

ショートノート

デルタ変調を利用した分析合成系のピッチ伝送方式

非会員 岩間 智史[†] 正員 島村 徹也[†]正員 鈴木 誠史[†]Transmission of Pitch Information by Delta Modulation in Analysis Synthesis
TelephonySatoshi IWAMA[†], Nonmember, Tetsuya SHIMAMURA[†]
and Jouji SUZUKI[†], Members[†] 埼玉大学工学部情報工学科, 浦和市

Faculty of Engineering, Saitama University, Urawa-shi, 338 Japan

あらまし 音声の分析合成系において、ピッチ情報の伝送にデルタ変調を行う方式を提案する。本方式は、ピッチ抽出誤りの自動補正および伝送に要するビット数の削減が可能である。本方式が有効なことを試聴試験により示す。

キーワード 分析合成系, ピッチ抽出, デルタ変調

1. まえがき

ピッチ駆動型の分析合成系では、有声音源の基本周波数または基本周期(ここではピッチ周波数またはピッチ周期と呼び、両者をピッチと総称する)の抽出精度が合成音声の品質に大きな影響を与える。しかし、完全に正確なピッチ抽出を行うことは、複雑な処理を行っても非常に困難である。ピッチ抽出誤りの多くは、倍ピッチや半ピッチのような大幅な誤りである。一方、実際の音声のピッチの変化には連続性があり、近接する抽出時点では極端なピッチの変化は少ない。この点に着目しピッチを抽出または補正する方式があるが^{(1),(2)}、複雑な処理を必要とする。

本方式では、ピッチの変化を3値デルタ変調により伝送することにより、簡単な処理でピッチ抽出誤りの補正を行う⁽³⁾。これにより、大幅な抽出誤りがあった場合においても、デルタ変調の変化幅以上にはピッチが変化しないためにピッチ抽出誤りの補正が可能である。

また、ピッチ情報が有声・無声の別を合わせて2ビット/フレームで伝送できるため、ピッチ情報に6~7ビット/フレーム、有声・無声の別に1ビット/フレーム必要とする従来法と比較し、大幅にビット数を削減できる。但し、通常の20msのフレーム周期では、ピッチの変化が荒くなり品質に影響がでるため、ピッチの情報だけを10msごとに伝送する。この場合でも通常の半分のビット数に削減できる。

2. 本方式の動作

本方式の動作を図1を用いて説明する。aは有声

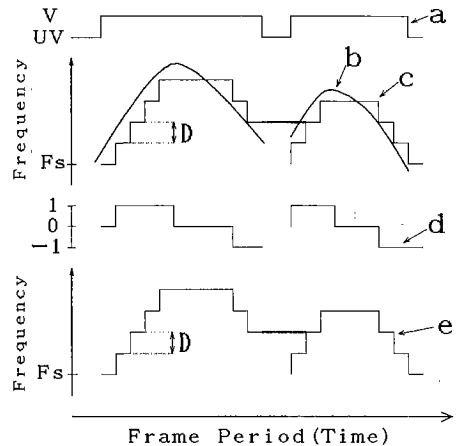


図1 デルタ変調による処理過程
Fig.1 Process of delta modulation.

(v)/無声または無音(uw)の別, bは抽出されたピッチ周波数, cは初期値が F_s である参照周波数, dは伝送される符号, eは復号されたピッチ周波数, Dは量子化幅を表す。なお, 図1ではピッチ周波数を変調, 伝送しているが, ピッチ周期を伝送する場合の動作も同様である。

有声のフレームで, ピッチ周波数 b が参照周波数 c より一定の量子化幅 D 以上大きい場合, 伝送される符号 d を +1 とし, 参照周波数 c を D だけ増加する。同様に, b が c より D 以上小さい場合, d を -1 とし, c を D だけ減少する。b と c の差の絶対値が D より小さい場合には, d を 0 とし, c は変化させない。

