

$C_{acc}(i)$ 以外のコストとして予測時間長との距離、接続する単位同士の接続点での対数 F_0 の距離、対数パワーの距離、メルケプストラムの距離を求め、各コストの重み付き総和が最小になる単位列を選択する。

5 実験

前項で述べた単位選択を用いて合成した1アクセント句の合成音を聞き、アクセントの型の判別を行う実験を行った。

5.1 実験条件

音声コーパスには女性話者1名の朗読音声(16 kHz, 16 bit)を用いた。全体のモーラ数は15832である。なお、このコーパスは韻律バランスを考慮して設計されたものではなく、従来法[2]においては予測韻律に対し十分に忠実な合成音は得られないものである。

2モーラから4モーラの長さを持つ名字17個の後ろに「サン」を付けたテキストについて、名字のアクセント型が0型から最大4型になるよう単位選択を行い合成音を作成した。例えば「フジイ」の場合「フジイサン」を合成テキストとし、0型から3型までの合成音を作った。「サン」を付けることで0型と3型の区別が可能になる。作成した合成音の内訳を表1に示す。

表1: 実験試料の内訳(計75文)

アクセント型	試料数
0	17
1	17
2	17
3	15
4	9

また前述のとおり、今回用いたコーパスは小規模なものであるため、提案手法が従来法に比べてどの程度改善が得られるかという観点から性能を知るための指標として従来の波形接続型音声合成システム[2]を用いて同コーパスで同様に合成音75文を作成した。

被験者は成人男女6名であり、各人に対しランダムに並び替えた150文をヘッドホンにより受聴し、アクセント型を判別してもらった。なお、判別の際合成音を何度も聞き直すことを許した。

5.2 実験結果

合成音の型判別の結果を表2,3に示す。各欄の値は合成時に指定した型ごとの回答の割合である。

表2: 提案手法による合成音の判別結果

回答の型	指定した型				
	0	1	2	3	4
0	97.059	7.843	17.647	34.444	31.481
1	0.000	74.510	1.961	4.444	1.852
2	1.961	7.843	62.745	24.444	9.259
3	0.980	7.843	8.824	23.333	12.963
4	0.000	1.961	8.824	13.333	44.444

表3: 従来法による合成音の判別結果

回答の型	指定した型				
	0	1	2	3	4
0	5.882	10.784	2.941	2.222	0.000
1	17.647	46.078	20.588	18.889	22.222
2	33.333	19.608	36.275	32.222	16.667
3	34.314	17.647	35.294	36.667	40.741
4	8.824	5.882	4.902	10.000	20.370

表より、提案手法は3型以外の型で従来法より良い正答率を得た。特にアクセント核が前にある語での改善が顕著であった。アクセント核が後の方にある場合については、0型(アクセント核がない)と誤判断された比率が高い。つまりアクセント核での F_0 の下降が十分でなかったと考えられる。そのため、アクセント核が後の方にあるほど S_{fall} をより低い値にするとといった改良が必要であると考えられる。

6 まとめ

本稿では、波形接続型音声合成においてアクセント型の正しい音声単位列をより効率よく選択する手法について検討した。実験では従来法に比べ3型以外の型で良い判別率を得た。今後の課題として、しきい値 S_{rise} , S_{fall} をアクセント核の位置により変化させること、 F_0 のデクリネーションも取り込んだ形で単位選択することなどが挙げられる。

謝辞 本研究の一部は科学技術振興事業団戦略的基礎研究推進事業(JST/CREST)の援助により行われた。

参考文献

- [1] 石井カルロス寿意, 峯松信明, 広瀬啓吉: "ピッチ知覚を考慮した日本語連続音声のアクセント型判定", 信学技報, SP2001-48, pp.23-30 (Aug.2001)
- [2] ニック・キャンベル, アラン・ブラック: "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, pp.45-52 (May.1996)