

Hybrid HMM/SVM を利用した筋電位に基づく日本語黙声認識

EMG-based Silent Speech Recognition for Japanese Language Using Hybrid HMM/SVM Architectures

学籍番号: 201121712

氏名: 大内 慶久

Yoshihisa OUCHI

近年、音声認識技術の発展により、音声入力インタフェースを用いたサービスが身近なものとなりつつある。しかし、公共の場所などでは、声を出して発話することがマナーの点で問題となることがある。また、音声発話の機能が失われた場合には代替システムが必要となる。これまでに、音声を用いずに発話内容を認識する研究が行われており、その一つとして筋収縮に伴って生じる筋電位信号を利用した筋電インタフェースがある。従来研究では日本語五母音あるいは二十単語程度を用いて高い認識精度を得ているが、大規模語彙による認識への拡張のために必要な子音の認識について、十分な検討を行った研究は見当たらない。

そこで本研究では、日本語で用いられる母音と子音を組み合わせた五十音を認識することを目的として、従来の認識手法を改良し、評価実験による検討を行う。具体的には、従来研究の日本語五母音の認識などで高い認識精度を得ていた隠れマルコフモデル (HMM) とサポートベクトルマシン (SVM) に注目し、時系列信号のモデル化に適した HMM の構造と、汎化性能に優れた SVM を組み合わせた Hybrid HMM/SVM を黙声認識のために利用することを提案する。Hybrid HMM/SVM は、HMM の各定常信号源からの出力において SVM による事後確率を利用し、比較的学习データが少ない場合においても十分な認識性能を示すことが期待できる。

日本語五十音の黙声認識実験として、表情筋及び頸部の五箇所筋電位を計測し、五十音のうち清音と撥音の四十六音のデータを用いて認識実験を実施した。実験では、まず四十六音について、それぞれの音が含む母音を正解として認識させた。例えば、「き」を「い」、「し」を「い」、「め」を「え」などとした。このとき、Hybrid HMM/SVM による平均認識率は 95.5% で、従来手法である HMM での 67.3%、及び SVM での 78.8% と比較して優れた認識性能が得られた。次に、四十六音をそれぞれの音（例えば、「き」は「き」、「し」は「し」、「め」は「め」など）として認識する実験を行った結果は、Hybrid HMM/SVM での認識率が 35.9% となり、HMM での 19.3%、SVM での 21.3% に比べて高い認識率を得た。これらの実験から、黙声認識における Hybrid HMM/SVM の利用が、従来手法に比べて有効であることが確認できた。

四十六音を各音として認識した結果を分析すると、「あ」「か」「は」「ま」「や」「ら」行の音が平均認識率以上であるのに対して、それ以外の行の音では認識率が 20% 以下となり、音声学における子音の調音位置による分類が影響している可能性が示唆される。今後は、子音をグループ化することでその認識可能な範囲を検討するとともに、より多くのデータを収集しながら認識モデルの効果をさらに検証する必要がある。

研究指導教員: 三河 正彦

副研究指導教員: 田中 和世