

# 縦続1次系に基づく音声合成システム - 子音の調音タイミングの検討 -\*

外園真実雄, 緒方公一 (熊本大)

## 1 はじめに

著者らは、音響管モデルに基づく音声合成システムの開発を進めており、声道断面積の時間変化に、縦続1次系関数を利用することで滑らかな断面積変化を実現している[1, 2]。特に子音の音声合成では、声道の閉鎖と声帯の振動との時間的相対関係等、時間構造の把握が不可欠であり、シミュレーションを通して合成に関与するパラメータの適切な設定を実現していく必要がある。本稿では、子音の合成のための断面積変化のタイミング等、関連するパラメータの設定を検討するためのGraphical User Interface (GUI)の開発とそれを用いたシミュレーションについて報告する。

## 2 音響管モデルと縦続1次系関数の利用

本音声合成システムは、SondhiとSchroeterにより提案された声道シミュレータ[3]を基に作成した音響管モデルによる合成システムである。声道については、本システムでは20個の円筒状音響管を接続した直管型の音響管として近似的に表現している。部位によりその直径が異なる音響管の形状を表すパラメータである(1)声道長 $L$ , (2)声道断面積 $A_n$ ( $n = 1 \sim 20$ )によって声道形状を制御している。

また、連続音の合成では、簡易に断面積の時間変化を表現する手段として、音響管の直径の変化に縦続1次系関数を適用し、その変化に伴う断面積変化を利用している。これは、子音を含めた調音運動の軌跡を、縦続1次系関数を用いて良好に近似できるという結果を積極的に利用したものである[4]。

Fig. 1に音響管の形状変化への縦続1次系関数の適用の様子を示す。図は、ある一つの音響管の直径が時間の経過とともに大きくなる様子を模式的に示したもので、音響管上部

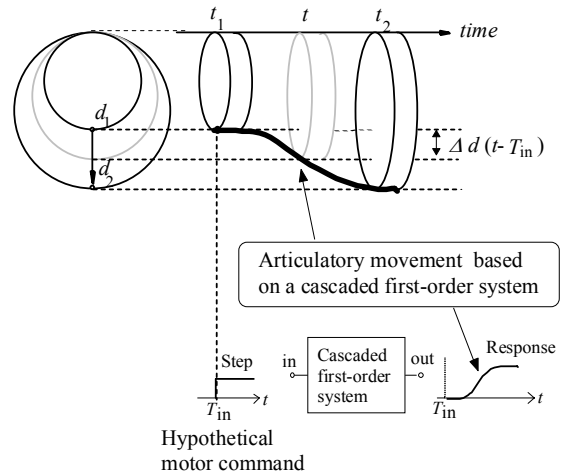


Fig. 1 Change in the area of an acoustic tube as a function of time.

が口蓋側、下部が舌表面にそれぞれ対応するような状況を想定している。時刻 $t_1$ ,  $t_2$ それぞれに対応する音響管の直径を $d_1$ ,  $d_2$ としている。時刻 $t_1$ から $t_2$ にかけて舌が下降し、このときの運動軌跡が、すなわち直径の時間変化が図右側下部に示した縦続1次系関数のステップ応答パターンで表現できるものとして、直径の変化にともなう断面積の変化を求めている。

声帯は2質量モデルで表現されており、著者らのシステムでは、(1)声の大きさに関係する声門下圧 $P_s$ , (2)有声無声の制御に関する声門中立面積 $A_{g0}$ , (3)声の高さに関係する声帯緊張パラメータ $Q$ の3つを可変として、調整できるようにしている。これらのパラメータでも、縦続1次系関数のステップ応答を利用して時間変化を表現し、応答の変化の速さに関する値である時定数は調整可能となっている。

## 3 GUI シミュレーションシステム

Fig. 2にGUIシステムの画面を示す。画面では、左側上部から下部にかけて、20区間のう

\* Speech synthesis system based on cascaded first-order systems – A Study of timing of articulation for consonants –,  
by HOKAZONO, Mamio and OGATA, Kohichi (Kumamoto University).

