

計算機制御された力学系による歌声生成の基礎研究

5 Z-5

荷本 優子 澤田 秀之 橋本 周司

早稲田大学 理工学部

1. はじめに

計算機を用いた音声合成はアルゴリズムによる手法が主流となっているが、その合成音は自然な歌声の生成に適当なものとは言えない。一方、人間の発声機構を物理的に構成することにより、より人間らしい音声を生産できると考えられる[1][2][3]。また、このような機械系に聴覚フィードバックをつけて計算機によりダイナミカルに制御することによって、人間が発声技術を獲得したり、或いは練習によって声色をまねる過程をシミュレートすることも可能であろう[4]。

我々は現在、計算機制御された力学系による歌声生成システムの製作を試みている。音声生成に必要なピッチ、ホルマント等を生成するためのパラメータと機械系の動作を対応付け、計算機で実時間制御することにより自然な歌声を生成することが目的である。本報告では、声帯部に人工声帯を用いて声の高さ及び音量を適応的に制御することにより、ハミングの歌声を生成するシステムについて述べる。

2. システム概要

2-1 人間の発声機構

人間の発声においては、肺からの呼気流によって声帯が振動し音源が生成される。更にこの声帯音源波に対して声道が音響フィルタの役割を果たすことによって、音素が構成される。このフィルタは声道内壁の形状などによって伝達特性が決まるが、主として、その非定常的な変化から子音が生成され、定常的な形状からは母音が生成される。一方、声の高さを決める基本周波数（ピッチ）及び声の大きさは、声帯音源波が持っている情報であるが、これらは肺

からの呼気流量や声帯の形状と弾性などによって調節されている[5]。

2-2 発声システムの構成

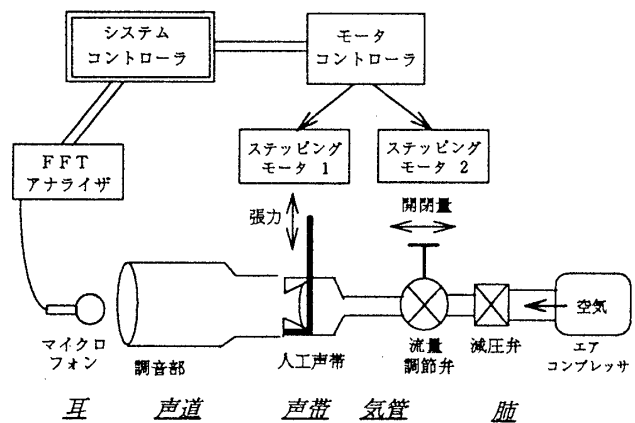


図1 システムの構成

本システムの発声部は図1に示すように、エアコンプレッサ、流量調節弁、人工声帯、調音部、マイクロフォン、FFTアナライザから構成されている。これらはそれぞれ、人間の発声器官の肺、気管、声帯、声道、耳、聴覚系に対応する。人工声帯には、咽頭摘出者が疑似声帯として用いるタピア笛を使用している。タピア笛は図2に示すように構成がシンプルで、ゴムの張力調整によって比較的簡単に振動特性が変えられるという利点もある。図3は人工声帯による音源波形の一例である。コンプレッサから送出される空気は、減圧弁を通して、外気圧に対して約+0.2気圧に減圧される。これはほぼ肺からの呼気圧に相当する。調音部からの生成音は、

