Vol.2014-MUS-103 No.13
2014/5/25

音声生成研究の経緯と音声合成に向けた展望

誉田雅彰[†1]

音声生成研究と音声合成研究は，音声特徴と言語情報との関係を明らかにし，それを元に音声を人工的に作り出す共通の課題を有している．一方，音声信号を表現するモデル，音声を記述するパラメータ，言語情報から音声パラメータを生成する方法など，音声を作り出すアプローチにおいて両者には大きな隔たりがある．その背景には，音声生成研究の主眼が人間の発声発話機構を解明することであるのに対して、音声合成研究の主眼が自然でかつ多様な音声を工学的に作り出すという、両者の研究のねらいの違いが関係している．本稿では，音声生成研究の経緯と現状を紹介するとともに，両者のアプローチの違いを埋める上での問題点や課題など，音声合成に向けた展望について述べる．

Review of Speech Production Research and Trends for Speech Synthesis

MASAAKI HONMDA[†1]

. Speech production research and speech synthesis research have a common target in a sense of artificially generating the speech signals from the phonemic information based on their relationship. There exists, however, a significant gap in the approach in terms of the speech signal model, the parameter representation and handing of the parameters for a given phonemic information. The gap seems to originate from each research object. Speech production research is aiming for clarifying the human speech production mechanism as while speech synthesis research is aiming for automatically generating a natural sounding speech. In this paper, we introduce a recent trend of speech production research and argue the problems in overcoming the gap between speech production and speech synthesis researches.

1. はじめに

音声生成研究と音声合成研究は，音声特徴と言語情報との関係を明らかにし，それを元に音声を人工的に作り出すという意味において共通の課題を有している．一方，音声信号のモデル表現，音声を記述するパラメータ，言語情報から音声パラメータを生成する方法論など，音声を作り出すアプローチにおいて両者の間には大きな隔たりも存在する．その背景には，両者の研究のねらいの違いが関係しているように思われる．音声生成研究のねらいは，合成音声の品質を高めることよりも，むしろ音声生成の流体音響レベル，発話動作の力学的生理的レベル，さらには運動制御や運動計画に関する脳神経レベルにおいて人間の発声発話機構を解明することにある．一方，音声合成研究のねらいは，許容されるメモリ容量や演算量の中で，言語情報から自然でかつ多様な音声を作り出すことにある．このような両分野の研究のねらいの違いは、近年音声生成研究分野の成果が必ずしも直接音声合成技術に還元されない 1 つの要因になっているように思われる。

本稿では，音声信号モデルと音声パラメータの動的表現形式について，音声生成と音声合成のアプローチを対比させながら紹介するとともに，音声生成研究を音声合成技術に生かす上での問題点や課題について述べる．

音声生成・音声合成研究の経緯

図 1 に示すように, 音声生成と音声合成の研究の流れを経時的に捉えると，1960 年代後半から 1970 年代にかけて, Fant の声道音響理論を基礎とする声道音響モデル[1],[2]とボコーダに代表されるソースフィルタモデル[3],[4]など音声を表現する信号モデルが確立され，これらのモデルに基づく音声合成の研究が盛んに行われた．また，声道断面積関数の推定[5],[6]や調音パラメータの推定[7],[8],[9]など，両者のモデルを関連づける研究も多くなされた．また，音源モデルに関しても，石坂・Flanagan の 2 質点モデルモデル[10]や Rosenberg の声帯音源モデル[11]など，音声生成研究と

†1 早稲田大学スポーツ科学学術院
Waseda University、Faculty of Sport Sciences

音声合成研究の間に一定の親和性が存在した．1980年代以降になると，音声生成研究の主眼は，人間の音声生成機構をより精密に表現することを目指し，声道の FEM 音響モデル[12][13]，舌の弾性体の構造や筋構造を精密に表現するモデル[14][15]，空気流体現象と声帯の力学的な構造を精密にシミュレートする声帯振動モデルなどの研究に展開していった．また，発話動作に関しても，Articulatory Phonetics[16]や CD Model[35]，Task Dynamic Model[17] [19]，DIVA Model[18]など，言語情報から調音運動を構造的に表現するモデルの研究がなされ，近年では脳内における運動計画や発話動作獲得メカニズムの解明に向けて，脳神経レベルでの研究が展開されている[20].

