論文 / 著書情報
Article / Book Information

論題(和文)	ハフ変換による基本周波数情報を用いた耐雑音音声認識の高性能化の検討
Title(English)	
著者(和文)	安井 英己, 篠田 浩一, 古井貞熙, 岩野 公司
Authors(English)	Hideki Yasui, Koichi Shinoda, SADAOKI FURUI, Koji Iwano
出典(和文)	日本音響学会2009年春季講演論文集, Vol. , No. , pp. 35-38
Citation(English)	, Vol. , No. , pp. 35-38
発行日 / Pub. date	2009, 3

ハフ変換による基本周波数情報を用いた耐雑音音声認識の高性能化の検討*

◎安井英己, 篠田浩一, 古井貞熙 (東工大), 岩野公司 (武工大)

1 はじめに

これまで, 韻律情報を利用した連続音声認識の研究としてスペクトルの調波構造を韻律情報として利用した研究 [1] が報告されている. また, 雑音環境下での連続音声認識において基本周波数 (F_0) 情報を用いて音声認識性能を向上させる研究が行われている [2, 3]. F_0 情報は句や単語境界の推定, 有声部と無声/無音部の境界推定に役立ち, 雑音環境下で頑健に抽出することが出来れば, 雑音重畳音声の認識性能向上に有効である. 岩野らは時間-ケプストラム平面をハフ変換することで得られる F_0 情報を利用し, 雑音環境下での連続数字音声認識において手法の有効性を確認している [4].

筆者らは雑音環境下での大語彙連続音声認識においても, ハフ変換による F_0 情報の利用による有効性を確認している [5]. スペクトルサブトラクション法 [6] を適用した音声波形から抽出した MFCC, F_0 情報特徴量を組み合わせることで, 全ての雑音条件において単語正解精度が向上した. 各雑音条件での単語正解精度の平均が MFCC のみでは 52.7%であったが, MFCC と F_0 情報特徴量を使用した場合では 54.2%に改善した. しかし, ハフ変換による F_0 抽出には大きな計算量を必要とするため, リアルタイムでの動作が困難である.

本稿では, リアルタイムでの動作を実現させるために, ハフ変換による F_0 抽出の高速化を提案する.

2 F_0 情報抽出

高木ら [1] はスペクトルの調波構造を特徴量として利用する際に, 各分析フレーム (10ms 程度) 毎のケプストラムの高次ピークの強さを利用している. しかし, 雑音環境下では, 求めるべき音声の基本周波数に対応するピークと, 雑音によって発生するピークが混ざり合ってしまう. そのため, 1 フレームのケプストラム情報からでは F_0 を頑健に抽出できない場合が多い. そこで, 関らは音声の F_0 パターンの時間連続性を利用した, 雑音に頑健な F_0 の抽出法を提案している [7]. この手法では, 適当な窓幅で時間-ケプストラム領域を切り出し, ハフ変換によりその中の最も優位な直線を取り出すことで, 時間連続性が考慮された, 雑音に頑健な F_0 抽出を行っている. 本稿でも, この手法を利用して F_0 抽出を行う.

2.1 ハフ変換

ハフ変換は画像処理の分野でよく利用される手法で, 雑音を含む画像から直線, 円, 楕円といったパラメトリックな図形の特徴を抽出するのに有効な手法である [8].

直線検出のためのハフ変換は, (傾き m)-(切片 c) 座標系を用いるのが基本手法である. 変換対象画像 (x-y 平面) に n 個の画素 (x_i, y_i) $(i = 1, \cdots, n)$ が存在するとき, 各点を次式を用いて m-c 平面上の直線に変換する.

$$c = -x_i m + y_i \quad (i = 1, \cdots, n) \tag{1}$$

この時, m-c 平面の直線上の点に, 点 (x_i, y_i) の輝度を累積する. この操作を m-c 平面への投票と呼ぶ. x-y 平面上の全ての点を m-c 平面に投票した後で, m-c 平面上で投票値の累積が最大となる点 $(\hat{m}, \hat{c})$ を選び, 以下の式で逆変換することで, 最も優位な x-y 平面上の直線を抽出する.

$$y = \hat{m}x + \hat{c} \tag{2}$$

2.2 ハフ変換による F_0 抽出

ハフ変換によりケプストラムから F_0 を抽出する過程を Fig. 1 に示す.

