音楽情報科学 16－2
（1996．7．27）

声道モデルの機械系による実現とその計算機制御

澤田秀之*、橋本周司

早稲田大学理工学部応用物理学科
〒169　東京都新宿区大久保3－4－1
TEL: (03)5286-3233
*日本学術振興会特別研究員
e-mail: sawa@shalab.phys.waseda.ac.jp　shuji@shalab.phys.waseda.ac.jp

あらまし：声道モデルを機械系によって構成することにより、人間の発声機構を再現する試みを行っている。これによってより人間らしい音声の生成が可能になるばかりでなく、“歌う”楽器を作ることができると考えている。本モデルは主に、肺、声帯、声道部、聴覚部から成り、現在はピッチを変化させて、ハミングによる歌声を生成することが可能となっている。ピッチの生成には複雑な声帯振動の解析が必要であるが、ここでは声帯摘出者が用いる人工声帯を使用している。本報告では機械系による声道モデルを紹介し、ピッチの学習アルゴリズムと聴覚フィードバックによる適応制御について述べる。

A Mechanical Model of Vocal Tract and its Adaptive Control

Hideyuki Sawada*　and　Shuji Hashimoto

Department of Applied Physics, School of Science and Engineering, WASEDA University
3-4-1, Okubo, Shinjuku-ku, Tokyo, 169, Japan
* Research Fellow of the Japan Society for the Promotion of Science
e-mail: sawa@shalab.phys.waseda.ac.jp　shuji@shalab.phys.waseda.ac.jp

Abstract: We are constructing a phonetic machine having a vocal chord and a vocal tract based on mechatronics technology, and have so far developed a pitch generation part as a subsystem for melody synthesis to sing in humming. In the pitch generation, the analysis and mechanical simulation of the behavior of the vocal chords are required. The fluid mechanical system is, however, less stable to make the control difficult. This paper presents a singing instrument system together with an adaptive tuning algorithm of the physical mechanism using an auditory feedback. The mechanical method is considered to be promising to generate more natural voice than algorithmic sound synthesis methods.

1．はじめに

人間の音声は、音声生成器官の複雑な働きによって作られる音である。発声器官は主に、肺、気道、声帯、声道、舌、口蓋とそれらを動かす筋肉などから成り、互いに適当な位置や形状を形成することにより言葉が生成される[1][2]。声は大抵の動物にあるが、言葉は人間にしかないコミュニケーション手段であり、音声生成のメカニズムや音声の認識手法などが古くから研究されてきた。特に最近のマンマシンインタフェース技術には、人間らしい音声による情報の提示や、音声による入力デバイスは不可欠な

要素となっている。

音声合成において、現在は計算機を使ったアルゴリズミックな音声合成が主流となっている[3][4][5]。しかしそれらの多くは波形レベルでの音声生成であるために、歌声のように自然性が重視されるような音声には適した方法とはいえない。一方で、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成できると考えられる[2]。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。人間の各器官を機械系により再現し、特定の発声行動を適応的に学習していく機構を解析することにより、音楽工学への応用が可能であるばかりではなく、ロボティクスや医療、福祉工学への適用も期待される。

我々は現在、声道モデルを機械系によって構成することにより、音声生成システムの製作を試みている。発声においては、声帯の振動から音源波が作られ、声道の共鳴管を駆動することによってホルマントが形成される。発声に必要な器官を声道モデルとして構成し、計算機で実時間制御して自然な歌声を生成することが目的である。そのためには、音声生成に必要なピッチ、ホルマント等を形成するための操作を、実際に出力される音声と対応付けることが必要になる。現在までに、ピッチを学習して計算機制御によって変化させる機構を構成し、ハミングによる歌声を生成することが可能となった。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、歌声を生成するシステムについて述べる。

2．人間の発声機構

人間の発声は大きく分けて、声帯振動による音源の生成と共鳴によるホルマントの付加という2つの働きによって構成されている。肺からの呼気流は気道を通って声帯の振動を引き起こし、音源を生成する。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。このフィルタの伝達特性は声道内壁及び舌の形状などによって決まるが、主として顎や舌の非定常な動作によって引き起こされる変化から子音が生成され、母音は定常的な声道形状を形成することによって生成される。更に器官組織の湿り気や粘性なども生成される音質に大きな影響を与えるが、これは風邪を引いて声が嗄れるといったときに経験することからもわかる。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。