一方，音声合成の研究は，調音結合によって生じる音声パラメータの変動を規則によって作り出すことの限界，またメモリ許容量の増大を背景として，音声単位を音節単位などに拡張して接続する方法[21]，大量の音声データベースから所望の言語情報に合致する音声単位を選択して接続するコーパスベースの音声合成法[22], HMM を用いて音声単位を安定かつ自動的に抽出し，効率的に接続する方法[23]など，統計的手法を基本として高品質の合成音声の実現に向けた研究が展開された．また，近年では，感情音声やパラ言語音声，会話音声を対象とする音声合成や，音声と発話時の顔映像を同時に合成するトーキングヘッドなど，音声合成技術の適用領域を拡大する方向で研究が展開されている．

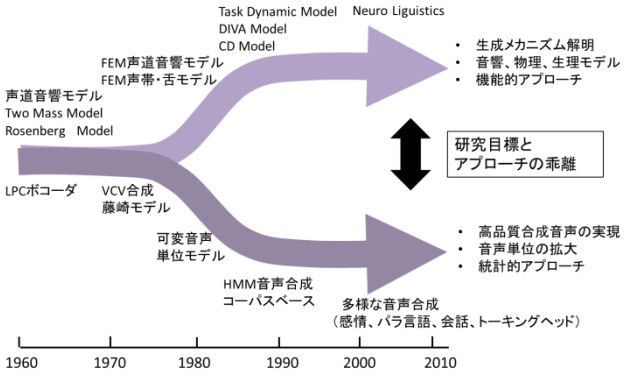

図 1 音声生成研究と音声合成研究の経緯
Figure 1 Trends of speech production research and speech synthesis research

音声生成モデルと音声合成モデル

音声生成と音声合成とでは，音声信号の表現モデルが異なる．音声生成においては，図 2(a)に示すように，人の音声生成過程を模擬する声帯・声道モデルを基本モデルとし，肺圧，声帯張力，声門開閉度を音源パラメータとし，声道断面積関数，調音器官の形態を表す調音パラメータ，あるいは調音器官の各筋肉の筋収縮度を声道パラメータとして音声生成モデルを制御する[10]．一方，音声合成では，図 2(b)に示すようにソースフィルタモデルを基本とし，音源振幅，基本周期，有声・無声判定からなる音源パラメータと音声スペクトルパラメータを用いて音声合成モデルを制御する[1].

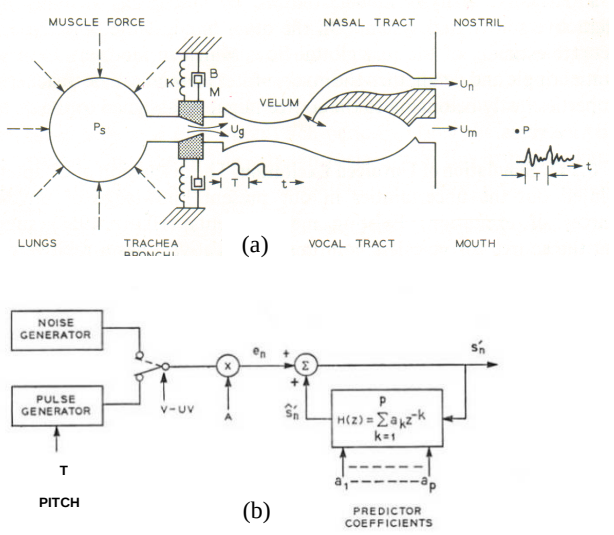

図 2 音声生成モデル（上）と音声合成モデル（下）(文献(1)より)
Figure 2 Speech production model and speech synthesis model. (Ref. (1))

音声生成モデルの長所と短所

音声生成モデルの長所を以下に示す.

(1) パラメータの軌道が時間的に連続しており，その挙動は比較的単純である.
(2) 音素特徴や調音結合が明示的にパラメータ上で記述できる.
(3) パラメータは発声発話動作に直接的に関係することからその物理的，生理的な意味づけが明確であり，許容されるパラメータ空間が定義しやすい.

