音声生成過程に基づく音声合成システムの開発*

藤井心平　緒方公一　園田頼信（熊本大学）

1. はじめに

本研究室では，音声生成時に重要な役割を担っている舌，口唇，下顎などの調音器官の挙動の把握に関する研究を行っており，その研究成果を基に音声生成過程に基づく音声合成のシステム開発を行っている．本稿では，人間の音声生成過程を模擬した調音モデルとして Sondhi と Schroeter により提案された Hybrid 型の声道シミュレータ[1]を計算機上に構築し，子音を含んだ連続音の合成実験を行う．

2. 音声合成モデル

音声の生成過程を模擬し，音声を人工的に合成するものに声道シミュレータがある．本稿では，SondhiとSchroeter により提案された声道シミュレータを用いる．このシミュレータは，Fig.1に示すように，1）声帯モデル，2）声道モデル，3）口唇放射モデルの3つのモデルから構成されている．声帯モデルは，Ishizaka･Flanagan の2質量モデル[2]を用いて時間領域で表す．声道・口唇放射モデルは，周波数領域で表わすHybrid 型の構成になっている．実際の声道・口唇が持つ損失は，周波数領域で声道モデルに組み込まれている．

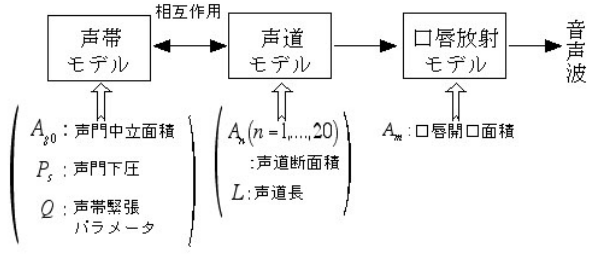


Fig.1 Configuration of speech synthesis system.

実際の声帯と声道・口唇の間には相互作用が存在する．このシミュレータでは，上の2つの領域をフーリエ変換と離散化畳み込みによって結合することによりこの相互作用を取り入れている．

2.1 声帯モデル － 時間領域モデル

声帯モデルは，Ishizaka･Flanagan によって提案された2質量モデルを用いている．2質量モデルは，位相差を持って振動する声帯上部および下部の運動を記述するために，声帯を等価的にFig.2に示すようなスチフネスk_cにより結合した上下2つの振動子によって表している．各々の振動子は，質量m_i，スチフネスk_i，及び粘性抵抗r_iにより等価的に表されている．ここで，添字i = 1, 2 であり，i= 1 のとき下部振動子，i = 2 のとき上部振動子を表す．正常な声帯は左右対称であり，その振動もまた左右対称であるとみなすことができるのでFig.2では片側のみを示している．

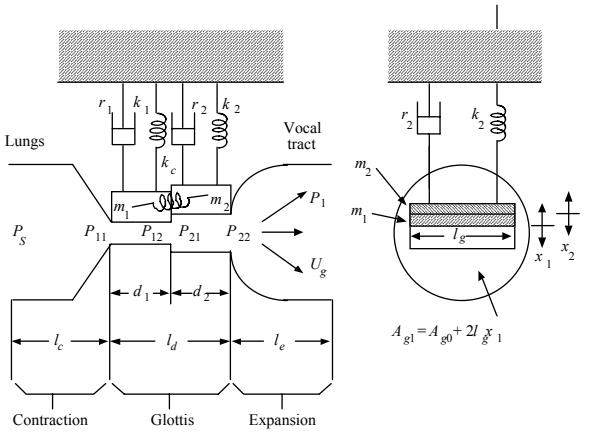


Fig.2 Two-mass model of vocal cords.

