音声特徴量の抽出

音声の生成



一方、言葉（言語情報の伝達に使われる音声）

「音素」に区分化される

日本語：あいうえお＋ん　と　子音により分類。

音声生成の信号モデル

音声信号出力

調音フィルタ

h(n)

パルス

（声帯振動）

インパルス

（破裂音）

ノイズ

調音フィルタ：全極型のシステムではないか？（共振特性のみで説明されるシステム）

↓

仮説の妥当性を説明するものとして

1. 調音器官を単純な音響管の接続とみると、鼻子音を除く音素には反共振が存在しない
2. 人間の視覚特性がスペクトルのピークに敏感であること

全極型システムはフィードバックループにより構成されるよ！

＊音声信号のスペクトル分析

与えられた信号を生成した調音フィルタの特性を信号により推定

→信号の周波数領域表現を用いる。

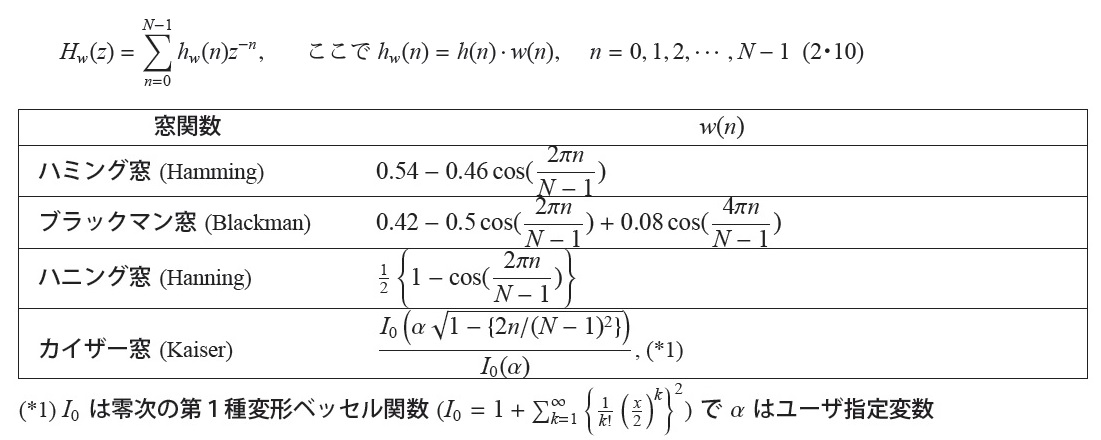
・分析するには

音声から連続したミリ秒単位の区間を切り出してそれが確率過程に従うと仮定してスペクトル解析を行う

信号に分析窓をかけて信号系列をとりだす。

この作業はフレーム化処理と呼ばれる。

・窓関数



フレーム化処理により短時間フーリエスペクトルがえられる。

（離散フーリエ変換により表現）

実際は高速フーリエ変換が用いられる。

音声認識に用いられるのはパワースペクトル（音声の振幅スペクトルの二乗値）

であり、複素平面上の距離であらわされる。（実部と虚部の二乗和）

＊音声の線形分析予測

音声が全極形伝達関数を持つ調音フィルタであることを前提に効率的に音声信号のスペクトルの概形を求める方法。

LPC分析とも呼ばれる。

＊音声信号のケプストラム分析

観測された音声から音源の信号スペクトルと調音フィルタの特性のスペクトルを分離する方法。（音素を特徴づけるものとして調音フィルタの伝達特性を抽出したい）

・ケプストラム分析のアルゴリズム

1. 出力信号はS＝G\*Hと、入力とフィルタの伝達特性の積であらわされる。

対数をとり、積を和の形にする。

logS=logG+logH

このとき、音源Gにくらべて音源Hは周波数に対して振幅の変動が激しい。

周波数領域のスペクトルを時間領域に見立ててフーリエ変換（逆フーリエ変換）を行うと、

Hの逆フーリエ変換は低いケフレンシー帯域に、Gの逆フーリエ変換は高いケフレンシー帯域にエネルギーが集中する。

G,Hを逆フーリエ変換した結果をそれぞれG,Hのケプストラム係数という。

ケプストラム係数の低次の項には調音フィルタの特性、高次の項には音源の特徴を見ることができる。

すなわち、ケプストラム係数を低次で打ち切れば調音フィルタの特性を取り出すことができる。

＊LPCケプストラム係数

音声の実際の特徴ベクトルとして扱われることが多い。

# 音声特徴抽出の実際

1. MFCCパラメータ（メル周波数ケプストラム係数）

もっとも一般的な音声特徴量である。