図 3 は, /ese/の発声時における音声波形, 舌先の舌運動, および声門運動の実測値を示したものである[24]．舌運動と声門開閉運動は時間的に滑らかで単純な軌道を示しているが，音声波形上では有声音から無声音への切り替えが生じ，連続した舌運動に対応したスペクトル変化を見出すことは難しい．この例に示すように，音声生成モデルでは音声の非線形現象を生じさせる要因がモデルに内在する形で組み込まれており，そのことが比較的単純なパラメータ制御によって複雑な音声現象を表現できることが長所となる.

生成モデルの 2 つ目の長所については，次節の調音運動軌道生成で述べることにする.

3 つ目の長所であるパラメータ空間に関しては，生成モデルにおいて一般的に言える点ではあるが，声道断面積関

数，調音パラメータ，筋収縮パラメータによってその程度は異なる．声道断面積関数は正の値をとり，その最大値が

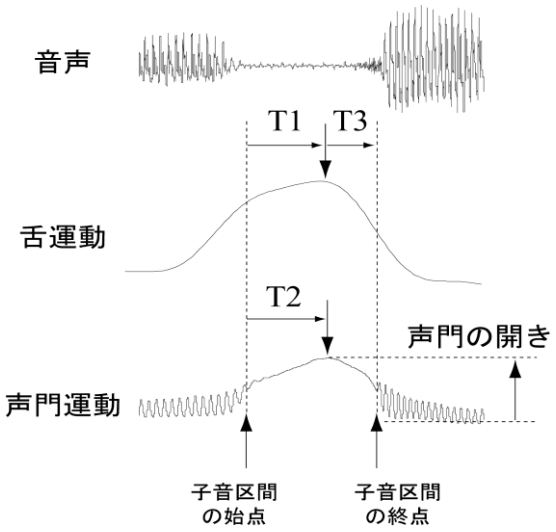

図 3 /ese/発話時における音声波形，舌運動および声門運動

Figure 3 Speech waveform and movements of the tongue tip and the glottal opening.

物理的に制限されるが，個々の値については制約が弱い．より制約を強める方法として，声道断面積関数をフーリエ級数展開する方法や，せばめの位置と大きさをパラメータとして関数近似する方法などが用いられる．

一方，調音パラメータは調音器官の取りうる形状に基づいて声道断面積関数の許容範囲を制約する．このような制約は不自然な声道形状を制約できる利点はあるが，その反面，物理的な制約が信号モデルとしての表現力を損なう面もある．また，調音器官の形状は個人差が大きく，話者依存の調音モデルを用いる場合，同一の座標軸で表されるスペクトルパラメータや声道断面積関数とは異なり，個別の座標系で音声を記述することになる．調音レベルにおける話者正規化の問題は音声生成モデルにおける課題となっている[25].

音声生成モデルを音声合成に適用する上での最大の問題は，音声からパラメータを高精度で抽出する Speech Inversion の問題が十分には解決されていない点にある．Speech Inversion を音声合成技術に応用するには，推定されたパラメータが，生成パラメータとしての長所を保持しかつ品質の良い音声が復元できることが必要となる．一般的に，物理的な制約を多く取り込んだ生成モデルに基づく Speech Inversion では，生成パラメータの長所は保持されるが分析合成系としての性能が不十分であり，制約の弱い生成モデルの場合はその逆の傾向がある．

音声合成モデルの長所と短所

ソースフィルタモデルに基づく音声合成モデルの長所と短所は，生成モデルと逆の関係にある．最大の長所は，精度の高いパラメータの分析手法が確立されていること，また分析合成系において一定以上の音声品質が担保されることである．ただし，分析合成音の品質は，音声合成音の品質の上限を規定するものであり保証するものでない．音声合成では，スペクトルパラメータとは独立に基本周波数を変形する必要があり，その場合の音声品質の劣化が問題となる．これまでの音声合成では，基本的に基本周波数毎のスペクトルパラメータを用いる方法でこの問題に対処している．

基本周波数の変形に伴う音声品質の劣化が，音声生成モデルにおいて解消されるかは未だ検討の余地が残されているが，基本周波数や肺圧（音声の振幅）の変形に伴う声帯音源スペクトル[1]の変化や，喉頭調節に伴う声道長や喉頭付近の声道形状の変化[32]など，生成的視点からより直接的に基本周波数の変形に伴うスペクトルの変化を表すことが可能となる．一方，ソースフィルタモデルでは音源特性と声道共鳴特性がスペクトルに混在した形で表現されていることが，基本周波数の変化に伴うスペクトルの合理的な変形規則を見出すことを困難にする１つの要因であると考えられる．

