

聴覚障害者のための音声認識システム — 多数決原理による認識率の向上 —

川辺弘之[†] 杉森公一[†] 瀬戸就一[‡] 下村 有子[†]
金城大学[†] 金城大学短期大学部[‡]

1. はじめに

本研究を含む進行中の研究プロジェクトの目的は、聴覚障害学生が大学の講義を不自由なく受講できるシステムの構築である。当初、この研究プロジェクトでは、多数の入力ボランティアのキーボード入力によるノートテイクを構想していた[1]。「質（正確さ）より量（人数）」の概念にもとづいたノートテイクシステムである。そこで問題になったのは、入力ボランティアを多数確保することと、キーボード入力の正確さであった。そこで、この問題を解決するため、キーボード入力を講師による音声入力に置き換えることを我々は考えた。音声認識率は約80%と高くはないが、初心者によるキーボード入力より優る。したがって、音声認識と「質より量」の多数決原理により上の問題を解決できる。

音声認識エンジンには、言語モデルや音響モデル、音声辞書、そして認識のためのパラメータを変える自由度が存在する。これらを変えることで同一の音声であっても音声認識エンジンは異なった認識結果を返す。この場合、多数の音声認識結果から最終的な音声認識結果を多数決で抽出することになる。

本研究では、まず、多数決原理による認識率向上のモデルとそのコンピュータシミュレーション結果を簡単に紹介する。次に、音声認識プログラムに異なった設定を施した場合の認識結果、そして、多数決原理で抽出した結果を与え、本手法の有効性を示す。

2. モデルとコンピュータシミュレーション

我々のシステムでは、特性の異なった音声認識プログラムが並列実行される。したがって、講師が発した文章データを複数得ることができる。この中には、正しく認識された単語もあれば、誤って認識された単語もある。このとき、動作する音声認識プログラムの数が増えれば、正しい単語も多くなることが期待できる。一方、

単語の認識誤りの傾向とその発生率や発生箇所はランダムで、全く同じ認識誤りは現れないと仮定する。すなわち、複数の単語データにおいて、2つ以上同じ単語データが現れたならば、それは正しい単語であると仮定する。すなわち、認識誤りの完全なランダム性を仮定する。そして、複数の認識単語データから多数決により正しい認識箇所を抽出し、つなぎ合わせることで、元の文章の再現が可能になる。

したがって、音声認識プログラムの実行数が増加すれば、正しく認識する確率が向上するであろう。コンピュータシミュレーションにおいて、音声認識プログラムが確率 0.5 で正しく認識できるという条件下で並列数を変えた場合、6~8 並列で十分な精度が得られている。さらに、10 並列程度で 95%を超える認識率となる[2]。

3. 調整パラメータによる音声認識への影響

音声認識システムは認識エンジンと認識エンジンの調節パラメータとから構成される。同一の音声であっても、異なったパラメータで調整された認識エンジンに与える と、異なった認識結果を得る。

本研究では、音声認識エンジンとして Julius[3]を用いた。Julius では、音響モデルや、言語モデル、デコーダを変更可能であり、また、解析する際、無音期間の長さや言語スコアの重み、ビームの幅などの音声認識のアルゴリズムでのパラメータを調整できる。

Julius に「およそ桃太郎の話を知らない人はいない」という文章を与え、種々のモデル、音声辞書、パラメータを変えて音声認識させた結果が表である。最初の行が原文、最後の行が共通部分、その他が音声認識エンジンで認識された文である。認識された文において、上段が認識された単語、下段が音声認識のスコアである。単語のうち原文に含まれている単語の数を認識率とした。

それぞれの認識率は 6/12 (0.50) から 11/12 (0.92) であったが、認識率が悪い結果を含め、多数決原理により共通部分を取り出すことで最終的に 11/12 (0.92) の認識率となった。なお、

Voice Recognition System for Hearing Impaired Student ---
Improvement of Recognition by Principle of Majority Rule ---
[†]H. Kawabe, K. Sugimori, Y. Shimomura · Kinjo University
[‡]S. Seto · Kinjo College

