

音声認識ICが言葉を聞き取るメカニズム

荒金 康人, 宮崎 善行

本稿では、音声認識ICが言葉を聞きとるメカニズムを解説します。人間が声を出すしくみと聞きとるしくみの理解が欠かせません。声を出すしくみがわからないと、音声データがどうなっているかわかりません。聞きとるしくみがわからないと、どのような処理を行えば人間の認識に近づけられるかわかりません。

実際には次のような処理で音声で認識しています。

- 入力された音声の特徴を抽出
- 事前に用意しておいた音声データ（音響モデル）と比較
- 確からしい音の確率計算
- 結果の出力

音響モデルはパソコンで繰り返し学習させて生成し、音声認識ICに格納しておきます。（編集部）



音声認識の分類と特徴

音声認識とは、音声から話者が話した単語や文を推定することです。最近ではCPUやアルゴリズムの進化によりワンチップで処理できる音声認識ICが登場しています。

音声認識は、図1のように大きく分けて認識と学習の二つの処理を行います。

認識処理では、入力された音声信号を分析して音声の特徴を抽出し、あらかじめ作成した音声データ（音響モデル）から抽出した特徴が生成される確率を比較、計算し、結果として出力します。

学習処理は、複数のサンプル音声から音声の特徴を抽出し、それらの特徴を生成する確率が最大となるように調整して音響モデルを作成します。

音声認識をパソコンで実現するときは、音響モデルの学習機能を備えたアプリケーション・ソフトウェアがあります。ICで実現するときは多くの場合、あらかじめ作成した

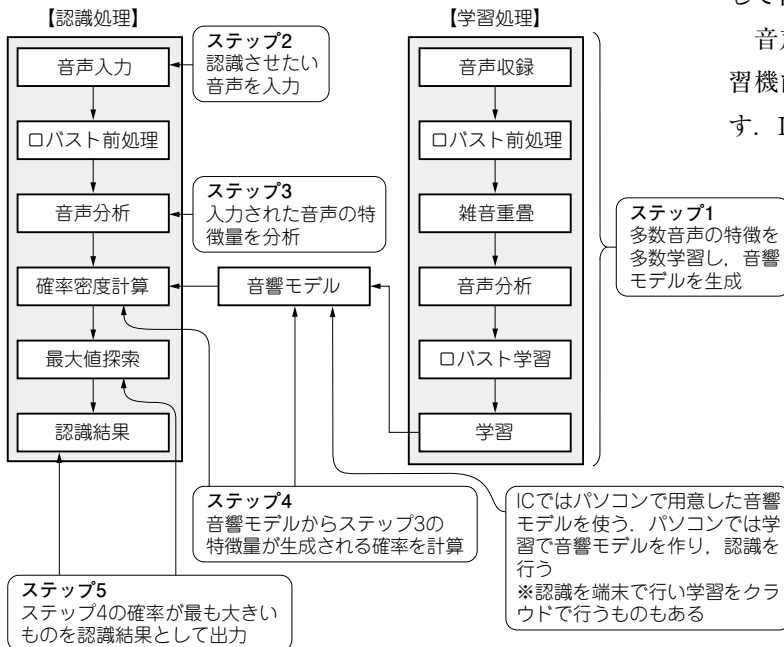


表1 日本語の主な音素はアルファベットで表現する

分類	音素
母音	/a/, /i/, /u/, /e/, /o/
子音	/k/, /s/, /t/, /c/, /n/, /h/, /m/, /r/, /g/, /z/, /d/, /b/, /p/
半母音	/j/, /w/
特殊モーラ	/N/, /Q/, /H/

図1 音声認識の基本的な処理。入力された音声の特徴を抽出して音響モデルと比較し、認識する

雑音ロバスト処理はAppendix 1参照