

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4回

話す人が違ってOK! 音や文章を抽出するためのキー・テクノロジー  
本格認識の前に…声の特徴を表すデータベース音響モデル入門

安藤 章悟  
小黒 玲

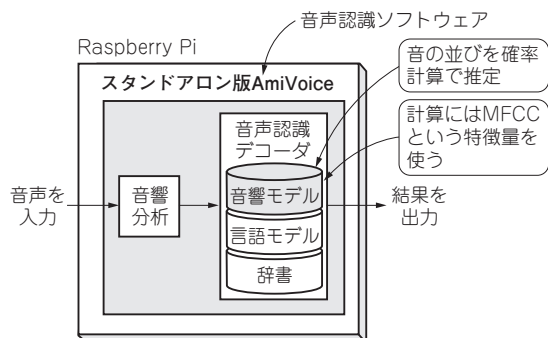


図1 シャべった音を認識するためには音響モデルというデータベースが必要



図2 音声波形は音素で構成されている  
「おはよー」としゃべった場合は五つの音素になる

音声認識では、音響モデル/辞書モデル/言語モデルという三つのデータを使用します。今回は声の特徴を表すための音響モデルについて解説します。

発話をコンピュータが認識するには、音の特徴を抽出して、データベースと照らし合わせて単語を推定します。このときにデータとして使われるのが、音響モデルです。また、音素の特徴検出にはMFCC (Mel-Frequency Cepstrum Coefficients:メル周波数ケプstram係数) という値を利用します。

今回は、音声認識の音響解析で重要な、音響モデル

とMFCCという二つのキー・テクノロジーを解説します。

## 発音を確認するためのデータベース …音響モデル

### ● 声の特徴は音素の組み合わせで表す

音響モデルでは音素を単位にモデル化します。例えば図2は「おはよー」としゃべった音の波形を耳で聞きながら音素ごとに/o/h/a/y/o/と区切ったものです。波形を見ると音素の切れ目で特徴が変わっています。この音素一つ一つが音響モデルを作る際の最小単位です。

図2から最初の音素「o」を拡大した波形が図3です。これを見ると、

- 前半・中盤・後半で特徴が異なる
  - 特徴の似た波形が何回も繰り返し表れている
- また、図3には表れていませんが、
- 音素の長さはしゃべり方やしゃべる人によって異なる

という傾向もあります。音響モデルではこのような音声の傾向を計算できるようにするために隠れマルコフ・モデル (HMM; Hidden Markov Model) という確率計算機を使います。

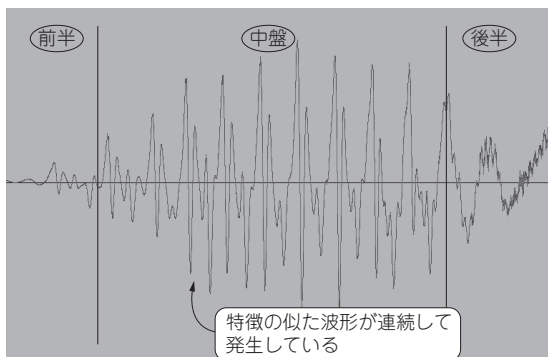


図3 一つの音素でも前半・中盤・後半で特徴が異なる  
「お」を拡大した波形