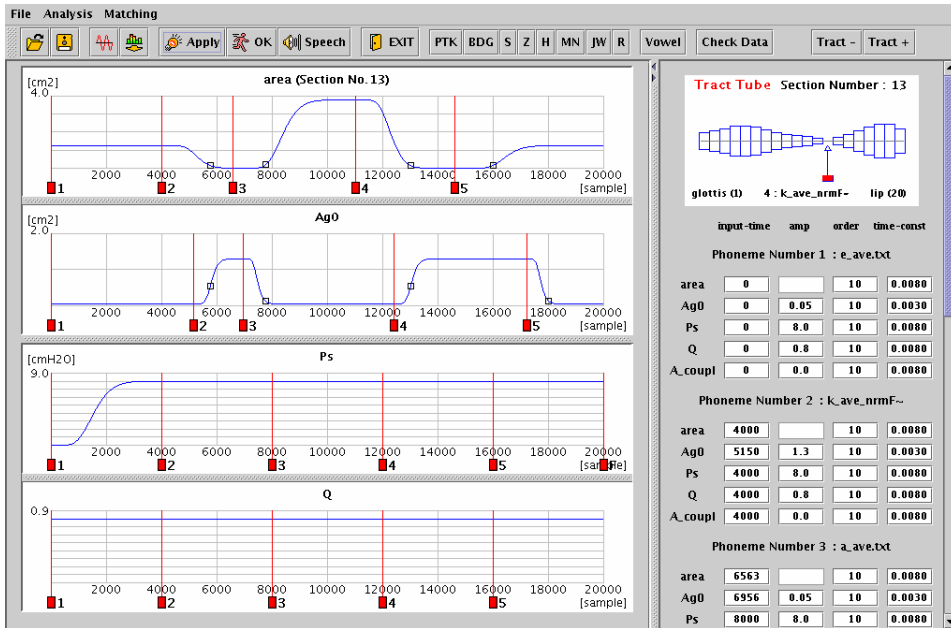


Fig. 2 GUI window for speech synthesis.

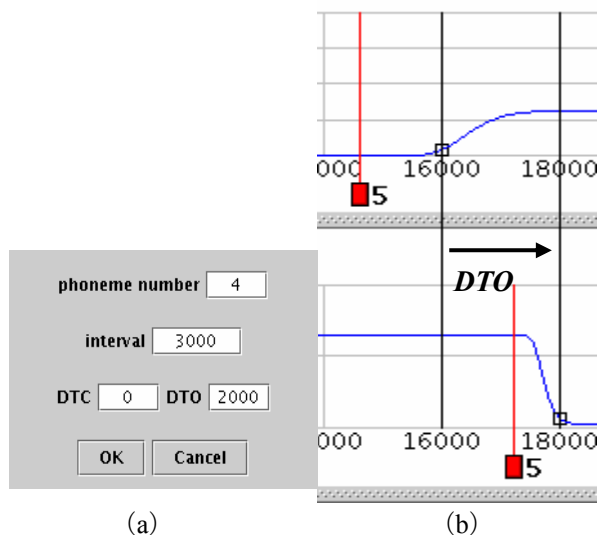


Fig. 3 Parameter setting window and the definition of *DTO*.

ちの選択した一つの音響管の断面積(*area*), 声門中立面積( $A_{g0}$ ), 声門下圧( $P_s$ ), および声帯緊張パラメータ( $Q$ )それぞれの時間変化が表示されている。画面右側の数値入力部では、各種パラメータの設定や変更が可能であり、それらの値が画面左側の時間変化パターンに反映される。画面左側に位置する番号付けされた複数の縦棒(マーカー)は、Fig. 1において断面積変化の原因となる、ステップ信号の入力時刻( $T_{in}$ )に対応するものであり、マウス操作あるいはウィンドウ右側の数値入力によってこれらの設定ができる。パラメータの設定後は、画面上部のアイコンをクリックすることで、合成音の生成やサウンドスペクトログラムの表示等が可能となっている。サンプ

リング周波数は 20 kHzである。

先に開発したプロトタイプシステムでは、入力時刻等のパラメータを設定することが可能であるが、本稿では、タイミングの設定の効果をより定量的に評価できるように、機能の充実を図った。