サンプリング周波数 16kHz の音声データを, 分析窓長 32ms, フレーム周期 10ms で 256 次元のケプストラムに変換する. ピークの探索範囲はケ

*Enhancement of noise-rubust speech recognition performance using fundamental frequency information extracted by Hough transform By Hideki Yasui, Koichi Shinoda, Sadaoki Furui (Tokyo Institute of Technology), and Koji Iwano (Musashi Institute of Technology)

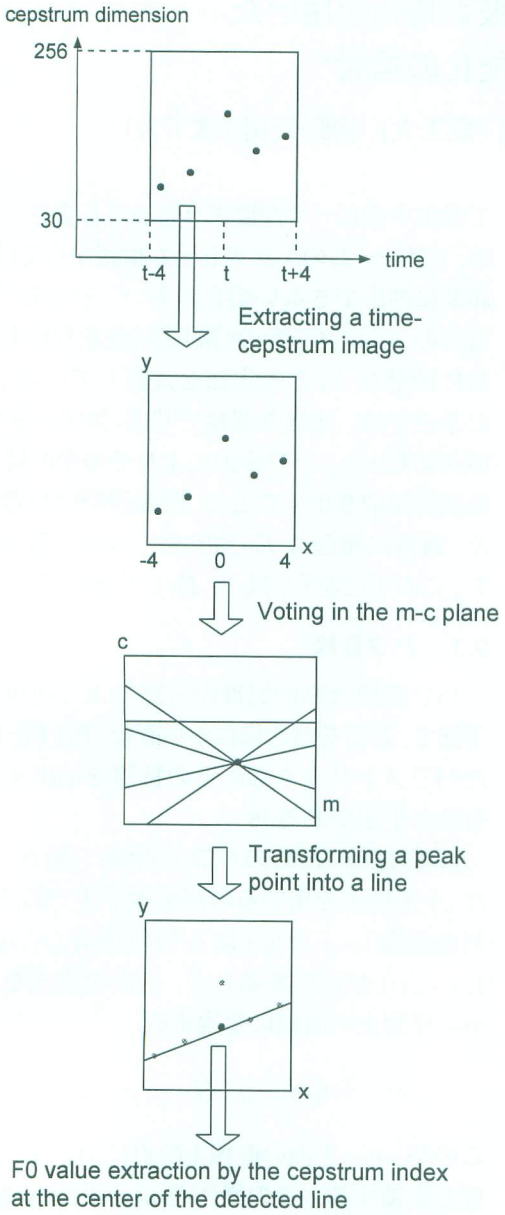

Fig. 1　ハフ変換により F_0 を抽出する過程.

プストラムの 30 次元目以上 (F_0 で 540Hz 以下) に限定する. さらに, 雑音の重畳した音声ケプストラムは, 低次部分ほどピーク値が大きくなる傾向がある. それを補正するため, 探索領域の低次部 (30〜140 次元) の d 次のケプストラムに次式で示す値 k_d を乗算しておく.

$$k_d = 0.6 + 0.4 \sin\left(\frac{d-30}{140-30} \times \frac{\pi}{2}\right) \qquad (3)$$

次に, F_0 を求めたいフレームを中心に, 前後 4 フレーム, 計 9 フレームの時間- ケプストラム画像を切り出し, ハフ変換を行う. この時, 各画素の輝度値はケプストラムの値であり, この値が投票値となる. ハフ変換によって得られた直線の中点のケプストラム次数を最終的に決定されたピーク箇所とし, F_0 を計算する. この操作を全てのフレームについて行うことで, 9 フレーム分の連続性が考慮された F_0 が抽出される.

3　ハフ変換による F_0 情報抽出の高速化

ハフ変換により 1 フレームの F_0 を抽出する際に, 複数フレームにわたる時間-ケプストラム画像に対してハフ変換を行い, m-c 平面を足し合わせる必要がある. そのため大きな計算量を必要とし, リアルタイムでの動作を困難なものとしている. そこで, 前フレームにおいて F_0 を抽出する際に用いた m-c 平面の投票値を再利用することで, 計算量の削減を行う.

時間-ケプストラム領域上の点 P について考える. 時刻 t で切り出した時間-ケプストラム画像において, 点 P に対応する x-y 平面上の点を $(x_i(t), y_i(t))$ とする. 点 P は m-c 平面上の直線 $c = -x_i(t)m + y_i(t)$ に変換されるが, このとき, c を t と m の関数とみなし, $c(t,m)$ と表す. すると, $c(t,m)$ は以下の式で与えられる.

$$c(t,m) = -x_i(t)m + y_i(t) \qquad (4)$$

ここで, 時刻 $t-j$ で切り出した時間-ケプストラム画像において, 点 P に対応する x-y 平面上の点を $(x_i(t-j), y_i(t-j))$ とすると, 以下の式の関係が成り立つ.

