

佐治 裕一朗 Pitoyo Hartono 橋本 周司 澤田 秀之*

早稲田大学理工学部応用物理学科 香川大学工学部知能機械システム工学科*

1 はじめに

人や動物の音声は、肺、気道、声帯、声道、舌やそれらを動かす筋肉などの複雑な働きによって作られる。本論文では、それらの発声器官の代わりにエアポンプ、人工声帯、声道モデルなどを用いて音声を生成する試みについて報告する。

我々は、今までに物理モデルによって生成される音声のピッチや大きさを制御することで、ハミングの生成に成功している[1]。ここでは、その音色を制御することが目的である。

このような機械的に音声合成を行う試みはいくつか報告されているが[2,3]、人間の発話動作が十分に解明されていないため、その制御は試行錯誤的なものが多い。

本モデルは、入力された音響のスペクトルを模擬した音響を生成することを目的としている。そのために、音声から発声動作を逆推定する処理にニューラルネットワーク(以下 NN)を用いた。これにより、個別の声道の形状と生成される音響との対応関係を学習させて、望む音響を生成することが可能である。

2 音声生成システムの概要

2.1 音声生成システムの構成

音声生成システムの構成を図1に示す。このシステムは、エアポンプ、人工声帯、声道モデル、サウンドレコーダ・FFTアナライザからなる。

エアポンプから送られた空気流が人工声帯を振動させることで、原音(音源波)が生成され、声道モデルの形状により共鳴特性を加えることによって調音される。

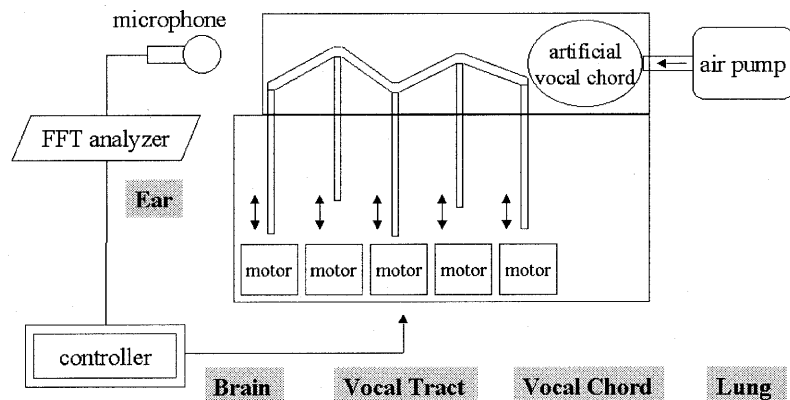


図1 音声生成システムの構成

また、聴覚部において音響の周波数特徴により、生成した音響の評価を行う。

2.2 人工声帯

ここでは、人工声帯の1つである口中笛を用いた。現在は、エアポンプから出される空気量、人工声帯のゴムの張力を一定としているので、生成される音響の大きさやピッチは一定である。まず、成人男性の平均的な基本周波数(約120Hz)を参考に、エアポンプの空気量、人工声帯のゴムの張力をそれぞれ手で調節した。

2.3 声道モデル

声道モデルを図2に示す。主な材質はジュラルミン樹脂、アクリル樹脂、ゴム状のプラスチックである。

声道部分は厚さ10mmの箱型で、内径の寸法は長さ(声道の方向)175mm、高さ25mmとなっている。また、この箱の下側面に声道変形部を挿入した。声帯側から声道の方向に5mm、39mm、73mm、107mm、141mmの5ヶ所で、真鍮棒を抜き差しすることによって、断面積が約1.3cm²から約8.8cm²のあいだで自由に変形を加えることが可能である。

2.4 聴覚部

生成された音響に以下の条件によるFFT解析を行うことにより、そのスペクトルを求める。

サンプリング周波数 ; 11.025[kHz]
量子化ビット ; 8[bit]
FFTのデータ数 ; 1,024
窓関数 ; ハミング窓

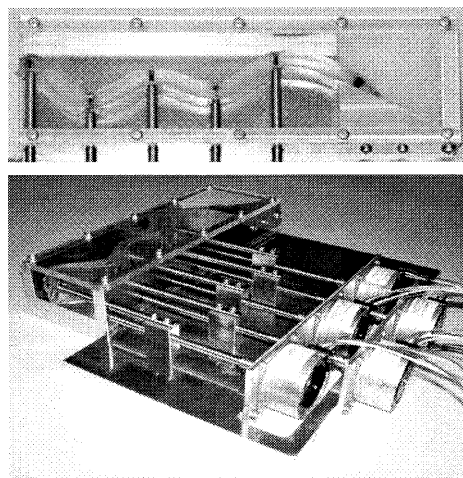


図2 声道モデル

3 声道モデルの制御

3.1 ニューラルネットワークの導入

声道断面積関数を推定する手法としては、PARCOR 分析法がよく知られている。しかし、PARCOR 分析法は 1 次元モデルであり、声道断面の 2 次元形状の影響は考慮していない。

そこで、NN を用い、ある音響とそのときの声道の形状との対応関係を学習させることで、本声道モデルに特化した声道断面積関数を導き出すを試みる。

3.2 学習法

学習時と音響生成時における NN の働きを図 3 に示す。

学習時において、NN はシステムが生成した音響の音響パラメータを入力とし、そのときの制御パラメータを教師信号とすることで、ある音響を生成するために必要な声道の形状を学習する。