一般にピッチ周波数は, 起声に伴って低い周波数から上昇するため, 参照周波数の初期値 F_s は, 最低のピッチ周波数の近傍に設定する。無声部分が一定の長さ(フレーム数: 500ms 相当)以上続いた場合には, 音声の休止があったと判断して参照周波数を F_s にリセットする。無声部分が一定の長さ以下であった場合には, 図1の点線部分のように参照周波数を保持する。

3. ピッチ変化量の統計

本方式では, 量子化幅 D の値が品質に影響を与える。D の値が小さすぎる場合には, ピッチの急な変化に追従できない。また, 大きすぎる場合には, ピッチが定常になる区間が長くなる。この D の最適値を求めるために, 実際の音声のピッチの変化の統計をとった。それぞれ4種類の男声サンプル(M1~M4), 女声サンプル(F1~F4)を標本化周波数10kHzで標本化し, フレーム周期: 10ms, 分析次数: 16次で k パラメータにより分析した。ピッチ抽出は分析残差信号の自

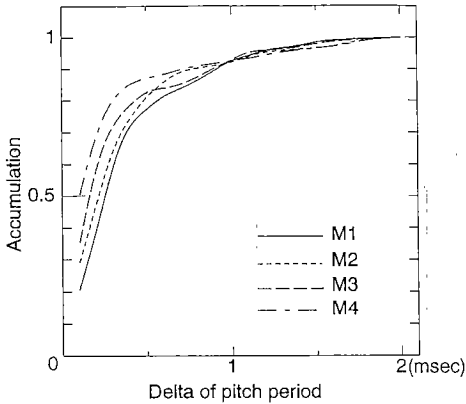


図2 ピッチ周期の変化の累積分布(男声)

Fig. 2 Accumulated distribution of pitch period on 4 male speakers.

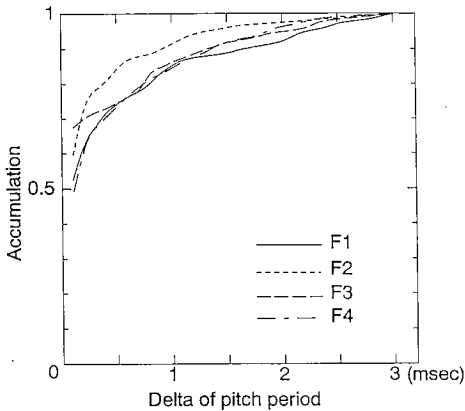


図3 ピッチ周期の変化の累積分布(女声)

Fig. 3 Accumulated distribution of pitch period on 4 female speakers.

己相関関数により行った。発声内容はいずれも、「屏風絵と如来像に対する興味」である。抽出したピッチ周期の隣合ったフレームごとの変化量を求め、その累積分布を図2(男声)、図3(女声)に示す。累積度は、全体を1に正規化している。なお、大幅なピッチ抽出誤りがあったものは除外している。

図2、図3において、累積度が0.5である変化量に D の値を設定し、デルタ変調を行った場合には、変化量が過大である場合と過小である場合とが同程度発生することになる。図2より、男声では約0.1 msから0.3 ms程度で累積度が0.5になっていることがわかる。また、図3より、女声では0.1 msで累積度が0.5を超えており、ピッチの変化が小さいことがわかる。

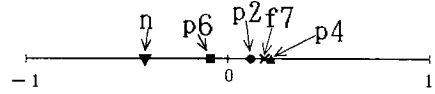


図4 対比較試験の結果

Fig. 4 Result of pair comparison test. (Thurston: case V)

4. 合成音の対比較試験による評価

3.で求めた値を参考に、ピッチ周波数と周期を伝送するそれぞれの場合について、 D の値を変えて数種類の合成音声を作成した。音声サンプルは3.で分析した男声4種類を用いた。合成フィルタの励振音源には、Rosenbergの声帯音源近似モデル⁽⁴⁾を通常のパルス波の代わりに用いた。なお、参照周波数(周期)の初期値 F_s は100 Hz(10 ms)である。

