波形接続型音声合成における位相情報を利用した音節境界位置の決定方法*

◎橋本浩志，村上仁一，池原悟 (鳥取大)

1　はじめに

音声合成の手法の 1 つとして音節波形接続型音声合成 [1] が提案されている．この手法は録音音声から音声波形を音節単位で分割し，接続することで合成音声を作成する手法である．信号処理を用いずに言語的な情報のみを用いて音節素片を選択するという特徴をもつ．

音声波形の音節開始位置・終了位置は，ATR から提供されている音節境界位置が記載されたラベル (以下，ATR ラベル) を利用している．しかし，ATR ラベルは人手で作成されているが，音節境界位置の精度が低いため，音声合成時に波形の接続点が不連続になり，音声品質が劣化する．そこで波形が滑らかに接続するように波形接続時に人手で音節開始位置・終了位置に修正を加えている．しかし，この修正作業にはコストがかかる [2]．

そこで，本研究では音節の精密な開始位置・終了位置を自動的に決める方法を提案する．具体的には，音節素片のパワーが最大となる周波数を求め，その初期位相が “-$\frac{\pi}{2}$” となる時間を音節開始位置にする．そして，提案方法を用いて合成音声を作成し，音声品質を調査する．

2　音節波形接続型音声合成

2.1　音節波形接続方式における音声合成方法

音節波形接続型音声合成における具体的な音声合成方法を以下に示す．

1. 作成したい単語の音節素片選択条件 (Table1 に示す．) を求める．
2. 1. で指定した条件と一致する音節素片を，ATR の単語発話データベース中から選択する．
3. 選択した音節素片の開始位置・終了位置を ATR ラベルに記載されている音節境界位置から得る．
4. 接続部が滑らかに繋がるように音節開始位置・終了位置を人手で修正する．
5. 修正した音節開始位置・終了位置を用いて音節素片を切り出し，音声合成を行う．

Table 1　音節素片の選択条件

1. 中心の音節
2. 直前の音素 (前音素環境)
3. 直後の音素 (後音素環境)
4. 単語のモーラ数
5. 単語のモーラ位置
6. 単語のアクセント型
7. 単語のアクセントの高低

「乗り物 (no/ri/mo/no)」という合成音声を作成する際の例を Fig. 1 に示す．Fig. 1 の「‾|_|¯」はアクセントの高低を表す．また，太文字で示している部分は，接続する音節素片である．

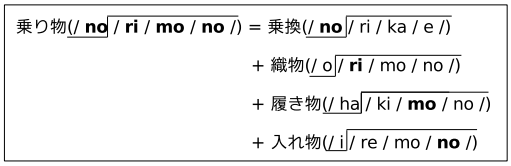

Fig. 1　音節波形接続型音声合成の音声合成例

2.2　音節境界位置の問題点

音節波形接続型音声合成では，ATR ラベルから得られる音節境界位置を利用して各音節の開始位置・終了位置を決めている．しかし，ATR ラベルは人手で作成されているが，波形接続型音声用に正確に記載されていない．そのため，音声合成時に接続点に不連続が生じる．そこで音節波形接続型音声合成では，ATR ラベルから得られる音節境界位置を元にして合成時に接続部が滑らかに繋がるように音節開始位置・終了位置を人手で修正している．

ATR ラベルを用いた合成例を Fig. 2 に示す．Fig. 2 は合成音声「形見 /ka/ta/mi/」の “ta/mi” の音声波形である．また，図中の縦線部は “ta” と “mi” の接続部を示している．接続部に注目すると，滑らかに接続されていないことがわかる．

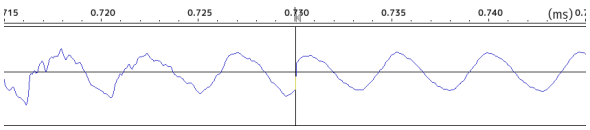

Fig. 2　ATR ラベルの音節境界位置を用いて「形見 /ka/ta/mi/」を音声合成した例

2.3　位相を用いた音節開始位置変更方法

2.2 節の問題点を解決するために，本研究では音節開始位置を，音節素片のパワーが最大となる周波数の初期位相が “-$\frac{\pi}{2}$” となる位置に決める．具体的な方法を以下に示す．

1. ATR のラベルから得られる音節境界位置を用いて音節開始位置・終了位置を得る．
2. 音節素片に対して FFT(フーリエ変換) を行う．
3. フーリエ変換によって得られるパワーが最大時の周波数 F_{max} における初期位相 P_0 を求める．
4. 初期位相 P_0 を “-$\frac{\pi}{2}$” にするように音節開始位置を変更する．変更する時間幅 T は $T = \frac{\frac{1}{F_{max}} * P_0}{2\pi}$ になる．なお，初期位相 “-$\frac{\pi}{2}$” では音声波形の振幅が-から+に変わる点である．また，位相 P_0 は連続値であるが，サンプリング周波数 16kHz の音声を用いるため，音節開始位置の変更値は約 0.0625 ms 毎の離散値になる．

2.4　音節終了位置変更方法

音節の終了位置は，次の音節の開始位置を 2.3 節の手順 1. から手順 4. を行って変更する．

---

*Syllable Boundary Position using Phase for Corpus Based Speech Synthesis. by HASHIMOTO Hiroshi, MURAKAMI Jin'ichi and IKEHARA Satoru(Tottori University)