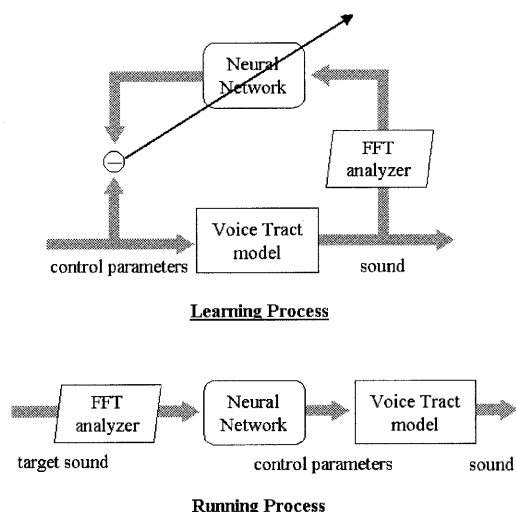


図 3 NN の動作

学習後、NN を声道モデルに対し直列につなぐ。NN に生成したい音響の音響パラメータを入力し、出力された制御パラメータから求められる声道の形状に変化させることによって、望む音響を生成することが可能となる。

4 実験

本論文で用いた NN は、3 層からなる多層型パーセプトロンであり、学習にはバックプロパゲーション法を用いた。NN の入力には音響パラメータ(15 ユニット)、出力には制御パラメータ(10 ユニット)を用い、中間層のユニット数を 50 とした。ここでいう音響パラメータとは、音響スペクトルの周波数領域(およそ 3kHz まで)を 15 等分してそれぞれの成分の平均パワーとし[0, 1]で規格化したもの、制御パラメータとは、5 ヶ所それぞれの真鍮棒の抜き差しを加減 4 段階を 2bit の 2 進数に対応させて二

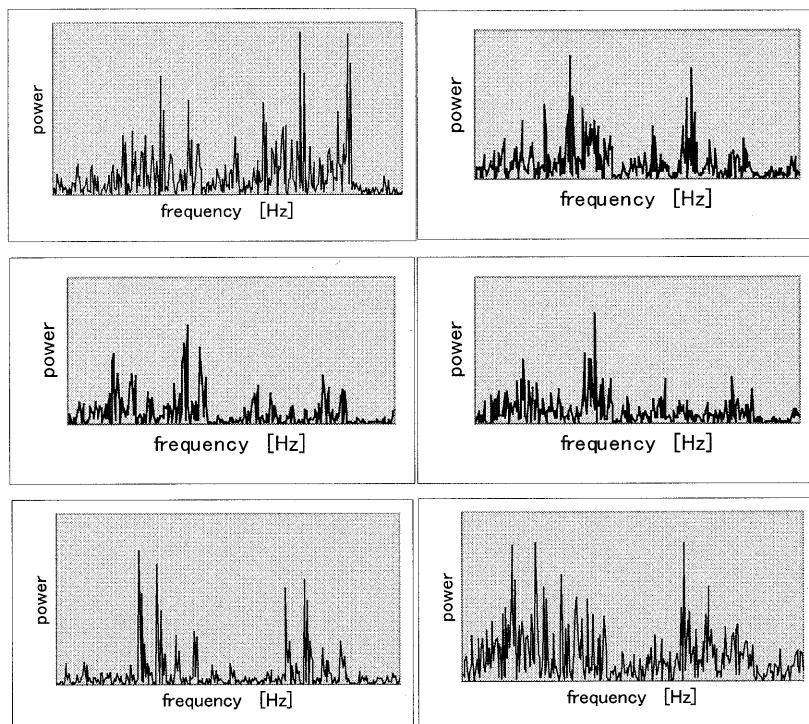


図 4 実音(左)と合成音(右)のスペクトルの比較
(上段「あ」;中段「う」;下段「え」)

値化したものである。

評価として、人の音声 5 母音を目的の音響とし、本システムによって生成させた。図 4 に人の母音「あ」、「う」、「え」に対する実音と生成音とのスペクトルを示す。

図 4 から、スペクトルにおいてある程度の一致が見られ、モデルが学習によって声道断面積関数を習得していることが分かる。誤差の低減のためには、制御点数の増加、聴覚フィードバックによる適応的な制御などが必要と思われる。また、より自然な音声を生成させるためにも、声道モデルの構成やその材質、声帯についての改良が必要である。

5 おわりに

人の発声器官を物理的に構築し、その制御に NN を用いたシステムについて報告した。また、その解析から本システムの有用性を示した。

現在、音響の大きさやピッチの高さなどを含む制御、外乱などに対して安定した音声を生成させるための音声のフィードバック機構を検討している。

参考文献

1. Hideyuki Sawada, Shuji Hashimoto: "Mechanical hideyuki construction of a human vocal system for singing voice production", Advanced Robotics, Vol.13, No.7, pp.667-661 (2000)
2. 大須賀公一・荒木祐之・澤田謙次・小野敏郎:「機械式音声合成装置の実現に向けて-第 1 報:構音器官の三次元形状の再現-」, 日本ロボット学会誌, Vol.16, No.2, pp.189-194 (1998)
3. Kazufumi Nishikawa, Kouichirou Asama, Kouki Hayashi, Hideaki Takanobu and Atsuo Takanishi: "Development of a Talking Robot", Proceedings of the 5th Seminar on Speech Production, pp.345-348 (2000)