$$x_i(t) = x_i(t-j) - j \qquad (5)$$
$$y_i(t) = y_i(t-j) \qquad (6)$$

式 (5), (6) を式 (4) に代入することで次式が得られる.

$$\begin{aligned} c(t,m) &= -(x_i(t-j)-j)m + y_i(t-j) \\ &= -x_i(t-j)m + y_i(t-j) + jm \\ &= c(t-j,m) + jm \end{aligned} \qquad (7)$$

したがって, $c(t,m)$ は $c(t-j,m)$ を用いて表すことができ, j フレーム前に対応する時間-ケプストラム画像をハフ変換した時の m-c 平面の投票値を利用することができる.

w フレームの時間-ケプストラム画像を用いるとし, 先頭・末尾フレームの計算時に窓幅で音声

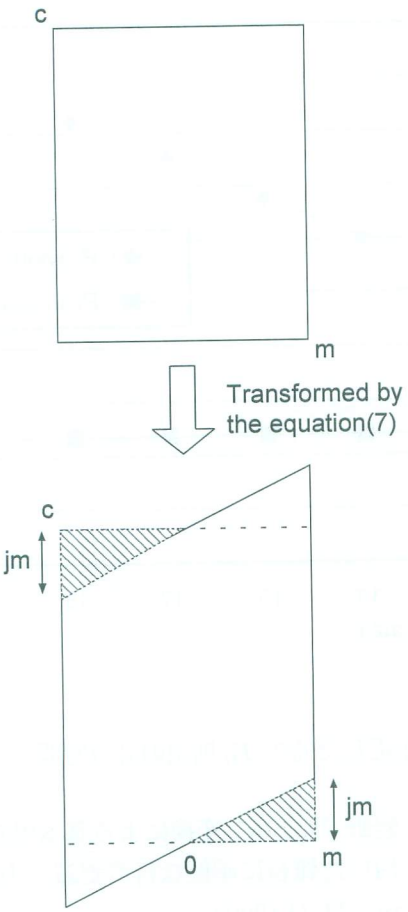

Fig. 2　m-c 平面変換.

波形が存在しない領域を 0 として計算することにすると, n フレームの音声に対して F_0 を抽出する際に, 従来手法では $w \times n$ フレームの m-c 平面を足し合わせる必要がある. 提案手法では 1 フレーム目で w フレームの m-c 平面の足し合わせを行い, 2 フレーム目以降は「不必要となった m-c 平面の減算」, 「式 (7) を用いた m-c 平面の変換」, 「新しく必要となる m-c 平面の加算」と 3 フレーム分の計算を行うため, $w + 3(n-1)$ フレーム分の m-c 平面の計算だけでよい. そのため, n が十分大きな場合には切り出すフレーム幅によらず計算時間が一定となることが期待できる. Fig. 2 に式 (7) を用いた m-c 平面変換の様子を示す. 変換を行う際に jm だけずれが生じるため, 図中の斜線部の領域において不足が起こる. そのため, 従来手法と比べて余分に m-c 平面の領域を用意する必要が生ずる.

なお, 提案手法では従来手法と同じ F_0 の値が得られるため, 文献 [5] で確認された認識性能の改善が, 提案手法でも同様に得られる.

Table 1　1 発声 (12.92 秒) に対するハフ変換による F_0 抽出の計算時間.

	Computational time(s)
Previous	8.74
Proposed	4.81

4　評価実験

4.1　実験条件

音声データとして, 新聞記事読み上げ音声コーパス (JNAS) 中の男性話者 1 発声 (12.92 秒) を用いた.

従来手法では m-c 平面への投票を行う際に $-20 \leq m \leq 20$, $30 \leq c \leq 256$ の範囲に m, c ともに 0.5 刻みで量子化されたビンを用意して投票を行った. 提案手法では時間-ケプストラム画像を切り出す窓幅に合わせて c の範囲の変更を行って量子化されたビンを用意した. 9 フレームの窓幅で切り出した時間-ケプストラム画像を用いる際は $-50 \leq c \leq 336$ の範囲に量子化されたビンを用意して投票を行った.

計算機は Intel Core 2 Duo 2.4GHz のものを使用した.

4.2　実験結果

まず, 9 フレームの窓幅で切り出した時間-ケプストラム画像を用いて F_0 抽出の高速化について性能比較実験を行った. Table 1 に提案手法による音声ケプストラムからハフ変換により F_0 を抽出する際の計算時間を示す.