音声の生成においては各組織の複雑な働きが必要であり、このような能力は幼児が言語を習得していく過程において発声と聞き取りの試行錯誤を繰り返すことによって獲得していくものである。

3．機械系による声道モデル

3－1．人工声帯

音声のピッチと大きさは声帯音源波の特性によってが決まるが、音源波は声帯の複雑な動きによって生成されている。人間の声帯は2枚の筋肉とそれを覆う粘膜から成り、その間を通り抜ける肺からの呼気流によって振動が引き起こされる。2枚の粘膜の擦れ合う音によって有声音源が生成されるが、その基本周波数は主に筋肉の緊張度によって決まり、また声の大きさは呼気流量によって変えることができる。このような声帯の運動を計算機によりシミュレーションする試みも幾つかなされている[6]が、ここではタピア式笛を用いて機械系モデルを構成し、声帯音源の生成を行った。タピア式笛は声帯摘出者が疑似声帯として用いる人工声帯であるが、図1に示すように構成がシンプルな上に、ゴムの張力調整によって比較的簡単に振動特性が変

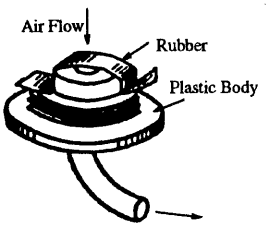

図1　人工声帯
Figure 1　An Artificial Vocal Chord

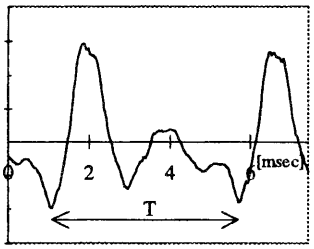

図２　人工声帯による音源波形
Figure 2　Sound Wave of Artificial Vocal Chord

えられるという利点もある。図２は人工声帯による音源波形の一例である。声帯音源波は、孤立三角波の繰り返しとして近似されることが多く、本人工声帯による波形は比較的実物に近いものとなっている。

人工声帯による音源を、長さLのゴム片の振動現象から得られる音波と考えると、生成される音の基本周波数fは、

$$f=\frac{1}{2L}\sqrt{\frac{S}{D}} \quad . \qquad (1)$$

と表される。この式は、基本周波数がゴムに加えられる張力S及び材質の密度Dによって変化することを示している。ここではゴム片に張力を作用させることによって基本周波数を変化させることを試みた。図３に、空気の流量を１０～１４[*l/min*]に固定したときの、張力とピッチの変化の関係の一例を示す。張力を変化させることによって、基本周波数が約１３０Ｈｚから３２０Ｈｚまで変化しており、１オクターブ以上のダイナミックレンジを持つことがわかった。また、ここで得られた関係は、式（１）に示すように張力Ｓの平方根に比例するものとはなっていないが、これは張力を加えることによってゴムの密度が変化してしまうことに起因すると考えられる。さらに、この結果は必ずしも繰り返し再現性の良いものではなく、また空気流量の変化にともない、生成される音源のピッチも不安定なものとなっている。しかし人工声帯は構造が単純であり、張力を作用させることによって比較的簡単に基本周波数を操作することが可能であるという利点を持っている。

３－２．調音部の構成

調音部には、長さ５００ｍｍ、径３０ｍｍの塩化ビニル製のパイプを用いている。生成される声質を柔らかくするために内外にウレタン材を張り付けており、このパイプは自由に曲げたり潰したりの変形をすることができる。これにより、外部から変形力を与えることによって声道断面積や共鳴部の形状が変化することになり、ホルマントを付加することができる。

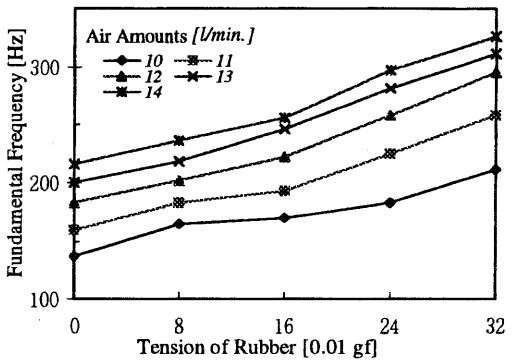

図３　ゴムの張力と基本周波数の関係
Figure 3　Relation between Tensile force and Fundamental Frequency