音声に対応するスペクトルパラメータ空間の定義が難しいことも，音声合成における課題の１つである．
音声合成では平均的な音声パタンを求めるためにスペクトルパラメータの平均操作が行われるが，平均操作が合成フィルタの安定性を阻害する場合があること，また音声生成面から見て単なる平均操作が平均的な音声パタンになるのかという疑問が生じる．特に子音の場合，生成パラメータでは調音位置と狭めの大きさに普遍的な子音特徴があり，それに関与しない部分は大きく変動する．したがって，生成パラメータ上での平均操作では普遍的な特徴が保存されることになる．一方，生成パラメータ上の一部の変動が全てのスペクトルパラメータに反映されることになるため，平均操作によって子音の普遍的な特徴を保持することにならない．

パラメータの動的モデル

音声生成モデルでは，パラメータの時間的な連続性，動的な挙動の単純性，音素に対応する安定な特徴などの長所があり，調音結合現象を構造的に表現するモデルの検討が進められた．一方，音声合成モデルでは，スペクトルパラメータの時間的挙動が複雑であり，またスペクトルパラメータと調音特徴との対応関係も複雑なため，多次元パラメータ空間上における時間パタンとして音素との対応を統計的に捉えるモデルの検討が進められた．

調音運動軌道生成モデル

ここでは，調音パラメータを生成パラメータとする運動軌道生成モデルについて紹介する．軌道生成モデルの研究では，音素に対応した運動目標（タスク）をどのように定義するのか，調音結合をどのように記述するのか，発話動作における協調動作をどのように記述するのか，軌道生成モデルが脳内における発話運動制御と合致するのかなどを課題として研究が進められた．

運動タスクに関しては，人間の発話運動計画にも関連する課題であり，モデル化と同時に多くの実験的な検討が行われ，現在でもホットなトピックとなっている．運動タスクの仮説としては，大きく分けて，Haskins が主張する Articulatory Gesture を運動タスクとする仮説と MIT が主張する Acoustic Target を運動タスクとする仮説がある．前者の仮説に基づく代表的なモデルが Task Dynamic Model[17]であり，後者の仮説に基づく代表的なモデルが DIVA Model[18]である．

Task Dynamic Model では，声道の狭めの位置と大きさを主要な運動タスクとし,それらを声道変数として定義する．発話動作は声道変数上にいて，その目標値を表すステップ関数と臨界二次系の動特性によって記述される．声道変数と各調音器官を表す調音変数は 1 対多の関係にある非線形関数で表され，そのヤコビアンの疑似逆行列を用いて声道変数から調音変数が決定される．唇閉鎖時における顎と上下唇間などの協調動作は，声道変数から調音変数を変換する過程で自動的に決定される．また，声道変数のステップ関数の時間長や強度を音素の調音特徴に応じた規則に基づいて指定することにより，音素間の調音結合を表現する．Task Dynamic Model では，声道変数の運動軌道が陽に指定されたが，声道変数による運動タスクを到達時点において指定し，運動規範最小化規範に基づいて調音変数の軌道を算出するモデルも提案されている．

DIVA Model をはじめとする Acoustic Target モデルでは，音響目標値をホルマント周波数によって指定し，音響誤差と調音運動の運動規範からなるコスト関数を非線形最適化する手法を用いて調音変数を決定する．このモデルでは，調音器官の協調動作の生成が主眼となっている．

脳内における運動計画のメカニズムの視点から見た場合，これらのモデルはいずれも脳内において発話時においてあらかじめ調音運動軌道全体が計算される内部モデル[26]の存在を仮定している．このような内部モデルに対し，平衡軌道仮説(Equilibrium point hypothesis)[27][28]では，発話時に脳内でそのような複雑な計算が行われているとは考えにくく，各時点で指定される運動目標に筋の固有受容系によるフィードバック制御に基づいて運動軌道が決定されるモデルも提案されている[29]．発話運動軌道を再現することに留まらず，脳内における発話運動計画と運動制御の実体を明らかにすることは，音声生成研究の重要な課題となっている[30].

