

自然音声合成に関する研究

— 発声メカニズムの解析 —



大阪大学工学部

助教授 工学博士

森田 龍 彌

1. はじめに

われわれは言葉を日常あまりに自由に使いこなしているから、発声の難しさに気付かないが、それは鍵盤の上を走るピアニストの指の動きにも例えられるだろう。われわれが幼時から知らず知らずのうちに身につけた発声法は容易に修正できない程に固定されている。大阪人の大阪弁は一生抜けないし、外国語を流暢に使いこなすことも至難の技である。

音声は、声帯の振動が声道（口腔、鼻腔）で共振、反共振を受けて口唇より放射されたもので、声帯による声の高さ（ピッチ）、下顎や舌等による音韻、口唇や歯による放射特性などの要素が関与している。言語を伝える目的からみれば音韻が重要に違いないが、言語らしい滑らかさや感情の動きにはアクセントやイントネーションが不可欠で、これは声帯における声の高さに由来する。

また、母音は定常音、子音は過渡音であるといわれるが、母音を発声する場合でも、アの発声中に次のイを発声するための指令が舌や下顎に送られ、運動を開始し始める。母音の区間でも大部分はアでもイでもない過渡音であるといつてよい。このような過渡的な発声では下顎、舌、口唇など各調音器官の運動と声帯振動とのタイミングが極めて重要な要素であるといえる。

ピアニストの指の動きは驚異的ではあるが、ロボットに置換えることは可能かも知れない。演奏内容については楽譜をプログラムしておけばよい。また、鍵盤にセンサを取り付けて打鍵

の強さやタイミングを計ることもできる。これに対し発声機構を機械化することは容易ではない。楽譜に相当する神経系からの制御指令は計り知ることができない。密閉された口腔内にある舌の運動は外部から計測することが困難である。音声は、人と人とのコミュニケーションを行うためばかりでなく、人間-機械間のインターフェイスとしても益々重要になると予想され、発声機構を解明することは発声障害者の補助代行システムや音声応答システムの開発に寄与するところが大きい。そこには極めて困難な障害があるといわざるを得ない。

発声器官の計測にはさまざまな方法があるが、まだ決定的なものはない。ビデオカメラなどを用いて口唇の動きを計測するコンピュータを用いた読唇術も、これだけでは十分な情報は得られない。核磁気共鳴像(MRI)を用いれば、口唇から咽喉部まで発声器官の断層像を立体画像として表示できるが、撮像に数十秒もの時間を要し過渡的運動を計測できる状況ではない。舌の表面に圧電薄膜電極を接着して舌の曲率変化を計測する方法は連続計測できるが、舌の先端部4～5cm程度の部分に限られる。したがって、さまざまな知見を総合的に解釈する以外今のところ方法がないといえる。そのような部分情報としては発声器官の運動に伴う組織の電気インピーダンス変化を計測する方法も連続計測が可能で有効と考えられる。また、発声された音声を分析して、声道変化の動特性を推定する方法も考えられる。以下にはこの2つの研究経過に

ついてまとめる。

2. 頸部インピーダンスによる音声ピッチの推定

筋が収縮した場合、長さが短縮し断面積が増加するためインピーダンスが小さくなる。頸部インピーダンス計は、この原理を喉頭部の筋に適用したものである。声帯は輪状甲状筋などの筋群によって制御されているが、頸部の深い所にあるため、そのインピーダンスを計測するのは難しい。しかし、下咽頭収縮筋は図1のように甲状軟骨が前下方に傾く(音声ピッチの上昇)と伸張し、インピーダンスが増加すると考えられるので、図2のように頸部の皮膚に4枚の電極を装着し、外側2枚の電極より一定振幅の交

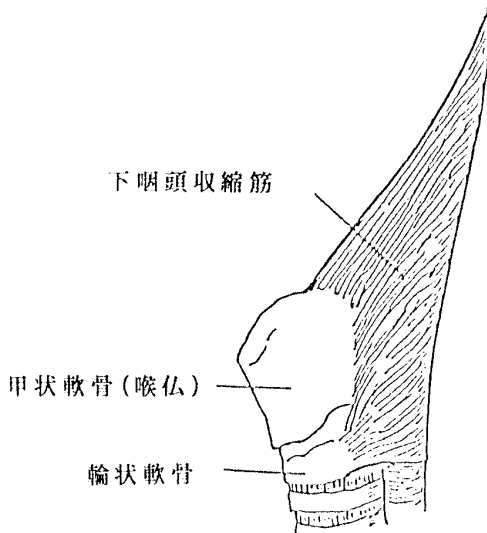


図1. 下咽頭収縮筋の位置

流電流を流し、内側2枚間の電位差を検出してインピーダンスの変化を計測した。また、同時にマイクで採取した音声から音声ピッチ変化を線形予測分析法により求め、頸部インピーダンスとの関係を調べた。

