

My アルゴリズムで聞き分け率アップ! ボイス制御マシンにも使える!?

ラズベリー・パイでトライ! 音声認識

お試しプログラムも用意!

第6回

データ量は生音の15%! 必要な値だけ計算すればOK

認識処理に使う重要データ! 音声特徴量MFCCを求める

安藤 章悟, 小黒 玲

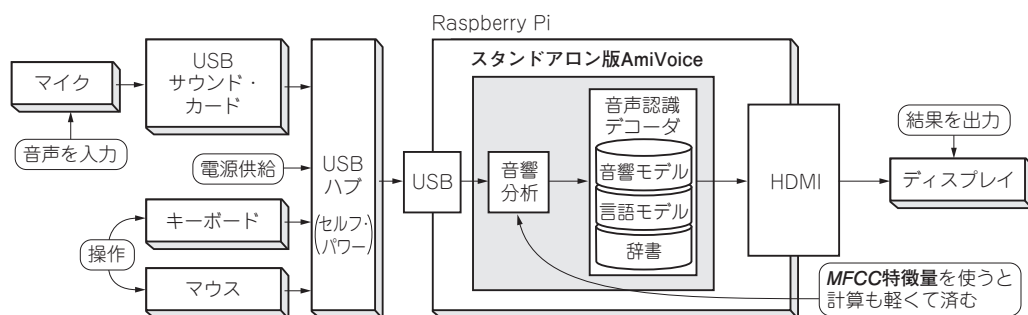


図1 音声認識では音声の特徴を抽出したMFCCという値を使う

音声認識ソフトAmiVoiceのRaspberry Pi版は本誌ダウンロードページ(<http://www.cqpub.co.jp/interface/download/amivoice.htm>)から入手できる

音声認識の処理では音声データをそのまま使わず、図1のように音声の特徴を抽出したMFCC(メル周波数ケプストラム係数)という値に変換します。

MFCCは生の音声データよりも声道特性(しゃべったときにどのような口の形をしていたか)などの音素ごとの違いが顕著な特徴をよく表現することができ、音声認識に適していると考えられています。また、生の音声データと比べてデータ量が小さくなります。

今回はこのMFCCについて、その計算方法を詳しく解説します。さらに、生の音声データと比較してデータ量がどのくらい小さくなるかを計算してみます。

特徴量計算の手順

MFCCの計算の流れを図2に示します。

- ステップ1…フレームの切り出し
- ステップ2…窓関数をかける
- ステップ3…離散フーリエ変換
- ステップ4…対数をとる
- ステップ5…メル・フィルタ・バンクにかける
- ステップ6…離散コサイン変換
- ステップ7…MFCC値を得る
- ステップ8…前後のフレームから Δ MFCCと $\Delta\Delta$ MFCCを計算
- ステップ9…MFCC, Δ MFCC, $\Delta\Delta$ MFCCを合

せて特徴量とする

まずステップ1で音声データを20～30msのごく短い区間(フレームと呼ぶ)に分割して切り出します。切り出したらその各フレームごとに計算処理をすることでステップ7のMFCCを得ます。

MFCCは12個程度の数値の組(12次元のベクトル)ですが、この組がフレームの数と同じ数だけ作られることになります。このMFCCを元に、前後の変化量を計算したステップ8の Δ MFCC, $\Delta\Delta$ MFCCという値も使います。これらのMFCC, Δ MFCC, $\Delta\Delta$ MFCCを合わせた数値の組を特徴量と呼びます。

実際の計算

それではMFCCの計算方法について説明します。

● ステップ1…フレームの切り出し

音声認識では音素を20～30msのフレームに分割して、各フレームごとに特徴量を抽出します。

図3は筆者の発音した「あ」という音です。各フレームは図3のように隣同士を重ねて10～15msずつずらしていきます。切り出したフレームは図4のようになります。

● ステップ2…窓関数をかける

フレームを切り出したら離散フーリエ変換を行いま