本システムにおいては、人工声帯出力端からの距離x_jの位置に、パイプの上方向から力$P_j(x_j)$を加えて、変形を行っている。

３－３．音声生成システムの構成

音声生成システムの構成を図４に示す。このシステムは、主にエアコンプレッサ、流量調節弁、人工声帯、調音部、マイクロホン、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。

人間の肺からの呼気圧は、外気圧に対して約＋０．２気圧である。エアコンプレッサ内の空気は約５気圧まで圧縮されるため、減圧弁を通して気道内の呼気圧程度まで減圧している。またこの減圧弁は、空気圧縮時に発生するコンプレッサ内の空気の脈動の低減にも有効である。減圧された空気は流量調節弁を通して人工声帯へ送られる。さらに調音部によってハミングの音質を変えることが可能となっている。

ハミングを生成するために必要なピッチ生成と音量調節のために、２つのモータを用いている。モータ１は人工声帯ゴムの張力操作用であり、モータ２は流量調節弁の開閉量調節用としている。ここではフィードバックモータを位置制御に用いて、モータコントローラからの指令

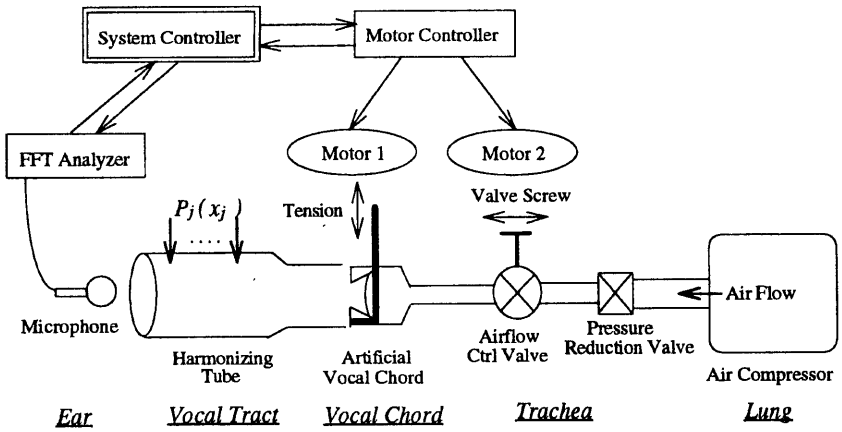

図4　音声生成システムの構成
Figure 4　System Configuration

によってそれぞれ張力の操作及び弁の開閉度の操作を行うことができる。一方、ＦＦＴアナライザは人間の聴覚部に相当し聴覚フィードバックに必要な機構であるが、声道モデル部により生成される音声に実時間フーリエ変換を施すことにより、基本周波数とスペクトラム包絡を抽出する。

システムコントローラは、ＧＰ－ＩＢを通してモータコントローラ及びＦＦＴアナライザとデータ通信を行うことにより、音声生成システムの管理を行っている。

4．歌声生成及びピッチ制御

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。本システムでは、聴覚フィードバックをつけてこの過程を模倣することによって適応的にピッチの学習を行う。ここで構成した機械系は、人工声帯に取り付けているゴムの振動によって音源を生成しているため、得られるピッチは必ずしも再現性の良いものではない。また、ある基本周波数の音声を保持する場合にも、空気流量の変動や張力のわずかな変化に対してもピッチが変動してしまう。このような外乱に対して安定した音声を生成するためにも、フィードバックアルゴリズムは不可欠であると考えられる。

歌声は、楽譜の音名とそのピッチを与えるモータ位置の対応関係を記述しているマップに基づいて生成される。このマップは、ピッチ学習モードにおいて作成される。楽譜情報は、音名と音長の並びとして記述されており、あらかじめシステムコントローラ内に与えられている。演奏モードでは、楽譜情報に基づいてマップを参照しながらモータコントローラに位置データを送出して歌声が生成されるが、出力は常にＦＦＴアナライザによって監視されており、ピッチのずれは適応的に修正される。

なお今回は調音部の制御は行っておらず、3ヶ所の適当な固定位置x_jに$P_j(x_j)$ (j =1,2,3) を与えることによりホルマントをつけている。

4－1．学習モードにおけるピッチの学習

ピッチ学習時における、システムコントローラ、声道モデル、ＦＦＴアナライザの働きを、アルゴリズムの流れとともに図5に示す。学習モードでは、点線内に示した各処理ユニットがシステムコントローラ内に生成され、ピッチの獲得をおこなう。