スペクトルパラメータの時間パタン表現

音声合成において用いられる MFCC などのスペクトルパラメータでは，調音結合を明示的に表現することが難しい．また，ホルマント周波数は声道共鳴特性を直接反映するスペクトルパラメータであり，声道断面積関数等の生成パラメータとの親和性が高く，Klat Talk など音声合成の初期研究段階では広く用いられてきたが，子音に対するホルマント分析の問題やホルマント合成音の品質の限界などにより最近では用いられることが少なくなってきている．

音声合成では，調音結合の明示的な扱いを避ける手段として，音節以上の大きな音声単位を用いる方法が広く用いられている．そこでの中心的な課題は，音声単位の設定方法，抽出方法，効率的な表現方法，合理的な接続方法などである．音声単位としては，CV 音節や VCV 音節，CVC 音節のように明示的に設定する方法から，COC 方式[21]のように，多様な音声パタンを最も効率よく表す可変長の音声単位をスペクトル歪最小化規範に基づいて抽出する方法や，基本周波数とスペクトルパラメータを対にした単位を用いる方法，また大量の音声データベースの中からコンテキストと韻律情報が一致する最長音声単位を選択し，音声波形レベルで接続する方法[22]などが提案されている．また，音声単位を効率的に表現し，個々の音声単位を合理的に接続するモデルとして HMM 音声合成が提案され，現在の音声合成技術の主流となっている．

このような研究の経緯をみると，音声合成の研究は，多様な調音結合現象を表現するための音声単位の効率的な拡大と，音韻情報と韻律情報を独立に制御するための音声単位の拡張および合理的な選択にあるといえる．また，これらの研究は,各目的に応じた合理的な統計的モデルの設定,許容されるメモリ容量の拡大，大量な音声データベースに支えられているといえる．このような音声合成の研究の流れは，テキスト情報から品質の高い音声を工学的に作り出すという目的に沿った自然な流れであるといえる．一方，音声単位の拡大と音韻情報と韻律情報の独立な制御に限定した機能の実現は，音声データベースへの依存度を高めるとともに，規則による音声合成の側面あるいは音声生成研究の中で見いだされる知見の適用性を弱める要因にもなっている．また，多様な会話音声の生成やパラ言語音声や感情音声など音声合成の対象を拡大する場合に，規則を見出すことなしに音声データベースに内在する現象を再現する方法で問題が解決するかは，検討の余地があるように思われる．

音声生成に基づく音声合成への展開

ここでは，音声生成研究と音声合成研究の目標とそれに向けてのアプローチの乖離が顕在化している現状において，音声生成研究が音声合成に貢献する可能性について私見を述べる．

その 1 つ目は，音声合成において用いられている音声信号モデルを音声生成モデルによって置き換えることである．音声生成研究では，声道音響モデルと声帯振動モデルの両面においてより物理的実体を再現するモデル化の研究が絶え間なく続けられている．特に，声道音響モデルに関しては，MRI による声道の 3 次元形状の動的な測定技術が現実的になりつつあり[33]，[34]，3 次元声道形状から音声スペクトルを求める手法も整備されつつある[12],[13]．このような計測技術の進歩と物理的実体を再現する音声生成モデルの進歩は，生成パラメータの実測値から音声信号を生成する手段を提供することになる．また，このような生成モデルに基づく Speech Inversion 技術は, 音声から大量な生成パラメータを得る手段を提供し，生成パラメータに基づく音声合成のパラダイムチェンジを可能にする．

ただし, このようなパラダイムチェンジを促すためには，物理的な音声生成モデルが，音声生成過程における様々な非線形現象を含めて実際の音声を精度よく再現できることが前提となる．

2 つ目のアプローチは，EMA 装置などを用いて計測された大量の調音データと音声データから調音・音響マッピング関数を構成し，それを元に生成パラメータに基づく音声合成を構築するアプローチである．図 4 に示す HMM 音声生成モデル[31]はこのような考え方に基づくモデルの 1 つであり，生成パラメータの長所を生かしつつ分析合成音と同等の音声品質も担保でき，また HMM 音声合成モデルとの親和性も高い．また，音声から調音パラメータを比較的高い精度で推定することも可能である．ただし，このような調音・音響マッピング関数に基づくモデルは声道断面積関数などの物理的実体を直接表現していないため，マッピング関数の近似精度不足や関数の定義域のあいまい性があり，任意のパラメータの操作を行った場合に合成音声の品質劣化を招くおそれがある．

