人 工 知 能 77−14
（1991. 7. 18）

連続音声合成への協調問題解決の応用

小川 均　　松村 雅史

立命館大学理工学部　　大阪電気通信大学工学部

母音を声道の形状モデルを用いた規則合成方式を対象とした自然な連続音声の合成を試みた．各母音に対する声道形状モデルは磁気共鳴映像法（ＭＲＩ）を用いた測定値を用いている．このモデルは各母音を単独で発声している形状を示している．自然な連続音声を合成するにはこれらの形状間の変化を舌や口唇の形の限界や動作速度を考慮して決定しなければならない．本研究では，協調問題解決を利用した．各母音の声道形状に対応するエージェント，時間を調整するエージェント，そして，結果を統合するエージェントがある．これにより，自然な音声が合成でき，さらに，声道中の変化により発声される「わ」「や」「ゆ」「よ」の合成ができた．

An application of cooperative problem solvers for consecutive phoneme synthesis

Hitoshi Ogawa

Ritsumeikan University,
Faculty of Science and Engineering
Kitaku, Kyoto 607

Masafumi Matsumura

Osaka Electro-Communication University
Department of Applied Electronics
Neyagawa, Osaka 572

A vocal tract model for Japanese vowels was built by measuring the 3-dimensional vocal tract shapes by magnetic resonace images (MRIs). To produce consecutive phoneme, there exist the interactions between the different vocal tract shapes of the different vowels, because there are constraints of the vocal tract shape change from one vowel to another vowel. The natural change of the vocal tract shape is obtained by cooperative problem solvers, which consist of three kinds of agents: time agent, vowel agent and unity agent. It can produce the pronunciation (e.g. "ya", "wa" and so on) uttered changing vocal tract shape.

1．はじめに

現在の音声合成は，ＬＳＩ（Large scale integration）等の記憶媒体にサンプリング（標本化）された実音声のデータを，そのままもしくは圧縮して記憶させておく方法がほとんどである．そのため，ある特定の言葉をある特定の調子で発声させることしかできないのが現状であり，あらゆる言葉を発声させるには至っていない．一方，人間の音声生成機構に基づいた音声生成の過程のモデル化を行い，またその方法による日本語の母音の音声合成も行なわれている．[1)2)]この方法は各母音に対する声道をモデル化している．他の方法に比べモデル化という方法をとっていることから，その入力の表現方法は非常に簡単であり，任意の母音を任意の順序で発声できた．しかしながら，音声の連続発声においては各母音の音だけでなく，各母音間のつながりを旨く決定しなければ自然な音声が得られない．一つの単語を発声する場合，各音素は，大きさ（振幅），高さ（周波数）などが違い，それらは均等の重みがあるわけではない．強調すべき音素やそうでない音素がある．また，連続音の発声要求に対して，舌や口唇等の形や動作速度に制限があり，これらの制限をすべて満たす必要がある．そこで，本研究では連続音声合成を各母音に対する声道形状間の干渉とみなし協調問題解決を利用する．すなわち，各エージェントは各母音の発声する声道形状を維持するために前後のエージェントとの調整を行う．これにより，物理的に無理のない声道の動きを実現でき，自然な連続音声の合成を行う．

本論では，最初に本研究で使用する規則合成方式を簡単に紹介し，協調問題解決のための表現方法についてのべる．さらに，連続音声合成に必要なエージェントとそれらの動作について述べる．

2．音声合成方法

2．1　音声合成の現状

現在まで行われている音声合成の方法は，録音編集方式，パラメータ編集方式，そして，規則合成方式の３種類に分類できる．録音編集方式が一番合成音の音質がよく，現在実用化されている音声合成のほとんどはこの方法である．しかしながら，この方法は基本的に人が発声した音声のつなぎ合わせによって音声を合成するものであるので，出力できる語彙が，蓄積した単位の組合せに限られてしまう．つまり，同じ言葉でも文章中の位置によってイントネーションやアクセントが異なるので，１つの言葉に対して複数のタイプを蓄えておく必要があり，それが欠点となっている．