---

* Development of a speech synthesis system based on speech production mechanism

By Shinpei Fujii, Kohichi Ogata and Yorinobu Sonoda (Kumamoto University)

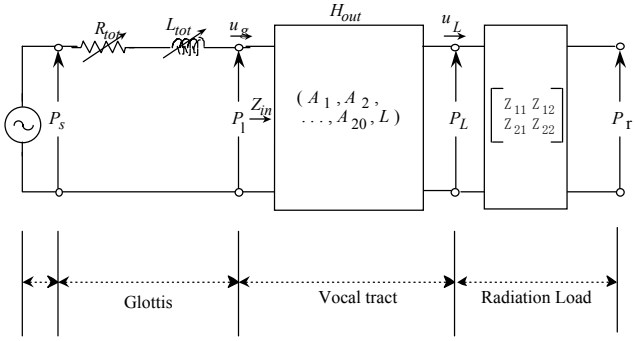


Fig.3 Equivalent circuit for speech production.

2.2 音声合成シミュレータの作成

Fig.3に声道シミュレータの電気的等価回路を示す．声道部分におけるせばめの程度から母音型や子音型のモデルとなる[1]．

母音型では，Fig.3の声道シミュレータの等価回路において成立する式(1)，(2)をサンプル時間 T_s 毎に解き，声門体積流 $u_g(n)$ 及び $p_1(n)$ が求まる．

$$p_1(n) - z_{in}(0)u_g(n) = \sum_{k=1}^{N-1} z_{in}(k)u_g(n-k) \quad (1)$$

$$T_s p_1(n) + denu_g(n) = T_s p_s(n) + L_{tot}u_g(n-1) \quad (2)$$

ここで，$den = T_s R_{tot} + L_{tot}$

得られた声門体積流と，声門から口唇側を見込んだ特性のインパルス応答との畳み込み演算により，音声出力を求めることができる．

摩擦音のような子音生成時には，声道途中のせばめによって雑音源が形成される．その等価回路はFig.4で表され，乱流雑音源の体積流がせばめから口唇側へ供給されることにより音声が生成される．乱流雑音源の体積流は，乱流雑音源の音圧 P_n と内部抵抗 R_n から決定される．乱流雑音源の音圧 P_n はレイノルズ数により決定され，レイノルズ数がある閾値を超えた場合に乱流が発生する．レイノルズ数や内部抵抗の値は，せばめが生じている区間の断面積や体積流に依存し，その体積流は，声門体積流，および声門とせばめにおける体積流伝達比の関係から求めることができる．

破裂音のように声道の閉鎖が生じている場合には音声出力は生じないが，閉鎖した声道を声門から見込んだインピーダンスを基に，声門体積流を求めることができる．

破裂音の生成では閉鎖後の開放により断面積が増加し，過渡的にせばめを形成する状況となる．すなわち，閉鎖終了後は摩擦型のモデルに移行することになり，破裂音の生成が完了する．

Fig.5に本システムにおける音声合成のフローチャートを示す．声道は20個の音響管で表現され，時間的に変化する断面積は次章で述べるように縦続１次系に基づいた調音運動を仮定することで求めている．断面積が最小となる音響管についてFig.5の条件判定を行い，せばめの面積の大きさに応じて，母音型，摩擦型，および閉鎖型に分岐して合成を行っている．

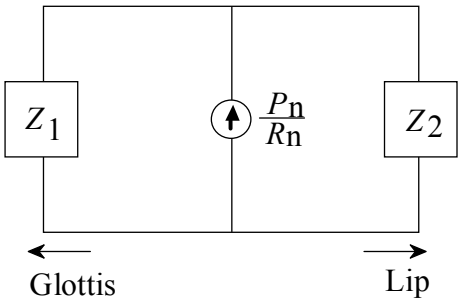


Fig.4 Constriction noise source in the vocal tract.