2.5 FFT の窓長

音節の基本周期は開始時と終了時で異なる．そのため，音節の開始位置から終了位置までの窓長でフーリエ変換を行うと，パワー最大の周波数の初期位相に誤差が生じ，正確な音節開始位置を得ることができない．そのため，本実験では FFT の窓長は音節の開始位置から 64 ポイント (約 4 ms) とする．

2.6 音節開始位置・終了位置変更後の波形例

上記の方法を用いて音節開始位置・終了位置の変更を行い作成した合成音声の音声波形例を Fig. 3 に示す．Fig. 3 は合成音声「形見 /ka/ta/mi/」の “ta/mi” 間の接続部の音声波形を示す．図中の縦線部は接続部を示している．接続部に注目すると，滑らかに接続されていることが分かる．

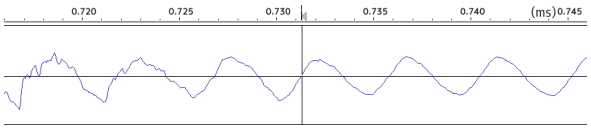

Fig. 3 提案方法を用いて「形見 /ka/ta/mi/」を音声合成した例

3 聴覚実験

3.1 実験データ

本実験において音声合成に用いる音節素片は ATR 単語発話 DB Aset(5,240 件) 内の単語を使用する．話者は女性話者 1 名 (FYN) とする．聴覚実験には，以下の 3 種類の合成音声を用いる．

- 提案方法を用いて作成した合成音声 (以後，提案合成音声).
- ATR ラベルから得られる音節境界位置を用いて作成した合成音声 (以後，ATR 合成音声).
- 波形を滑らかに接続するように人手で作成した合成音声 [3] (以後，人手合成音声).

なお，接続部に注目して音声品質の違いを調査するため，3 種類の合成音声にはすべて同じ単語を使用する．

3.2 評価方法

合成音声の音声品質を評価するために，音声研究に関わった経験のない 5 名を対象に，オピニオン評価実験及び対比較実験を行う．評価単語数は各 100 単語とする．また，オピニオン評価は自然音声を含めて行う．

4 実験結果

4.1 オピニオン評価実験

提案合成音声の音声品質を調査するために，オピニオン評価実験を行う．結果を Table2 に示す．

4.2 対比較実験

提案合成音声と従来合成音声の音声品質を比較するために，次の 2 種類の対比較実験を行う．(1) 提案合成音声と ATR 合成音声，(2) 提案合成音声と人手合成音声．結果を Table3 に示す．

4.3 実験結果のまとめ

聴覚実験の結果，提案合成音声は従来合成音声より高い音声品質が得られた．また，提案合成音声と人手合成音声の音声品質にはほとんど差がなかった．

Table 2 オピニオン評価実験の結果 (各評価単語数 : 100)

音声の種類	オピニオンスコア
提案合成音声	3.83
人手合成音声	3.93
ATR 合成音声	3.31
自然音声	4.71

Table 3 対比較実験の結果 (各評価単語対 : 100)

比較対象 1	比較対象 2
提案合成音声 : 78.4%	ATR 合成音声 : 21.6%
提案合成音声 : 45.2%	人手合成音声 : 54.8%

5 考察

提案合成音声の音声品質が劣化した音声波形を調査した．Fig. 4 に例を示す．Fig. 4 は「頑固/ga/N/ko/」の “ga/N” の提案合成音声である．Fig. 4 の縦線部より，音節素片が滑らかに接続されていないことがわかる．原因を以下に示す．

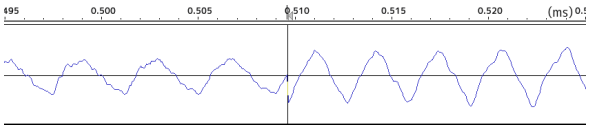

Fig. 4 音声品質が劣化した音声波形 (合成音声「頑固 /ga/N/ko/」) の例

本実験では，音節の開始位置から 64 ポイント (約 4 ms) の窓長でフーリエ変換を行った．この場合，周波数分解能は 250 Hz になる．周波数分解能が低いため，パワー最大の周波数の初期位相に誤差が生じ，正確な音節開始位置を得ることができない．そのため，音声品質が劣化した合成音声が作成されると考えられる．

6 おわりに

本研究では，音節開始位置・終了位置を波形接続方式用に正確に決める方法を提案した．その結果，従来より高い音声品質を得ることができた．

しかし，フーリエ変換の窓長が 64 ポイントでは，周波数分解能が低い．そのため，パワー最大の周波数に誤差が生じ，音節開始位置を変更する時間幅に誤差が生じる．この問題点を修正することで，今回の実験結果より高い音声品質を得ることができると考えている．

謝辞 本論文を執筆するにあたり，聴覚実験に協力してくださった研究室の方々に深く感謝いたします．

参考文献

[1] 村上他: “音節波形接続方式による単語音声合成”, 信学論 D-II, Vol.J85-D-II, No.7, pp.1157-1165, 2002.

[2] 居村他: “波形接続型音声合成のフレーズへの適用”, 言語処理学会第 14 回年次大会, pp.965-968, 2008.

[3] 石田: “アクセントを考慮した波形接続型単語音声合成”, 鳥取大学大学院工学研究科修士論文, 2004.