断面積と声門中立面積のそれぞれの時間変化パターンに基準となる時刻を設け、その時間的相対関係

を種々に設定し、その音響効果を検討するために設けた機能をFig. 3に示す。図(a)の数値入力ウィンドウ下部は、断面積と声門中立面積それぞれに設けた閾値に値が一致する時刻の時間差(*DTC*, *DTO*)を数値入力するためのものである。この設定を時間変化パターンに反映したものが図(b)であり、この場合、それぞれにおける閾値の時刻(グラフ上の□印)の時間差*DTO*が、図(a)で示した値となっている。このように、*DTO*は、断面積が閾値の値をとる時刻を基準として、声門中立面積が閾値をとる時刻で定義されており、*VOT* (voice onset time)[5]に関係する量となっている。なお、断面積(*area*)の閾値は  $0.016 \text{ cm}^2$ としている。この値は、本システムにおいて、せばめの面積の大きさに応じて母音型、子音型のモデルを選択して合成を行うが、その境界の値とほぼ同じ値となっている。また、声門中立面積( $A_{g0}$ )に対する閾値の決定については4で示す。一方、*DTC*は断面積閉鎖と無声化に関する時間差であり、取り扱いはいは*DTO*の場合と同様である。

## 4 シミュレーション実験

### 4.1 音響管の声道断面積の設定

合成には、音素に対応した声道断面積が必要となる。声道断面積は、著者らによって導出された平均的な声道断面積[6,7]を利用した。

### 4.2 声門中立面積( $A_{g0}$ )の閾値決定

声門中立面積( $A_{g0}$ )は、声帯の振動の生成や

停止，すなわち有声無声の制御に関するパラメータである。破裂音から母音に移行する過程では， $A_{g0}$ を大きな値から小さな値へと変化させ，無声から有声への変化を生じさせる必要がある。無声から有声に転じる声門中立面積( $A_{g0}$ )の値，すなわちFig. 3(b)で示した閾値を求めるため，語頭を中性母音/e/とした/eCV/(C = p, t, k, V = i, e, a, o, u)の音形について検討した。

各音形において，声道形状は時間的に変化していくが，子音閉鎖の解放の時刻付近でどのようなタイミングで声帯振動が開始されるかを把握する必要がある。そのため，各子音に対し閉鎖の生じる音響管の断面積(area)が3で述べた閾値  $0.016 \text{ cm}^2$ に達したときの声道形状を求めた。この声道形状を用いて，声門中立面積( $A_{g0}$ )の値をFig. 3(b)下部のように大から小へと変化させ，声帯振動を起こす，すなわち無声から有声に変化する境界値を調べた。有声無声の境界時刻は，サウンドスペクトログラムおよび声門体積流の波形の視察により求めた。

Fig. 4に，上記の音形に対して求められた，閾値を示す。子音/p/では，他に比べて母音による値の違いが小さく， $0.13 \text{ cm}^2$ 程度の値を示している。また，/t/や/k/では，母音の違いによるばらつきは/p/の場合よりも大きい，5母音平均の値としては， $0.14 \text{ cm}^2$ 程度の値となっている。ここでは，各子音について5母音の平均値を求め，その値を閾値として採用することにした。前述のFig. 3(b)では，このようにして求めた $A_{g0}$ の閾値の値を示しており，このときの時刻と，音響管の断面積(area)の閾値に対応する時間差を*DTO*として定義している。

#### 4.3 時間差 *DTO* の設定に関する検討

時間差 *DTO* を種々に設定し，適切な値について検討した。Fig. 5に，/epa/の音形について *DTO* の値を，9, 24, 60 ms としたときの/pa/の

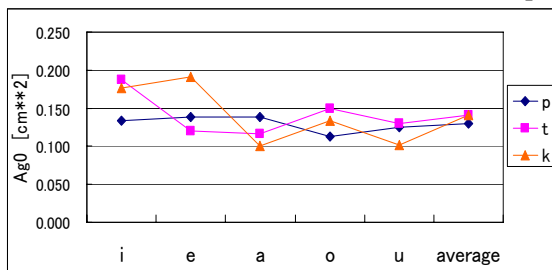


Fig. 4 Threshold for  $A_{g0}$  .