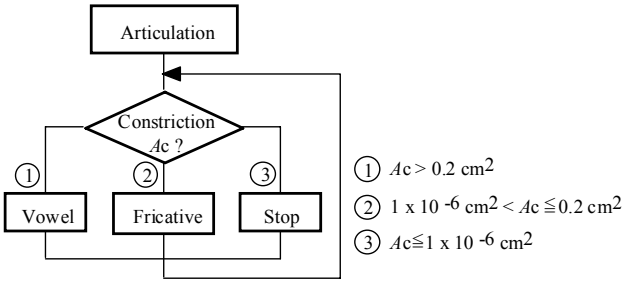


Fig.5 Flowchart of speech synthesis.

3. 縦続１次系を用いた声道形状表現

声道シミュレータを用いて連続音を合成する場合，そのパラメータとして連続な時系列データとしての声道断面積が必要となる．本研究室では，磁気センサを用いてダイナミックな調音運動の計測を行っており，縦続１次系の関数によって調音器官の運動を良好に近似できることが報告されている[4]．本シミュレータでは，声道断面積の時間変化を縦続１次系の関数を用いて表現している．Fig.6は調音運動に伴う声道の断面積変化を模式的に示したものである．この例では，上部が口蓋側に，下部が舌などの器官に相当し，時刻t_1からt_2までの下降運動によって断面積が拡大している様子を示す．この運動が縦続１次系の応答に従うものとして断面積変化を取り扱っている．これまでに単母音および連続母音の合成が可能な音声合成システムの開発を行っており，GUIを活用したインタラクティブなシステムとなっている[5]．

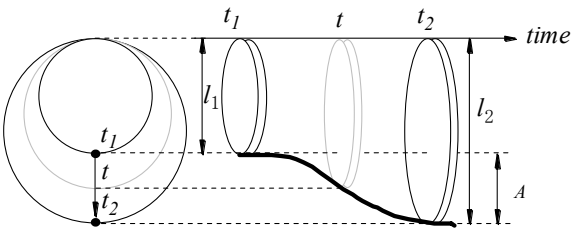


Fig.6 Change in the area of one of acoustic tubes. Increase of the area is described based on the step response of the cascaded first-order systems.

4. 連続音声の合成実験

今回の合成実験では摩擦音を含む連続音/aʃi/と破裂音を含む連続音/etete/の合成を試みた．ここでは連続音/aʃi/について述べる．Fig.7は摩擦子音を含む/aʃi/について，声道断面積の時間変化を示したものである．母音/a/,/i/の断面積については定常母音発話時のMRIデータから求めた断面積を利用し，子音/ʃ/の断面積については，Fantのデータ[3]を参考にした．連続音発話時の断面積変化は３で述べた縦続１次関数を用いて表現されており，滑らかな形状変化が表現されている．Fig.8に声門体積流u_gの時間推移を示す．合成の際には，合成音声の波形が実音声波形に類似するように，声門下圧P_s，声門中立面積A_{g0}等の調整を行った．/a/および語尾/i/の部分では，声門体積流に脈動が見られ，有声音の生成が行われており，中央の/ʃ/においては，体積流が流れ続け無声子音の生成が行われていることを表している．

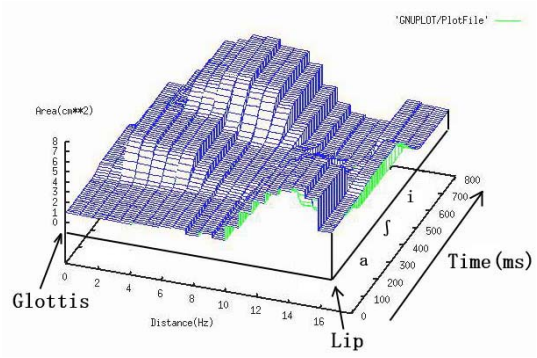


Fig.7 Change in vocal tract shape for the utterance /aʃi/.