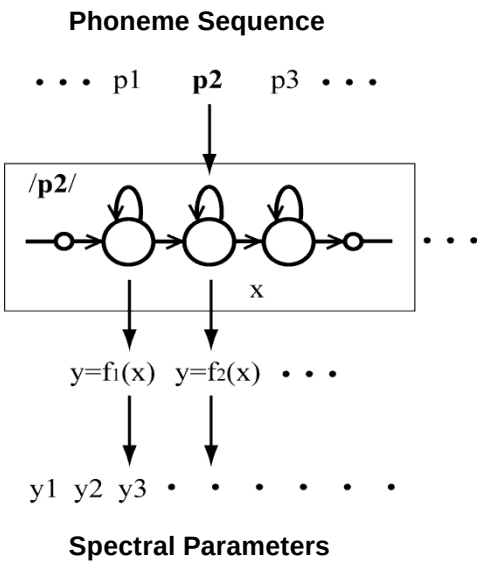

図 4 HMM 音声生成モデル
Figure 4 HMM speech production model

3 つ目のアプローチは，音声生成研究における知見を間接的に音声合成に生かすアプローチである．例えば，前述した基本周波数の変化に伴い音声スペクトルが変化する要因を，音声生成における声帯音源特性や声道形状の変化から探るのも 1 つのアプローチである．喉頭管の形状が F3 付近スペクトルの強度に深く関係し，それが音声の個人性を関連しているという知見[32]は，音声スペクトルを構造的に捉える上で有効な手掛かりになると思われる．また，発話様式の違いなど，その特徴が発話動作において直感的捉えられる対象に関しては，スペクトル分析だけに留まらず，発話動作と音声の同時観測データを基に，発話動作と対比しながらスペクトルの特徴を見出すアプローチが有効になると思われる．

むすび

音声生成と音声号合成に関して，音声モデルとパラメータの表現方法の違いについて両者を対比させながらその特徴について述べた．また，両者の研究の目標とアプローチが乖離しつつある現状を指摘するとともに，音声生成研究が音声合成に貢献できるアプローチについて述べた．両研究の乖離は，ある意味で必然的な面もあるが，音声が作り出される仕組を明らかにし，それを人工的に実現するという両者を共通の課題を目指して，今後両研究がより密接に関連していくことが期待される．

参考文献

1) J. L. Flanagan, Speech Analysis, Synthesis and Perception, Springer-Verlag, Berlin・Heiderberg・New York, 1972.
2) M. M. Sondhi and J. Schroeter, A hybrid time-frequency domain articulatory speech synthesiser, IEEE Trans. ASSP, vol.35, pp.955-967, 1987.
3) 板倉文忠，斉藤収三，統計的手法による音声スペクトル密度とホルマント周波数の推定，電子通信学会論文誌，53-A, 1, pp.35-42,1970.
4) B. S. Atal and S. L. Hanauer, Speech analysis and synthesis by linear prediction of the speech wave, JASA, vol.50, pp.637-655, 1968.
5) M. R. Schroeder, Determination of the geometry of the human vocal tract by acoustic measurements, JASA, vol.41, pp.1002-1010, 1967.
6) M. M. Sondhi.and J. R. Resnick, The inverse problem for the vocal tract: Numerical methods, acoustical

experiments, and speech synthesis, JASA, vol.73, pp.985-1002. 1983.

7) 白井克彦，誉田雅彰，音声波からの調音パラメータの推定，電子通信学会論文誌，61-A， no.5, pp.409-416, 1978.

8) B. S. Atal, J. J. Chang, M.V. Mathews, and J. W. Turkey , Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique, JASA, vol.63, pp.1535-1555, 1978.

9) J.Schroeter and M.M.Sondhi Speech Coding Based on Physiological Models of Speech Production, in Furui, S. and Sondhi, M.M. (eds.), Advances in Speech Signal Processing (Marcel Dekker, New York, Basel, Hong Kong): 231-268, 1992.

10) K. Ishizaka and J. L. Flanagan, Synthesis voiced sounds from a two-mass model of the vocal cords, BSTJ, vol.51, no.6, pp.1233-1268, 1972.

11) S. Rosenberg, Glottal pulse shape and vowel quality, JASA, vol.49, pp.583-590, 1970.