区間におけるサウンドスペクトログラムを示す。*DTO*が増加するに従い，破裂区間が広がり，最終的に破裂区間が終わり，破裂区間と母音区間の間に無音区間が生じるという形態になっている。*DTO*が9 msのときは破裂が生じるか否かの境界であり，この境界での値から *DTO* をどれだけ延長するかによって，破裂の持続時間が決まり，その値を適切に設定すれば，良好な合成音が得られることになる。/epa/, /ete/, /eka/, /epe/, /ete/, /eke/について，*DTO*の区間を延長し，音声波形やサウンドスペクトログラムの視察，聴覚的印象により，その子音として妥当と思われる最大の延長時間を求め，破裂開始からその最大延長時間までの時間区間を導出したものをFig.6に示す。境界での *DTO* は/t/と/k/では同程度の値であるのに対し，/p/では15 ms程大きな値となっている。また，後続母音の違いに対する *DTO* は/p, t, k/それぞれにおいて同程度である。すなわち，/p/は，/t/や/k/と比較して声帯振動の開始が遅れる傾向にあり，今回扱った母音に対しては，後続母音の違いによる変化は小さいものと見られる。したがって，各音形に適切なタイミングを得るには，*DTO*をFig.6に示す区間内に設定すればよい。

Fig. 5において，破裂の時刻におけるエネルギーの分布に着目すると，低い周波数領域にその集中が確認できる。Fig. 7(a)に/eta/の合成音のサウンドスペクトログラムを示すが，

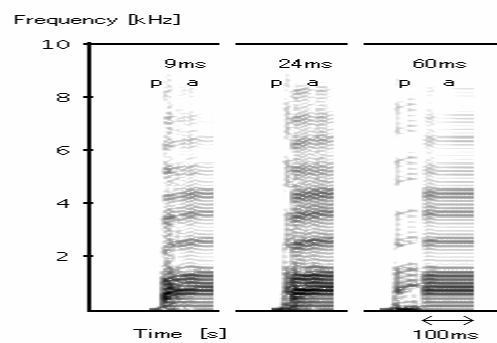


Fig. 5 Sound spectrogram for /epa/.

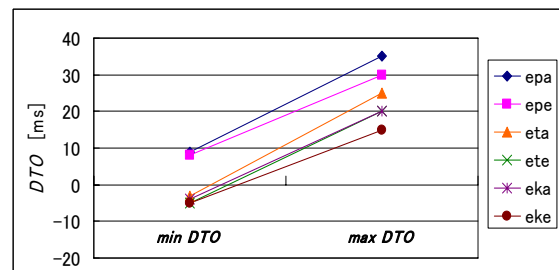


Fig. 6 Time interval for *DTO* .

破裂時におけるエネルギーの集中は高域側に見られ、実音声の分析による破裂音の特徴[5]と一致する傾向である。(b)の/eka/については、高域側に集中しすぎる傾向があり、断面積等の更なる検討が必要と考えられる。

今回、求めた適切な $DTO$ の値を用いて音声を合成するために、システムに新しく自動調整の機能を取り入れた。破裂子音の場合は無音区間となる時間間隔(interval)と声道断面積と正門中立面積の閾値の時間差( $DTC$ ,  $DTO$ )をFig. 3(a)に示すウィンドウで指定することで、3で述べた縦続1次系関数の仮想的入力時刻を自動で調整することが可能となっている。Fig. 8は自動調整の機能を使った/etate/の合成シミュレーションの様子とその合成音声のサウンドスペクトログラムである。時間間隔を2000 sample (100 ms)とし、最初の/t/では $DTO$ を200 sample(10 ms)、後の/t/では100 sample(5 ms)とした例である。グラフから、音響管の断面積(area)と声門中立面積( $A_{g0}$ )の閾値間の間隔が、設定した時間間隔と $DTO$ で再現されているのがわかり、この機能の利便性が理解できる。

## 5 まとめ

本稿では、音響管モデルに基づく音声合成システムを用いて子音を合成する際の調音タイミングについて検討した。子音閉鎖に関する音響管の断面積と声門中立面積とにそれぞれ閾値を設け、この時間的相対関係を調整する機能をGUIに設けることで、子音破裂のスペクトル構造の時間変化など、パラメータの設定の効果がより把握しやすくなった。また、断面積と声門中立面積の時間変化パターンは縦続1次系関数を利用して記述されるが、設定しようとする上記の所望の時間差に応じて縦続1次系関数の仮想的入力時刻を自動的