比較実験において, 提案手法では従来手法との比較で F_0 抽出時の計算時間が 45.0%削減された. 発話時間との比較で提案手法では 37.2%の時間で F_0 が抽出可能となり, 単純なタスクでは F_0 情報を利用した音声認識の実時間動作が可能である.

次に, 時間-ケプストラム画像を切り出す際の窓幅を変更した時の F_0 抽出の計算時間について性能比較実験を行った. Fig. 3 に F_0 を抽出する際の計算時間を示す.

従来手法では窓幅が増加するにつれて, 計算時間も増加するが, 提案手法では窓幅によらずほぼ

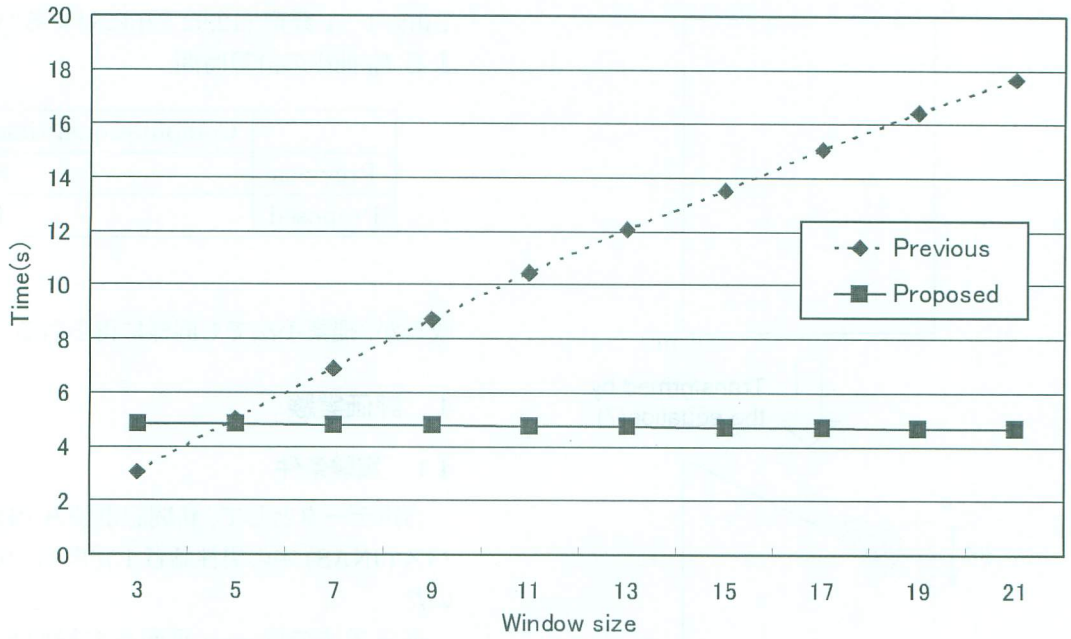

Fig. 3　1 発声 (12.92 秒) に対する窓幅を変更した時の F_0 抽出の計算時間.

一定の計算時間となった.

5　おわりに

リアルタイムでの F_0 情報の利用を実現させるために, ハフ変換による F_0 抽出の高速化を提案した. 9 フレームの時間-ケプストラム画像を用いる際, 従来手法との比べ, ハフ変換による F_0 抽出の計算量が 45.0%削減された.

今後の課題としては, m-c 平面の量子化の幅などを変化させて F_0 抽出を高速化させた時の認識性能の検証やリアルタイムでの F_0 情報を利用した音声認識システムの検証などが挙げられる.

参考文献

[1] 高木 他, "音声認識のためのスペクトルの調波構造の利用," 秋季音講論, pp. 3-4 (1997).

[2] L. Gu *et al.*, "Perceptual harmonic cepstral coefficients for speech recognition in noisy environment," *Proc. ICASSP2001*, pp. 125-128, 2001.

[3] A. Zolnay *et al.*, "Robust speech recognition using a voiced-unvoiced feature," *Proc. ICSLP2002*, pp. 1065-1068, 2002.

[4] 岩野 他, "ハフ変換による基本周波数情報を用いた雑音に頑健な音声認識," 秋季音講論, pp. 23-24 (2002).

[5] 安井 他, "スペクトルサブトラクションとハフ変換による基本周波数情報を用いた耐雑音音声認識," 秋季音講論, pp. 3-6 (2008).

[6] S.F.Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans.ASSP*, vol. 27, no. 2, pp. 113-120, 1979.

[7] 関 他, "ハフ変換による雑音に頑健な基本周波数抽出法," 情報処理学会研究報告, vol. 2001, no. 100, pp. 9-14 (2001).

[8] P.V.C.Hough, U.S. Patent #3069654(1962).