12) 松崎博季，元木邦俊，三木信弘，有限要素法における3次元声道モデルの曲り及び断面形状の簡略化に関する検討，音響学会誌，vol.59, no.8, pp.440-449, 2003.

13) H.Takemoto, P. Mokhtari, and T.Kitamura, Acoustic analysis of the vocal tract during vowel production by finite-difference time-domain method, JASA, vol. 128, pp.3724-3738, 2010.

14) J. Dang, J. and K. Honda, A physiological model of a dynamic vocal tract for speech production, J. Acoust. Soc. Jpn (E), 22, 415-425, 2001.

15) S. Buchaillard, P. Perrier and Y. Payan， A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning, JASA, vol.124, pp.2033-2051, 2009.

16) C. Browman and L. Goldstein, Tiers in articulatory phonology, with some implications for casual speech, in Papers in Laboratory Phonology, Ed. By J. Kingston and M. Beckman, Cambridge University Press, 1990.

17) E. Saltzman and J.A.S. Kelso, Skilled actions: a task dynamic approach, Psychological Review, vol.94, pp.84-106, 1987.

18) F. H. Guenther, A neural network model of speech acquisition and motor equivalent speech production. Biological Cybernetics, vol.*72*, pp.43-53, 1994.

19) T. Kaburagi and M. Honda, Dynamic articulatory model based on multidimensional invariant-feature task representation, JASA, vol.110, pp.441-452, 2001.

20) F. H. Guenther, S. S. Ghosh, and J.A.Tourville Neural Modeling and Imaging of the Cortical Interactions Underlying Syllable Production, Brain and Language, vol.96, pp.280-301, 2006.

21) 中嶌信弥，浜田洋，合成単位を自動生成する規則合成法，信学技報，SP87-15, 87,31, pp.57-64, 1987.

22) Y. Sagisaka, Speech synthesis by rule using an optimal selection of non-uniform synthesis units, Proc. ICASSP, pp.679-682, 1988.

23) 徳田恵一，HMM による音声合成の基礎，信学技報，SP2000-74, 100,392, pp.43-50, 2000.

24) 藤野昭典，鏑木時彦，誉田雅彰，村野恵美，新美成二，無声子音における舌・唇と喉頭の調音運動の時間関係の分析．日本音響学会誌，vol.59, no.3, pp.121-130, 2003.

25) 北村達也，竹本浩典，本多清志，母音発声時の声道断面積関数の個人差について， 日本音響学会講演論文集，pp.659-660, 2010.

26) M. Kawato, Internal models for motor control and trajectory planning. Current Opinions in Neurobiology, vol.9, pp.718–727, 1999.

27) A. G. Feldman, Once more on the equilibrium-point hypothesis (λmodel) for motor control. J Mot Behav 18: 17-54, 1986.

28) E.Bizzi， N. Accornero, W. Chapple, N. Hogan, Posture control and trajectory formation during arm movement. J Neuroscience, vol.4, pp.2738-2744,1984.

29) Payan, Y. and Perrier, P. ,Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis, Speech Communication, 22(2/3), pp.185–205, 1997.

30) P. Perrier, Gesture planning integrating knowledge of the motor plant's dynamics: A literature review from motor control and speech motor control, in Speech planning and Dynamics, ed. by S. Fuches etc., Peter Lang, 2012.

31) S. Hiroya and M. Honda, Estimation of ariculatory movements from speech acoustics using an HMM-based speech production model, IEEE Trans. on Speech and Acoustic Processing , vol. 12, no.2, pp.175-185, 2004.

32) H. Takemoto, T. Kitamura, K. Honda, and S. Masaki, Deformation of the hypopharyngeal cavities due to F_0 changes and its acoustic effects, Acoustic Science and Technology, vol.29, no.4,pp.300-303,2008.

33) S. Masaki, M. K. Tiede, and K. Honda, et al.., MRI-based speech production study using a synchronized sampling method. J Acoust Soc Jpn（E）, vol.20（5）, pp.375-379, 1999.

34) B. Sutton, C. Conway, Y. Bae, C. Brineger, Z. P. Liang, and D. P. Kuehen, Dynamic imageing of speech and swallowing with MRI, Proc. Of IEEE Eng. Med. Biol. Soc., pp.6651-4, 2009.

35) Fujimura, O.: C/D model: a computational model of phonetic implementation; in Ristad, Language computations,pp.1–20, 1994.