研究概要報告書

研究題目	声道物理モデルの機械系による構築と聴覚フィードバックに基づく発話獲得に関する研究	報告書作成者	澤田 秀之
研究従事者	澤田 秀之、東本 敏男		
研究目的	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。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。特に最近のヒューマン・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が不可欠な要素となっている。 人間の発声機構は大きく分けて、声帯振動による音源の生成と、声道部共鳴によるホルマントの付加という二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。音声の生成においては各発声器官の複雑な働きが必要であり、このような能力は幼児が生まれてから数年のあいだに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されるものである。また音声獲得後であっても、人間は練習によって声まねをすることができる。これは、発声のために必要な器官が生まれながらに備わっている一方で、発声技術は後天的に獲得して行くものであるためと考えられる。 本研究では、発話動作をおこなう器官をエアポンプ、人工声帯、声道共鳴管などを用いて機械的に構成し、計算機による聴覚フィードバック制御によって自らが音声を獲得、生成することができる発話ロボットの開発をおこなう。機械システムが人間のように音声から発話動作を適応的に学習し、再現することにより、人間どうしのように発話動作をともなった音声コミュニケーションの実現が可能となるであろう。また、新しい音声エンターテインメントの提案やアミューズメントへの利用も可能であると考えている。 従来の音声合成では、サンプリング法、規則合成、物理モデルといったソフトウェアによる手法が主流となっている。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、物理的な発話動作から得られるような自然性が重視される、歌声のような音声には適した方法とはいえない。一方で、機械的に音声生成を行う試みがこれまでにいくつか報告されている。しかし人間の発話動作が充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。 申請者はこれまで、人間の発声器官の肺、気管、声帯、声道、耳、聴覚に対応する機能をそれぞれ機械系によって構成することにより、音声の生成を試みてきた。調節弁のスクリュー、人工声帯の振動部のマニピュレーション、および共鳴管変形のためのピストンの操作用に、それぞれモータを用いた。機械系から生成される音声をマイクロフォンから計算機に入力し、音響解析部で解析を行う聴覚フィードバック機構により、人間が目標となる音声を与えるだけで、自己学習によって音声を獲得できる可能性をひらいた。		

研究内容	声道物理モデルの構築と聴覚フィードバックに基づく発話獲得に関する研究について、以下の 4 点について検討をおこなった。 **1) 機械式声道物理モデルの構築** 人間の発声器官である肺、気管、声帯、声道、鼻腔、耳、聴覚に対応する機能を、それぞれエアポンプ、空気圧･流量調節弁、人工声帯、調音用共鳴管、鼻腔、マイクロフォン、音響解析器によって構築した(図 1)。人間の発話においては、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、一方の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。空気流量調整や人工声帯、共鳴管などの機械系各部の操作にモータを用い、計算機により制御をおこなう。特に音質を決定する声帯と声道部について、材質、形状および制御点数の検討をおこない、人体皮膚程度の柔度をもつシリコンゴムと骨密度に近い石膏を用いて整形した(図 2)。また磁気共鳴画像(MRI)を基に、子音を明瞭に発音させるための口唇の形状及び鼻腔構造を決定し、発話動作を可能とする制御機構を実現した(図 3)。 **2) 聴覚フィードバック制御による発話動作の学習と適応制御の確立** 機械式声道モデルから生成される音声をマイクロフォンから入力し、音響解析部で解析を行う聴覚フィードバック機構により、人間が目標となる音声を与えることにより、機械系が自ら音声および発話動作を獲得できる学習機構を実現した。聴覚部として、音声からの実時間ピッチ、ホルマント抽出アルゴリズムを実装し、これらの音声特徴とモータ制御量の関係を適応的に学習できるニューラルネットワーク(NN)を構築した(図 4)。更に、NN 学習に基づく聴覚フィードバック制御による発話動作の学習機構を実現した。 自己学習過程で、機械系の音声が目標として与えた人間の音声に近づいていく過程の解析を進め、目標音声と獲得された音声の特徴についての比較、検討をおこなった。また人間が入力する音声から、NN の学習によってシステムが適応的に発話動作を推定、再現できることから、声真似を実現した。 **3) 歌声の生成** 声帯および声道形状を適応的に制御することにより、歌声の生成を実現した。あらかじめ、楽譜に含まれる音高、音長、歌詞の抽出をおこなう。まず歌声学習モードでは、楽譜中の音高および歌詞を発声するための、各モータ制御量の獲得と関連付けをおこなう。歌唱モードでは、楽譜を参照しながら適応的にモータの制御をおこない、歌声を生成する。歌唱中には、聴覚フィードバックにより音程のズレが適応的に修正される。これにより、人間と同様な発話動作に基づく発声と歌声の生成が可能な歌唱ロボットが実現できた。 **4) 聴覚障碍、発話障碍患者のための発声訓練システムの構築** 聴覚障碍、発声障碍患者の発声訓練は、発話中の口内形状が見えないことから、言語聴覚士が一人一人の患者に対して症状に応じた発声訓練をおこなっている。本発話ロボットは、人間の入力音声から機械系が発話動作を推定し、音声を再現できる。聴覚障碍者や発話障碍者が、見本となる発話動作や口内の動きを見ながら、対話的に発話訓練をすることができるシステムへの応用が期待できる。

