ISS-A-065

機械読唇を用いた母音識別および子音識別による発話語推定

阿部　聖志朗　　平原　誠
法政大学大学院　理工学研究科　応用情報工学専攻

1. はじめに

機械読唇とは，動画像情報のみで発話内容を推定する技術である．これは，雑音化の音声認識や聴覚障がい者のコミュニケーション支援として研究が進められている．しかし日本語における機械読唇は母音のみの認識が主流であり，発話内容を十分な精度で推定するまでには至っていない[1][2]．

本研究では，母音認識で得られた音素ごとの母音列と子音認識で得られた結果を用いて発話内容推定を行うことを目標とする．

2. 提案手法

本システムは母音認識部と子音認識部，発話内容を推定する文字列推定部の3つから成る．

母音および子音認識部は RNN(Recurrent Neural Network)を用いて学習により構築した．特徴量抽出には唇，鼻，輪郭，眉などの座標点を取得出来るソフトウェアライブラリを使用した．得られた座標点から「口唇中央の縦幅」と「口唇中央の横幅」に加え「上唇中央から顎の中央までの縦幅」の3つを特徴量として抽出した．母音認識部は，5つの母音に閉口状態を加えた[a,i,u,e,o,N]の6種類を分類した．子音認識部は，子音の行である[k,s,t,n,h,m,y,r,w]の9種類を分類した．

文字列推定部への入力は現状では母音認識部の出力のみを用いている．母音認識部から出力された時系列データを数音の母音列に変換した．その母音列とコーパス中の単語の母音列とのレーベンシュタイン距離が小さいものから列挙する．

3. 実験方法

母音認識部の学習データは，各音につき 50 個の動画を撮影して作成した．子音認識部の学習データは，各行につき合計50個になるように動画を撮影して作成した．テストデータにはコーパスの中からランダムに50個の単語を抜き出し発話した動画を用いた．

RNN による母音と子音の認識率はそれぞれ約 90%と約30%であった．図1に子音認識部にテストデータ「すみません」を入力した場合の出力結果を示す．「み」と「ん」の口を「ま行」として認識していることが分かる．

文字列推定部の結果にはテストデータの母音列とコーパスの母音列が完全に一致するものは存在しなかった．変換した母音列が部分的に誤変換を起こしている箇所が多かった．

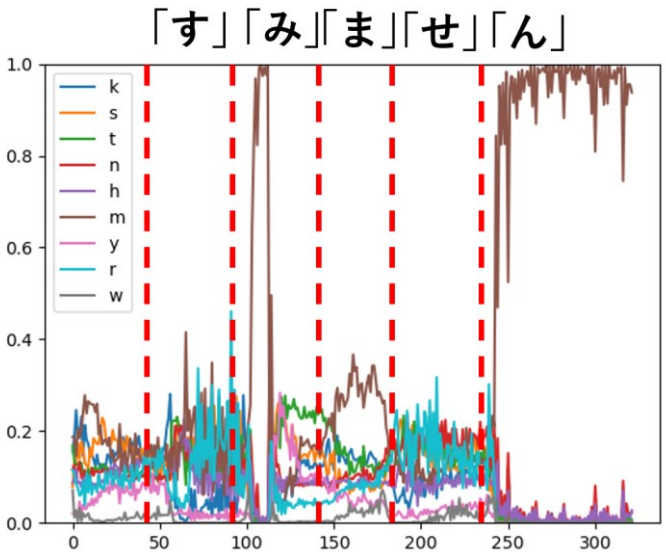

図1.「すみません」の子音認識結果

4. 考察

図1のように「ま」など唇を使って調音する子音は認識可能であることが分かった．また「ん」を「ま行」として認識していることから，子音認識部は口を閉じている動作を認識していることが分かる．そのため子音認識においては口を閉じる動作を含む「ま行」，「ば行」，「ぱ行」，「ん」の有無を認識できる可能性を秘めている．

母音認識部の認識精度は高い．しかしコーパスと比較するために母音列へ変換する際に，誤変換が多く起こりコーパスとの比較で違うものが選ばれていた．このため文字列推定部の推定精度は低かった．

5. まとめ

母音認識部における母音列への変換での誤変換が多いため，変換の際の精度向上が必要である．

子音認識部においては「ま行」，「ば行」，「ぱ行」，「ん」の有無を認識できる可能性を秘めているので，文字列推定部で反映させるシステムを作成する．

参考文献

[1]間瀬健二,アレックス ペンドランド:”オプティカルフローを用いた読唇”,テレビジョン学会技術報告会 ITEJ Technical Report,vol.12,no.44,pp.7-12,(1989).

[2]齊藤剛史,小西亮介:”トラジェクトリ特徴量に基づく単語読唇”,電子情報通信学会論文誌 vol.J90-D no.4,pp.1105 - 1114,(2007).