

音声信号の解析による発声器官パラメータの推定

横田敏明 * 砂川祐樹 ** M.R.Alsharif***

*琉球大学理工学研究科 **琉球大学工学部情報工学科 ***琉球大学情報工学科教授

1 はじめに

通常、人が人の声を認識する際、空気の振動を感知して得た音声信号とその抑揚、視覚から得られる話者の唇の動きや仕草などから、総合的に意味を認識する。また、機械による音声認識においては、話者の声帯や口腔内の器管から発せられた信号のみを処理し、発音を解析する。本文では、音声認識の際に人が得る情報に視覚情報を追加することの有用性と、雑音などが含まれる環境において、人が音声認識する際に視覚情報がある場合と無い場合の認識率の変化について述べ、その事実をどのように工学分野へ応用できるかを提案する。

2 背景と目的

アニメーション製作において、キャラクターの唇の動きと声優の音声の同期は非常に手間のかかる作業であるが、近年では音声信号を元に自動的にアニメーションを生成する手法がとられている。唇のアニメーションと音声同期すれば、視聴者はキャラクターのセリフを理解しやすくなり、作品の質も向上する。アニメーションの製作現場に限らず、雑音の含まれる環境において、人は話者の唇の動きなどから総合的に判断して意味を読み取ることができることが知られている。本研究の目的は、音声信号から口の動きを自動生成する手法の提案としている。

機械による人の発音を解析する方法として、隠れマルコフモデルとよばれる確率的状態遷移モデルが用いられている。連続音声認識の手法は、対象とする問題にどのようにアプローチするかを分類することができる。言語的な知識を用いずに、音響的解析によって音声を認識する方法である連続単語音声認識に基づく解析手法を用いれば、唇の動きと音声信号の同期が可能である。また、唇の動きを同期させるだけならば、音響的な解析のみを用いて主観的に発声器官の状態を推定することができる。

3 提案手法

音声信号から口の動きを生成する過程において、隠れマルコフモデルを用いて話者の口の動きを推定する手法と、音響的な解析によって口の状態を推定する手法を提案する。音響的な解析とは、確率モデルに頼らずに各器管の状態を推定することを指すものとする。

3.1 処理の流れ

処理の流れを図1に示す。

入力される信号は音声信号とし、出力は各器管の状態をあらわすパラメータとする。まず、入力信号の発音を解析する。その発音をもとに、各器管の状態を推定し、状態変数のコードを生成する。これらは時間によって変化するパラメトリックコードである。このコードをもとにアニメーションを生成する手法をとる。

3.2 パラメータの推定

人が声を発するとき、発声器官の状態を関数として、声帯から発せられた振動や肺から吹き出される空気が関数に入力され、体外へ声として放たれる。ここで述べる発声器官とは、歯や舌、鼻腔なども含むものとする。例えば、サ行の子音部分の発音は声帯が振動する必要はなく、肺から送り込まれた空気が歯の隙間を通ることによって乱流が起

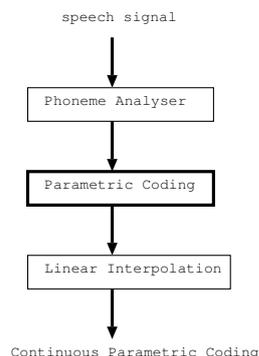


図1: アニメーション自動生成プロセス

こり、比較的高い周波数の振動が起こるものである。視認可能な話者の口の動きは、唇・歯・舌などであり、それらの状態を音声から推定できれば、音声から口の動きをある程度再現できる。すなわち、

- 唇の幅と高さ
- 上側の歯と下側の歯の位置
- 舌の形状と位置

等のパラメータを推定すればよい。ここでは口の動きのみを扱うが、歯の動きは顎の動きと連動しているため、歯の動きから顎の動きを再現することも可能である。

3.3 パラメータの連続的な変化

会話などでは発音が連続的に変化する。意図的に断続的な発音をすることもできるが、口の動きは必ず連続的に変化がおこる。そのため、発音をコード化した後に、モーションの線形補間を行う必要が生ずる。無音状態において、話者の口の動きは未来の発音に依存する。このような状況では動きを推定することが極めて困難であるが、分音声認識のように、言語的知識を用いて予測することはある程度可能である。

4 今後の課題

隠れマルコフモデルによる解析方法を検討しているが、唇の動きなどのパラメータ推定の際には、音響的な解析のみでもある程度可能であることから、そのようなアプローチの有用性も十分検討に値すると思われる。よって、今後は音響的な解析と、確率モデルを組み合わせたハイブリッド方式などを実装し、検証したい。現在は、入力された信号をバッチ処理した後にモーションを表示させる方法についての研究を行っているが、入力に対してリアルタイムでモーションを表示させる方法も検討したい。

参考文献

- [1] 森北出版「音声情報処理」古井貞熙著
- [2] John Coleman “Introducing Speech and Language Processing”
- [3] Lawrence R. Rabiner “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition” IEEE