研究概要報告書

研究のポイント	本研究では、発話動作をおこなう器官をエアポンプ、人工声帯、声道共鳴管などを用いて全て機械系によって構成し、計算機による聴覚フィードバック制御によって自らが音声を獲得、生成することができる発話ロボットの開発をおこなってきた。人間と同等の器官を物理的に構成することにより、人間の様に学習によって人間らしい発話動作および音声を獲得することができる適応制御機構を実現しようとするところ、更にここで得られる学習結果をもとに人間の発話学習機構を解明しようとするところが、本研究の独創的な点である。更に人間の音声学習機構との比較、考察を行っていくことにより、人間の脳における情報の認知･生成メカニズムの解明に貢献できると考えている。またソフトウェア合成アルゴリズムでは困難であった自然な歌声の生成を、物理的な発話動作によって実現しようとする点も独創的である。 一方、発声障碍、聴覚障碍患者の発声訓練は現在、ST(言語聴覚士)が一人一人の患者に対して症状に応じた発声訓練をおこなっている。発話ロボットを用いた訓練装置の実現によって、新しい対話型発話訓練手法を提案しようとする点が、独創的かつ新規な点である。
研究結果	人間と同等の発話器官を機械モデルによって構成し、可動部をモータにより制御することにより、人間の様に聴覚フィードバック学習によって発話動作および音声を獲得することができることを示した。これまでに 5 母音、鼻音、破裂音、一部の摩擦音の発声を実現した。更に聴覚フィードバック学習とピッチ制御により、歌声の生成を可能とした。 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、人間どうしの特に論理的な対話においては、音声が重要な役割を果たしている。人間との共存、対話を可能とする人間型ロボットの研究がさかんに進められているが、それらの多くは形態、動作の模倣に関するものであり、音声生成は専らソフトウェアによる合成音声が用いられている。本研究成果によって、人間と同様な発話動作および発声、さらに声真似を用いてコミュニケーションできるロボットの実現を示すことができたと考えている。更にこれらの成果は、音声を通したヒューマンインタフェースデバイス構築のための要素技術の蓄積にも貢献できると考えている。
今後の課題	今後は歯を付加することにより、サ行の摩擦音の生成を実現していく。また柔軟に形状変化をする舌を付加することにより、より明瞭な母音･子音の発声をめざしていく。更に聴覚フィードバック制御による発話動作の学習機構を確立する。自己学習過程で機械系の音声が目標として与えた人間の音声に近づいていく過程を解析し、目標音声と獲得された音声の違い、学習によって獲得できない人間の音声特徴などについて詳細な検討を進め、人間の音声獲得機構の解明につなげていく。 また音声の特徴から、その音声を生成するのに必要な発話動作を再現できることから、聴覚障碍者や発話障碍者のための発話訓練システムの構築が可能となる。高齢社会の到来に伴い、加齢による発話障碍や疾病による発声障碍、脳性麻痺患者の増加が見込まれることは想像に難くない。福祉や高齢者支援の視点から、音声の矯正、発声訓練システムの開発、新しい音声アミューズメントへの応用、提案を積極的に進めていきたい。