チューニングマネージャは、全ての処理ユニットの管理を受け持っており、目標のピッチと出力音声のピッチとの誤差を無くすようモータ操作量の演算を行いながら、ピッチの学習を進めていく。学習結果は、ピッチ－モータ対応マップとしてシステムコントローラ内に保存さ

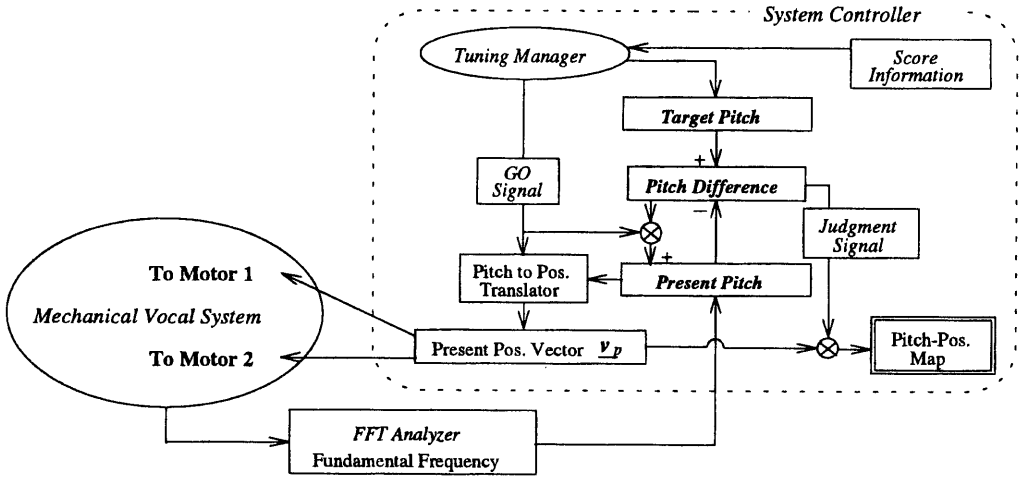

図5　ピッチ学習アルゴリズム
Figure 5　Diagram of Adaptive Tuning

れる。マップ生成の準備として、チューニングマネージャはまず楽譜情報ユニットを参照し、楽曲に含まれる全ての音名を抽出する。これらはピッチの低い順にソートされピッチの学習を行っていき、マップには音名と2つのモータ位置の組が記述されていく。

まずチューニングマネージャがモータ位置ベクトル $\underline{v}_p = (p_1, p_2)$ に任意の値をセットし、モータコントローラに送出することにより学習を開始する。同時に、ターゲットピッチユニットには目標となるピッチ値を設定する。ここで位置ベクトルの要素 p_1, p_2 は、それぞれモータ1、モータ2の位置である。

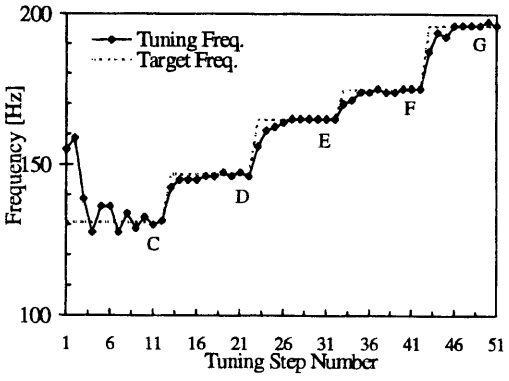

図6　ピッチの学習結果
Figure 6　Result of Pitch Tuning

モータコントローラは、受け取った位置ベクトルに従って各モータを移動するとともに、FFTアナライザは出力音声のスペクトルから基本周波数を算出し、出力ピッチユニットにセットする。ピッチ差分算出ユニットは常にターゲットピッチとの差分を検出しているが、マネージャからのGO信号を受け取るとその差分値を出力ピッチユニットに転送し、新しいモータ位置ベクトル $\underline{v}_p$ が算出される。出力ピッチからモータ位置ベクトルを算出しているのはピッチ－モータ位置変換ユニットであるが、その演算には図3の関係を定性的に用いている。以上のプロセスを繰り返すことにより、ピッチ差分がある閾値よりも小さくなったときに差分判定信号がアクティブとなり、目標の音名に対応するモータ位置としてマップ内に記述される。現在、閾値は2Hzに設定している。これを全ての音名に対して行うことにより、ピッチ－モータ対応マップが得られる。

