

# 自主課題研究

## 特殊な音声に対する音声認識の精度に関する調査

情報システム工学科3年 072 志馬弘基

### 1 はじめに

ヒトの話す音声言語をコンピュータによって解析し、話している内容を文字データとして取り出す処理を音声認識という。近年の音声認識技術の向上により、音声入力を用いた実用的な対話システムの実現を目指した研究が進められている。しかしながら、実用的な音声対話システムの構築には、音声認識誤りへの対応や音声認識そのものの精度向上が不可欠な状況である。本稿では、音声認識システム理解のための基礎的な研究として、既存する音声認識ソフトウェアを用いて、歌唱音や合成音などといった特殊な音声に対する認識精度の調査を行うことを目標とした。具体的には、以下の3つの特殊音声を用意し、それぞれについて音声認識精度の調査を行った。

1. 音程を変化させた音声
2. 音節間を伸ばした音声
3. 合成音声

尚、今回は音声認識ソフトウェアとして Dragon Naturally Speaking、音声合成ソフトウェアとして Aques Talk を用いた。

### 2 実験

#### 2.1 音程を変化させた音声に対する認識率

まず、男性が「金沢大学」と発声し録音した音声ファイルに対して全体の音程を変化させた場合の認識率について調査を行った。その結果、 $-4 \sim +5$  の範囲で音程を変化させた場合、ほぼ正しく認識を行うことが確認できた (1=半音)。この結果について離散フーリエ変換 (DFT) を適用すると、 $126[\text{Hz}]$ (オクターブ2のB)  $\sim 209[\text{Hz}]$ (オクターブ3のG#) の範囲で正しく認識できることが確認できた。

次に、各音節ごとに分割して音程をそれぞれ変化させた場合の音声認識率について調査した。その結果、ある程度滑らかに音程を変化させた場合は正しく認識できるが、急激に変化させた場合は認識が困難であることが確認できた。

#### 2.2 音節間を伸ばした音声に対する認識率

音節間を伸ばして発音した場合の認識率を調べた。その結果、通常に発声した時に比べて認識率は下がるが、5割程度は正しく認識されることが確認できた。

#### 2.3 合成音声に対する認識率

音声合成ソフトウェアを用いて発音させた場合の認識率を調べた。その結果、男性合成音では約8割認識可能であったが、ロボット合成音では全く認識が行えなかった。また、女性合成音でも認識できず、合成音声に対しては認識が困難であることが確認できた。

### 3 まとめ

本研究では、特殊な音声に対する認識率の精度を調査した。現在の音声認識ソフトウェアは、ユーザが単語を区切って発音したり、ゆっくり発音することなどを前提として提供されているため、やはり意図的に作成した特殊な音声に対してはそれほど高い認識率は得られなかった。今後の課題としては、今回取り上げた3つの特殊な音声を組み合わせた音声に対しての認識率の調査などが挙げられる。例えば、音声合成ソフトウェア VOCALOID を用いた時の認識率などは調査の余地があると考えられる。また、実際に音声認識を行うシステムの構築を行うことも今後の課題として挙げられる。