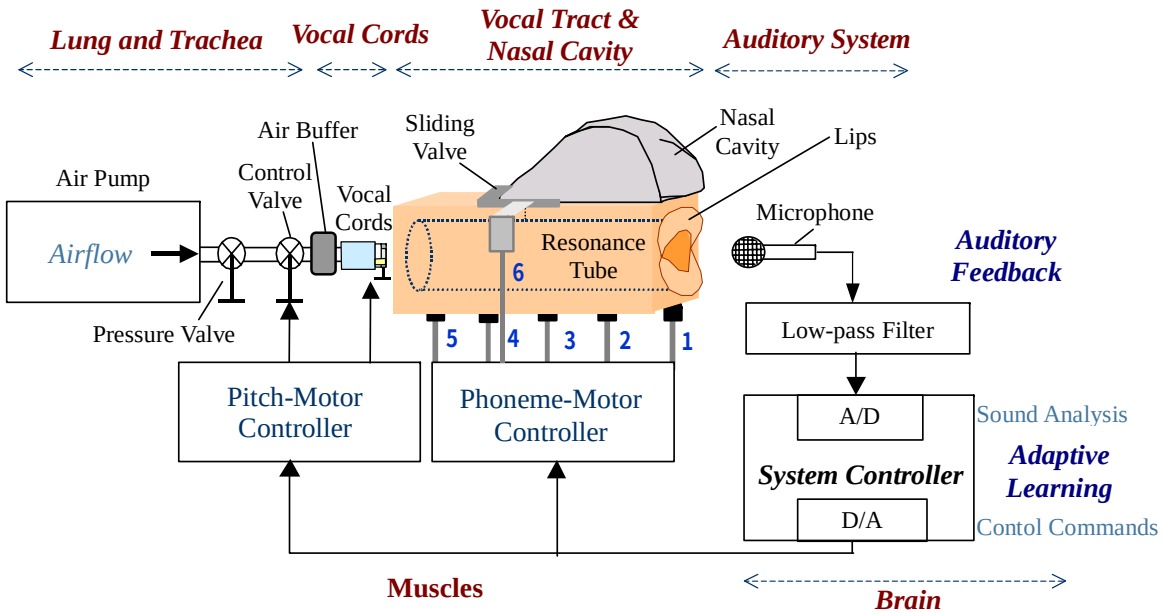


図 1 声道物理モデルの構成図

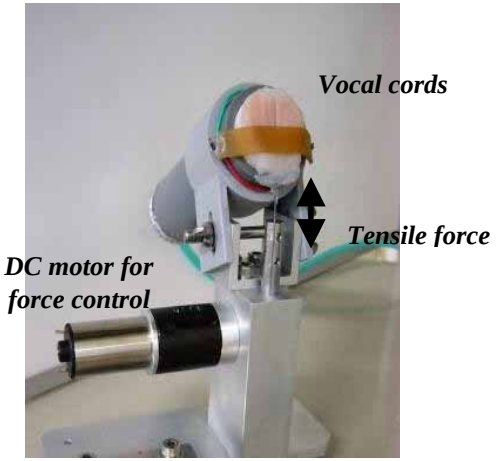
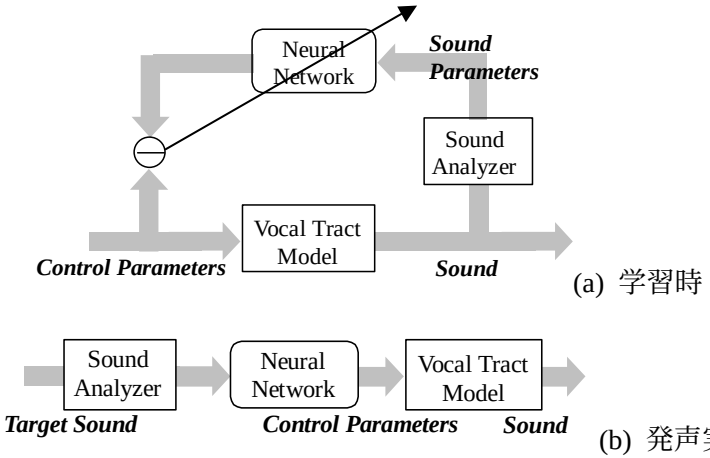


図 2 声帯部

図 3 声道部



(a) 学習時

(b) 発声実行時

図 4 ニューラルネットワークの動作