被験者(男子学生4名)にはある基準音から高音階へ上昇する場合と、高音階から基準音に下降する場合について繰り返し発声させ解析した。

分析の結果、音程変化前後の定常部のピッチのオクターブ差とインピーダンス計出力の差の間にはほぼ線形性が認められた。一方時間関係については、インピーダンスの方がピッチ変化より遅れて発生する傾向が見られたが、さらに詳細にみると、発声者によっては急速なピッチ変化の後、音程を補正するようにピッチが2段階的に変化する場合があり、この場合インピーダンス変化は緩やかなピッチ変化とほぼ同位

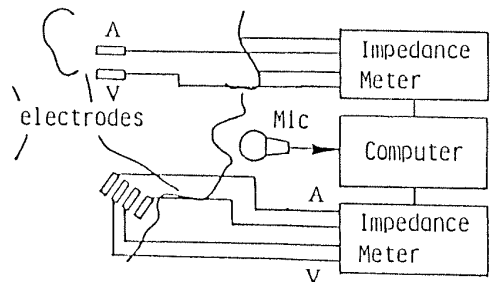


図2. 頸部、顔間インピーダンス同時計測システム
A : 電流端子 V : 電圧端子

相で発生する。また、ピッチ下降時にはピッチ上昇時よりも速やかに変化することができるが、インピーダンス変化もこれと対応した特性をもつことが判明した。以上から頸部インピーダンスは音声ピッチ変化を反映しているといえるが、ピッチより遅れることから、その発生源は声帯を直接制御する筋群ではなく、甲状軟骨の動きによって受動的に動かされる下咽頭収縮筋である可能性が高いと考えている。

3. 調音パラメータ動特性の最小2乗推定

線形予測分析法の開発は音声分析の発展のきっかけとなったが、この分析法により、声道特性を表す PARCOR 係数と声道を駆動している残差信号（声帯振動に対応）が推定できるようになった。またこの PARCOR 係数は、声道の音響反射係数に対応し、舌や下顎などの発声器官の運動に密接に関係している。したがって、音声分析によってもある程度、発声器官の運動が推定できるといえる。

発声においては慣性の大きい下顎によって声道の概形が形成され、舌によって細かい調整が行われると考えられる。したがって、この2器官の特性を反映した緩徐、急速成分が PARCOR 係数の変化に含まれていると考えて図3のようなモデルを作成し、力学パラメータとこれを駆動する神経系からの制御指令を推定した。すなわち、モデルの出力と音声分析から得られた PARCOR 係数との2乗誤差が最小となるように力学特性、制御指令のタイミング、大きさなどの未知パラメータを修正した。

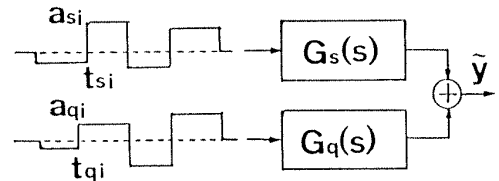


図3. 調音パラメータ動特性推定モデル

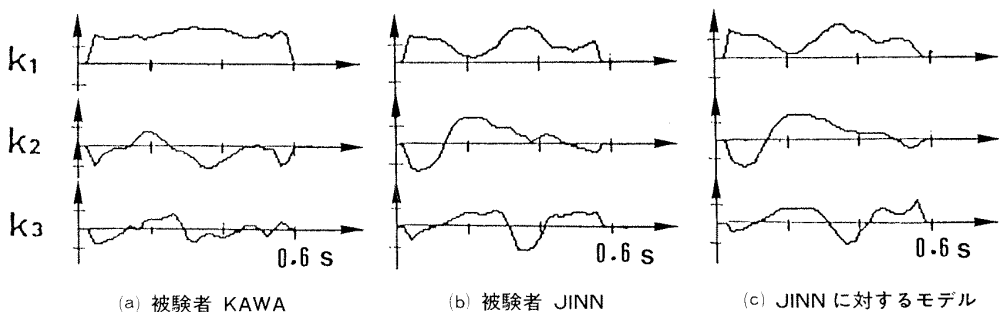


図4. PARCOR 係数の個人差とモデルの応答
第1～3係数のみを示す。

PARCOR 係数は同じ連続母音でも、**図 4** に示すように発声者により異なったパターンを示し、個人差が大きい。分析された力学特性にはほとんど差異がなく、PARCOR 係数パターン上の個人差は主として、神経系による発声器官制御のタイミングと大きさによるものであることが示唆された。また、逆に合成した音声は元の発声者の声と極めてよく類似しているため、解析結果はほぼ正しいと考えている。

謝 辞

末筆ながら、本研究の遂行に助成戴きましたサウンド技術振興財団と共同研究者、大阪電気通信大学工学部松村雅史助教授に深謝致します。

発表論文

1. 森田、松村：調音パラメータ動特性の最小自乗推定、システム制御情報学会論文誌、Vol.4、No.10、(1991.10)
2. 森田、松村：頸部電気インピーダンスを用いた音声ピッチの推定、電子情報通信学会論文誌、A (採録決定)