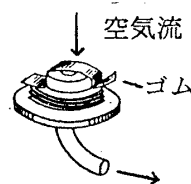


図2 人工声帯

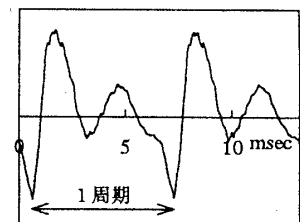


図3 声帯音源波形

A Singing Voice Generation Using Computer-Controlled Mechanical Vocal System

Yuko Kamoto, Hideyuki Sawada, Shuji Hashimoto

Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo

F F Tアナライザによって周波数解析され、随時システムコントローラにフィードバックされる。

2-3 音程制御システム

声帯音源のピッチ及び音量を変化させるために、モータ1（回転角0.9 deg/pulseの2相ステップモータ）、モータ2（回転角0.36 deg/pulseの5相ステップモータ）をそれぞれ人工声帯ゴムの張力及び、流量調節弁の開閉量の調節用として用いた。2つのモータは、モータコントローラからのパルスによって操作量が与えられる。図4は、ゴムの張力を0gfから0.32gfまで5段階に設定したときの、弁の開量と声帯音源の基本周波数の関係を示すものである。なおこれらの関係は必ずしも再現性の良いものではないため、音声生成過程では聴覚フィードバックによる調整が必要となってくる。図4で判るように、2つのモータの操作量によって声帯の振動数が約130Hzから320Hzまで変化しており、通常的人工声帯でも1オクターブ以上のダイナミックレンジが得られる。

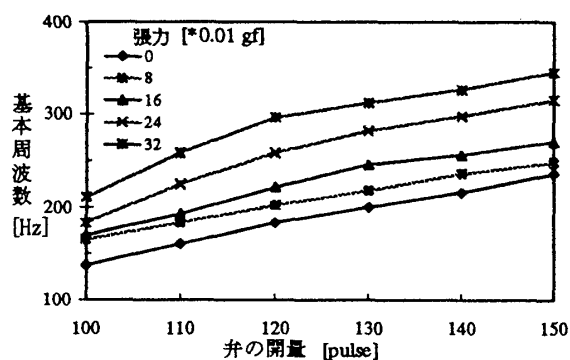


図4 弁の開量と基本周波数の関係

3. 歌声生成アルゴリズム

3-1 ピッチの学習

人間が歌の練習においてピッチを習得して行く過程では、自分の出しているピッチを耳で聞き、目標となる理想のピッチと比較することによって誤差をなくすように学習して行く。本アルゴリズムは、この過程を模倣することによってピッチの学習を行う。あらかじめシステムコントローラには、目標とな

る音名c音からh音までのピッチの基本周波数値を与えておく。まず適当な張力及び弁の開度で発声させる。F F Tアナライザは音声波形の周波数スペクトルを算出し、基本周波数値をシステムコントローラに転送する。システムコントローラは、図4の関係を定性的に用いて発声音と目標のピッチとの誤差をなくすように操作量の演算を行い、モータコントローラに送る。このプロセスを繰り返すことにより、c音から順に楽音ピッチの学習を行い、音名とモータ操作量のマップを作成する。

3-2 メロディー生成

システムコントローラには、楽譜情報が音名及び音長の並びからなる時系列データとして与えられている。前節で得られたマップに基づいて、音名に対応するモータ操作量が順次モータコントローラに送出され、メロディーが生成される。なお出力音声の基本周波数値は常にF F Tアナライザによって監視され、演奏中にもピッチのずれは適応的に修正されている。

4. おわりに

人工声帯を用いて歌声を生成するシステムと、その基礎研究の結果について述べた。今後は調音部の制御を行い、音声のホルマントを生成することによって、人間のように話し歌うことのできる新しいアコースティックインストルメントへ拡張する予定である。

参考文献

- [1] J.L.Flanagan, "Speech Analysis Synthesis and Perception", Springer-Verlag, 1972
- [2] 井澤他、「機械系制御による音声合成」、第3回ロボットシンポジウム、1993
- [3] 大須賀他、「機械式音声合成器について」、第13回日本ロボット学会学術講演会、平成7年
- [4] F.H.Guenther, "A neural network model of speech acquisition and motor equivalent speech production", Biol. Cybern. pp.43-53, 1994
- [5] 林 義雄、「こえとことばの科学」、第5版、昭和56年、鳳鳴堂書店