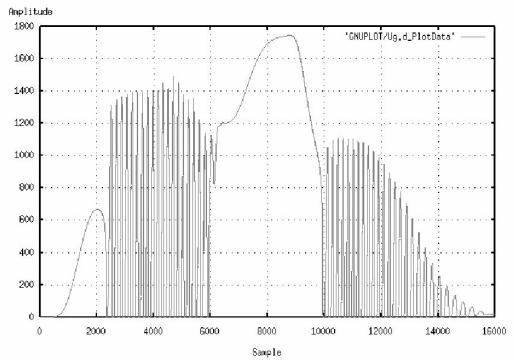
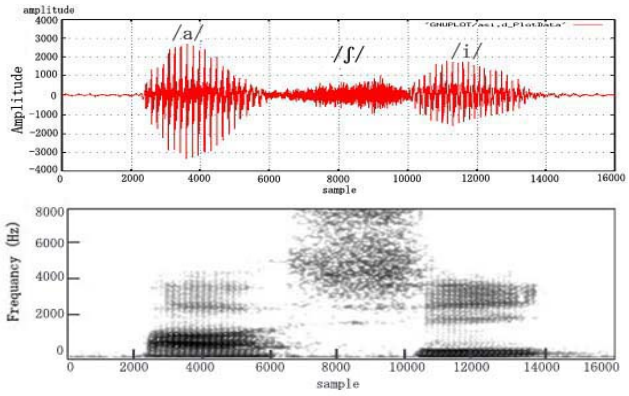


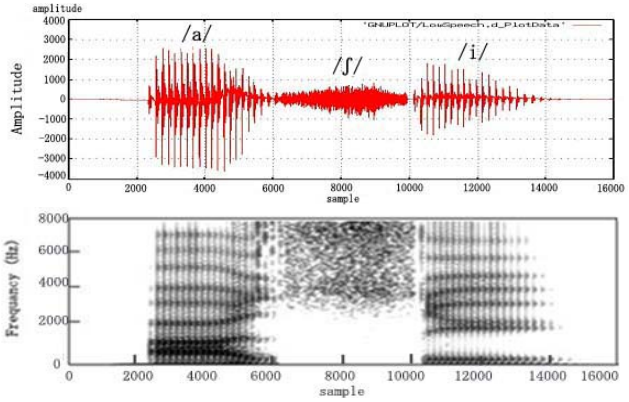
Fig.8 Glottal volume velocity u_g ug for the utterance /aʃi/.

Fig.9に実音声および合成音声それぞれについて音声波形とサウンドスペクトログラムを示す．多少の違いは見られるもの

の，両者の特性は類似したものになっている．



(a) Real speech



(b) Synthetic speech

Fig.9 Speech waveform and its sound spectrogram for the utterance /aʃi/.

5. まとめ

人間の音声生成過程に基づいた音声合成シミュレータを用いて子音を含む連続音の合成を試みた．声道断面積の時間変化は縦続１次系に基づいた調音運動を仮定することで表現した．音質的には改善の余地があるが，得られた合成音の聴覚的印象は比較的良好なものであった．

今後，声帯の振動を伴う有声子音，鼻音などの合成を試みる予定である．また今回使用したプログラムはCUI(Character User Interface)ベースのシステムとなっており，プログラム実行時のパラメータの設定等は，コマンドラインやテキストファイルからの入力により行っている．そこでシステムの操作性の向上のためGUI(Graphical User Interface)を活用した音声合成システムへと改良を進める予定である．

参考文献

[1]M. M. Sondhi and J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer", IEEE Trans. Acoust., Speech & Signal Process., ASSP-35, 7, pp.955-967 (1987).

[2]K.Ishizaka et al., “Synthesis of voiced sounds from a two-mass model of the vocal cords”, Bell Syst.Tech.J., Vol. 51, No.6, pp.1233-1268 (1972).

[3]G.Fant, ”Acoustic theory of speech production” Mouton,TheHague (1970).

[4]緒方公一，園田頼信，“縦続１次系による調音運動のモデル化”，音響学会誌 55，pp.156-164 (1999).

[5]緒方公一，園田頼信，”調音に基づく音声合成システム−GUIを用いたシステムの開発−“，信学技報，SP2002-76, pp.29-34 (2002).