

周波数領域における残差信号解析による高品質音声合成

5V-9

太田垣浩之 清水忠昭 菅田一博 井須尚紀

鳥取大学 工学部 知能情報工学科

1. はじめに

線形予測分析による音声分析合成系は、音声生成過程を音源部と調音部に完全分離したモデルを仮定している。線形予測分析は、音源部と調音部に関する特徴パラメータにより、少ない情報量での音声符号化が可能であるが、合成音声の品質に限界がある。合成音声を高品質化するには、大きく分けて、調音部の改良と音源部の改良の2方向が考えられる。本研究では、残差信号とインパルス列の短時間スペクトル分析による新しい音源を提案し、合成音声の高品質化を図る。

2. 音源生成手法

2.1 従来の音源生成手法

音声分析合成系における音源として、残差信号をそのまま利用すれば、原音声と区別つかない高品質の合成音声を得られるが、情報圧縮にならない。

残差信号は、音源パラメータを用いて、有声音区間に対してはインパルス列(以下、パルス列)、無声音区間においては白色ランダム雑音で近似するのが一般的である。音源パラメータは、以下の3つである。

1. 有声音/無声音の区別
2. 基本周期(ピッチ)
3. 音源振幅

音源パラメータによる音源は、大幅な情報圧縮が可能となるが、残差信号を極端に簡素化するために、合成音声の品質の劣化が避けられない。

High Quality Speech Synthesis by the Residual Signal Analysis in the Frequency Domain
Hiroyuki Otagaki, Tadaaki Shimizu,
Kazuhiro Sugata, Naoki Isu
Department of Information and Knowledge, Faculty of Engineering, Tottori University

2.2 本研究で考案した音源生成手法

パルス列は、残差信号の高周波成分の特徴を全くとらえていない。本研究では、パルス列に残差信号の高域の周波数成分を付加した音源を作成することを考案した。

音源生成手法を図1に示す。時間領域、周波数領域の変換には、FFT, IFFTを用いる。短時間スペクトル分析により、周波数領域で、パルス列スペクトルに残差スペクトルの高域部を付加し、合成スペクトルを作成する。合成スペクトルを、時間領域に変換し、新しい音源とする。

本研究で考案した音源は、残差情報を付加するため、情報量の増加が避けられない。そこで、残差情報を付加する際に、時間領域または周波数領域における情報圧縮を試みた。

時間領域では、TDHS アルゴリズム¹⁾を用いて、残差信号を平均ピッチ毎に半分に圧縮することにより、音源パラメータの総量を半分に圧縮した。周波数領域では、残差スペクトルの高域成分である実部、虚部それぞれを、適応差分符号化した。

3. 合成音声品質評価

合成音声は、分析次数12次のLSP分析により作成した。サンプリングは、10 kHz, 12bits, LSP分析・短時間スペクトル分析は、ハミング窓、フレーム長25.6ms, フレーム更新周期20.0msとした。従来の音源パラメータは、変形相関法により求めた。付加する新しい音源パラメータは、残差スペクトルの全周波数成分の約1/4である高周波数成分(3867~5000Hz)とした。

合成音声品質を検討するために、従来の音源パラメータから作成した合成音声と本研究で考案した音源パラ

メータから作成した合成音声の対比較試験を行った。音声資料は、男性4名、女性7名の単母音、有声音のみの連続音声、無声音混入の連続音声とした。対比較試験は、27名の被験者を対象に行った。

従来の音源による合成音声の総得点を1とした時の本研究で考案した音源による合成音声の総得点を図2に示す。図2より、本研究で考案した音源は、高品質音声合成に非常に有効であるといえる。本研究で考案した音源は、連続音声に対して高得点であったことから、わたり部、無声音区間にも有効であるといえる。

本研究で考案した音源は、女声よりピッチ抽出の精度が良い男声に有効であった。

情報圧縮した音源パラメータによる合成音声も、従来の音源による合成音声より良好な結果が得られた。

4. おわりに

従来の音源は、残差信号の大局的な特徴はとらえているが、細かい特徴は全くとらえていない。本研究では、周波数領域で、従来の音源に残差信号の高周波数成分を付加することにより、残差信号の時間領域における細かい特徴を与えた。

実験結果より、残差信号の高周波数成分は、合成音声の高品質化ならびに子音合成に有効であることがわかった。

今後、人間の聴覚特性を考慮した残差信号の高周波成分の抽出・符号化法ならびに高精度なピッチ抽出法の開発が望まれる。

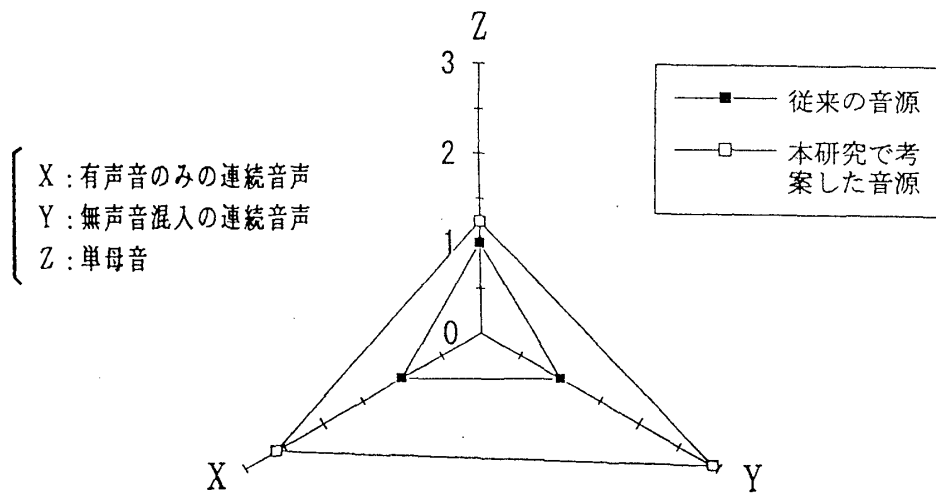


図2 従来の音源と本研究で考案した音源による合成音声の得点比

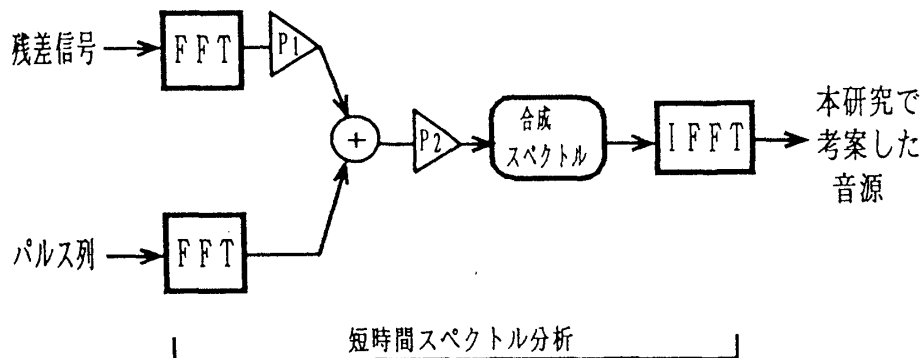


図1 本研究で考案した音源生成手法 (P1,P2: 振幅スペクトル比調節定数)

参考文献

1) 古井貞熙：デジタル音声処理, pp.122 ~ 124, 東海大学出版会, (1985).