本方式の分析合成音は、ピッチ抽出誤りの存在する分析合成音に比し、誤りに起因する不自然さ、雑音が少ない。ここで、ピッチ周期を伝送した場合の合成音声は、ピッチ周波数を伝送する場合より平均して品質がよく、また、 D の値は周期では0.4 ms程度、周波数では7 Hz程度が適していると主観的に判断できた。そこで D を周期で0.2 ms(p_2 とする、以下同様)、0.4 ms(p_4)、0.6 ms(p_6)、周波数で7 Hz(f_7)に設定した場合と、デルタ変調を行わない場合(n)の5種類の合成音について、対比較試験を行った。各合成音対は6回ずつ、6名の被験者に呈示された。評価基準は「自然で聴きやすい」である。結果をサーストンの心理尺度(ケースV)で表したものを図4に示す。また、 p_4 (太線)とデルタ変調を行わない場合(細線)のピッチの変化を図5に示す。

この結果から、本方式による合成音声は、デルタ変調を行わない場合より評価が高いことがわかる。特に、 p_4 、 f_7 の評価が高く、不自然なピッチの変化に対する平滑化があり、また、本来のピッチの変化に対する追従性も良いと言える。

5. むすび

ピッチの変化分をデルタ変調で伝送することにより、ピッチ抽出誤りの補正が可能であることを、対比較試験により確認した。また、伝送に要するビット数が節減できるため、これを他のパラメータに振り分けるか、全体のビットレートを小さくすることが可能である。

男声の場合の変化幅 D の最適値を求めたが、初期値 F_s の最適値および男女共通の各パラメータの最適値も同様に求めることが必要である。また、ピッチが一

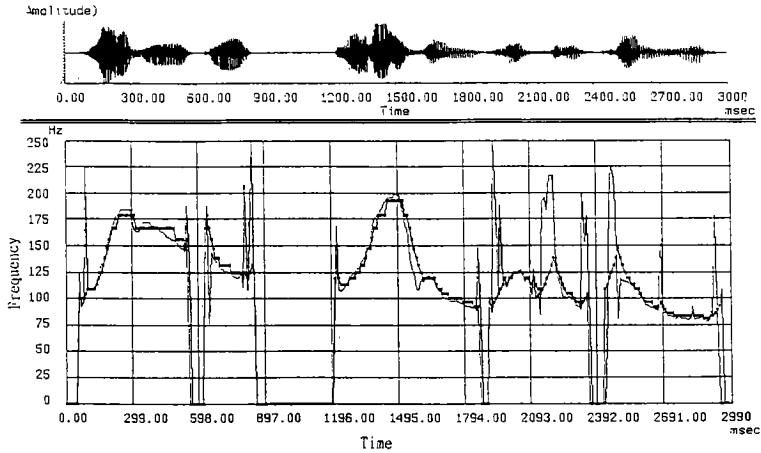


図 5 ピッチ周波数の変化 (細線: 抽出されたピッチの変化, 太線: 伝送されたピッチの変化)

Fig.5 Traces of pitch (fundamental) frequency. Thin line: extracted, Solid line: transmitted.

定になる部分を生じやすいため、ピッチの変化に揺らぎを与える等の処理を検討中である。

文 献

- (1) 伏木勝信: “ピッチ周期系列最適選択の一方式”, 日本音響学会講演論文集, pp. 35-36(1977-10).
- (2) 谷戸文広, 来山征士, 博松 明: “ボコーダの適応ピッチ抽出法”, 音響学会音声研資, S 79-10, pp. 73-80(1979-05).

- (3) 鈴木誠史: “音声分析合成系のピッチ周波数伝送方式”, 公開特許公報, 昭 56-121098 (1981-09).
- (4) Rosenberg A. E.: “Effect of glottal pulse shape on the quality of natural vowels”, 77 th ASA meeting, E 6 (1969).

(平成 5 年 1 月 21 日受付)