

バーバルインターフェイス

# バーバルインターフェイス

- 音声認識
- 音声合成

# 入力手段としての音声の利点

- 誰でも容易に使用できる
- 手や目がふさがっていても使用可能
- 直接指示可能(固有名称など)
- 遠隔操作が可能
- 入力速度が速い

# 入力手段としての音声の欠点

- 周囲雑音の影響大
- 消去ができない
- 機密保持性が低い
- ブラウジングができない
  - 書き物における斜め読みのようなもの

# 音声の特徴と性質

音声 = 音韻性(言語性)  
+ 個人性  
+ 情緒性

音声認識とは、音声から音韻性情報のみを  
取り出すこと

# 母音とフォルマント構成

- 音声スペクトルを観察すると、特定のピークが時間的に移動していることが分かる。このピークをフォルマントという
- 周波数の低い順に第一フォルマント、第二フォルマント...と呼ぶ
- 母音が同じならば似たフォルマントのパターンが観察されるので、フォルマント周波数を計測すれば母音の判別ができる

# 音声認識システムの基本構成

- 音声分析 : 音声スペクトルを求める
  - FFT、線形予測分析など
- 照合 : 標準的な音声と入力音声との比較照合処理
- 言語処理 : 構文解析と意味解析
- 判定 : 最終的に入力音声内容を判定

# 音声認識装置の分類

- 特定話者と不特定話者
- 単語音声認識と文章音声認識
- 孤立音声認識と連続音声認識
- 認識対象語彙数



# 現状の音声認識のレベル

- 離散単語、不特定話者で、約95%程度の認識率
  - 認識率は状況によって大きく変化するのであまりこれにこだわっても意味が無い
- 対象を限定すれば実用的に使える

# 音声認識の問題点

- 雑音対策
- 未知音声、無意味語対策
- 意味認識なども含めた認識性能の向上

# 録音再生方式による音声合成

- あらかじめ録音された音声をそのまま再生する
- 音声の品質は高い
- 決まったことしかいえない
  - 内容が変わるたびに再度録音必要
- 天気予報など

# 録音編集方式による音声合成

- 音声を単語や文節単位などで録音しておき、必要な音声を接続して再生する
- 接続における連続性の良否が音声品質を左右
- 同じ単語でも文中で使用する位置により、何種類かの声の高さの録音が必要

# パラメータ編集方式による音声合成

- 単語や文節単位などで音声を録音したものを音声の生成モデルを使用して分析し、パラメータ時系列の形で蓄え、このパラメータ時系列の形で合成する
- パラメータで情報を貯えるため、大幅に圧縮して保存できる
- 声の高さや発話速度の変更が容易

# 規則合成方式による音声合成

- アクセント、イントネーションなどの韻律情報の生成を規則によって行い、音素や音節などの単位を合成することにより自動的に音声を合成する方法
- テキストデータからでも音声を合成できる
- 少ないデータ量でよい
- 音声の品質は低く、やや不自然