このフィードバックプロセスによって得られたピッチの例を図6に示す。

4－2．演奏モードにおけるメロディ生成とピッチ制御

メロディ生成の流れと、ピッチ制御のダイヤグラムを図7に示す。聴覚ユニット内のFFTアナライザ及び、声道モデル以外の各ユニット

は、システムコントローラ内に生成されるものである。ここでは、演奏マネージャが全ての処理ユニットの管理を受け持っており、それは図中に太線で示したメロディ生成プロセスと、演奏中のピッチの適応制御の２つに分けられる。

演奏マネージャはメトロノームに対応するクロックを内部に持っており、まず楽譜情報ユニットの音長情報を参照して演奏テンポのプランニングを行う。このプランに従って演奏マネージャはＧＯ信号を送出し、音名に対応したモータ位置情報がモータコントローラに送られ歌声が生成されていく。

演奏中には、空気圧の変動やゴムの張力の微妙な変化などが原因となって、出力音声のピッチが変動してしまう。そこで、聴覚フィードバックによるピッチの適応制御が必要になる。聴覚ユニットはＦＦＴアナライザによって出力音声のピッチを算出し、楽譜音名とのズレを常に監視している。演奏マネージャは、補正が必要と判断した場合にチューニング信号を出力し、システムはピッチの補正を行う。同時に、マップも補正後のモータ位置に更新される。

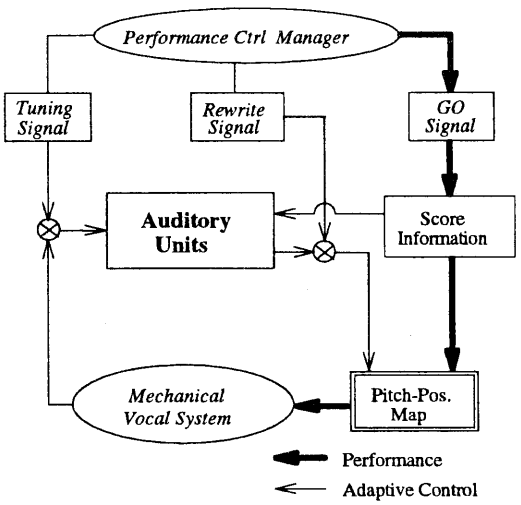

図７　メロディ生成とピッチ制御の流れ
Figure 7　Adaptive Control of Singing Performance

ここで構成した機械系による声道モデルは、外乱に対して不安定なものであるが、このようなアルゴリズムによって、ロバストに歌声を生成することが可能となった。

５．おわりに

声道モデルを機械系によって構成し、音声を生成するシステムについて報告した。音源に人工声帯を用い聴覚フィードバックを付けることによって、ピッチを獲得して、ハミングによって歌声を生成することが可能となった。

現在は内部クロックによってメトロノームのようにテンポ一定で出力しているが、今後は人間の歌のようにテンポにおける表現を付加することも必要であろう。更に声道部に適応制御機構を付加して調音部を構成することによって、声色を変えたり言葉を学習していく過程がシミュレーションできると考えている。また、声質を更に人間に近づけるために、声帯の構造や振動機構を詳細に解析して再現すること、声道共鳴官の材質を改良することなどを検討中である。

人間の発声器官を機械系により実現し、発声行動の学習機構を解析することにより、新しい楽器の構築、さらには新しいマンマシンインタフェース構築への足掛かりとしていきたい。

[参考文献]

[1]　林義雄, "こえとことばの科学, 1979

[2]　J.L.Flanagan, "Speech Analysis Synthesis and Perception", Springer-Verlag, 1972

[3]　広瀬啓吉, "音声合成の研究の現状と将来", 日本音響学会誌, ４８巻１号, pp.39-45, 1992

[4]　X.Rodet and G.Benett, "Synthesis of the Singing Voice, Current Directions in Computer Music Research, PIT Press, 1989

[5]　Ph.Depalle, G.Garcia and X.Rodet, "A Virtual Castrato", Proc.ICMC, pp357-360, 1994

[6]　K.Ishizaka and J.Flanagan, "Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords", Bell Syst. Tech. J., 50, 1223-1268, 1972