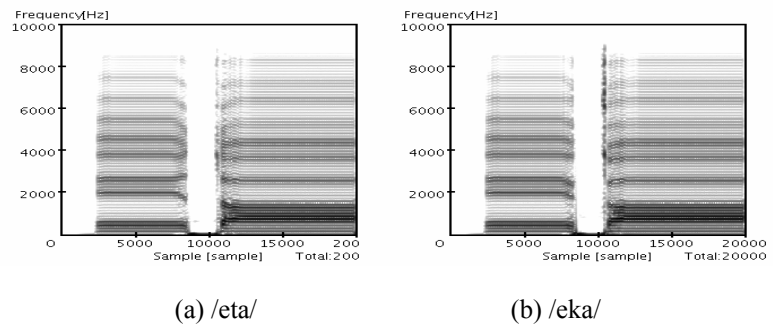


Fig. 7 Sound spectrogram for /eta/ and /eka/.

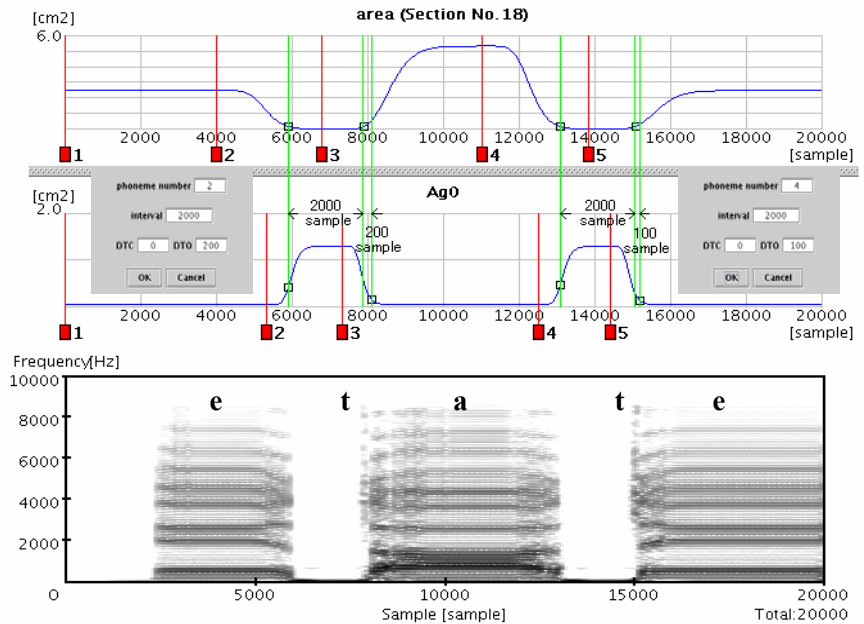


Fig. 8 An example of automatic parameter tuning for /etate/.

に調整する機能をGUIに組込んだ。この実現により、従来は試行錯誤的に調整してきたパラメータの自動調整が可能となり、より簡易で確実に子音の合成ができるパラメータを得ることが可能となった。

今後、他の子音についてもシミュレーションを進める。

## 参考文献

- [1] 緒方他, 信学技報, SP2004-30, 7-12, 2004.
- [2] 緒方, 増矢, 音響学会誌, vol. 62, no. 3, 199-207, 2006.
- [3] M. M. Sondhi and J. Schroeter, IEEE Trans. Acoust., Speech & Signal Process., ASSP-35(7), 955-967, 1987.
- [4] 緒方, 園田, 音響学会誌, vol. 55, no. 3, 156-164, 1999.
- [5] R. D. Kent and C. Read 著, 荒井, 菅原 監訳, 音声の音響分析 (海文堂, 東京, 2004) .
- [6] 緒方, 大塚, 音講論集 (秋), 165-166, 2006.
- [7] K. Ogata and B. Yang, Proc. of 19th International Congress on Acoustics, CD-ROM CAS-03-010, 2007.