パラメータ編集方式は，人が発声した音声波を分析してパラメータ系列で蓄積しておき，それをつなぎ合わせ，音声波形による音声合成を行う音声合成器を使用して出力するという方法である．この方法では，実際に出力するための音声合成に，音声のフォルマント周波数（共鳴周波数）に注目し，その波形を接続することによって連続発声を行う合成方法，もしくは線形予測分析に基づく分析合成の方法が良く用いられる．この方法は，ハードウエアが複雑になる代わりに記憶容量が少なくてすみ，さらに音素間の接続が滑らかになるようにパラメータの操作・調整を行ったり，音声のピッチや発声速度を制御することも容易にでき，柔軟性のある音声合成が可能となる．基本的に人が発声した音声のつなぎ合わせで音声を合成を行うので，出力できる語彙が，蓄積した単位の組合せに限られてしまうという欠点がある．

規則合成方式は，文字列あるいは音素記号列から音声学的・言語学的規則に基づき，音声合成器を使用して出力する方法である．この方法は上記２つの方法とは異なり，人間が発声した音声をそのまま用いてはいない．人間の音声発生のしくみをモデル化し，発生音を合成する方法である．したがって，出力できる語彙が限定されることもなく，パラメータ編集方式以上に柔軟な発声が可能となる．しかしながら，人間の口から喉における形状（すなわち，声道の形状）をそのまま実現するのは不可能であり，また，実現したとしても実時間で計算が終了しない．したがって，実時間で計算でき，しかも人間の声道形状を旨く反映するモデルが必要となる．

２．２　声道のモデル化

母音発生に必要なパラメータは，口唇の反射係数，口唇の断面積，声門パルスのパルス幅と振幅，そして，声道の形状（声道の断面積）である．声道は，その形状に応じた共振周波数を持つ音響管である．したがって，声帯振動波が声道を通過する際に共鳴が加えられることにより，フォルマント周波数（声道の共振周波数）が発生する．フォルマント周波数は，音韻の種類によって特有の値を持つ．そのため，このフォルマント周波数を決定する要素，つまり声道の形状及び音源である声帯振動をモデル化することが出来れば，音声の発声が可能であり，また声道の形状を滑らかに変化させることで，連続音声の発声も容易である．さらに，合成音声に個人情報を付加できる可能性も生まれる．

この方法を用いたシステムを開発するには，声道形状の十分な解析が必要である．そこでＸ線像により解析が行われてきたのだが，被験者のＸ線被爆量によりデータ数が制限されるため，この方法を用いることは困難であった．しかし，近年磁気共鳴映像法（Magnetic Resonance Imaging：以下ＭＲＩと略す）の発達により，安全に声道の断層像を得ることが可能になり，それによって声道形状の計測も十分に行われるようになった．声道の形状は基本的に舌と口蓋の距離により決定される．口蓋の形は変化しないので，舌の形を得ればよい．しかしながら，舌の形は様々に変化するので決定が難しい．舌の付け根の位置が固定，舌の長さが一定であることを考慮すると全舌部の２点より舌の形状が求まることが分かった．これにより，単母音の合成はつぎのステップで行なう．

（１）与えられた前舌部の２点から舌形状を推定する．
（２）舌と口蓋の位置から声道断面積の関数を得る．
（３）フォルマント周波数を計算する．

２．３　連続音声合成の問題点

本研究で用いる音声合成では，ＭＲＩによる測定値から得られた声道モデルを用いている．しかしながら，連続音声合成の場合は各音素間における声道モデルが必要である．ＭＲＩは静止している物体に対しては有効であるが，移動したり，形状が変化する物に対しては使用できない．したがって，ある声道形状から他の声道形状への会場の変化を求める必要がある．人間の体はその構造から，不可能な形や部分間の関係があり，その機能から動作速度の制限がある．声道においては，舌の形状，舌と口蓋の関係，口唇の断面積（口の大きさ）などや，舌と唇の動作速度な

どに制限がある．また，音素はすべて均等に発生されるわけではない．ある音素は必ず発声されるが，別の音素は隣接する音素と融合したり，異なる音に変化する場合もある．これらの関係をうまく反映し，上記の制限を満足する音声合成を行なう必要がある．各音素に対して発声時間，アクセント，声の大きさ等の決定に対して，各音素の強調の度合や上記の制限および各音素間の関係を考慮しなければならない．音素の数が大きくなればこのためのアルゴリズムは不明確となる．

3．協調問題解決器

2．3で述べたように，連続音声合成を一括してプログラムするのは困難であるので，音声合成の問題を各音素を発声する声道形状間の干渉と捉え，協調問題解決を用いることにした．以下では，使用したシステムＰＳＡ（Problem Solving Agent）の概略を簡単に紹介する．

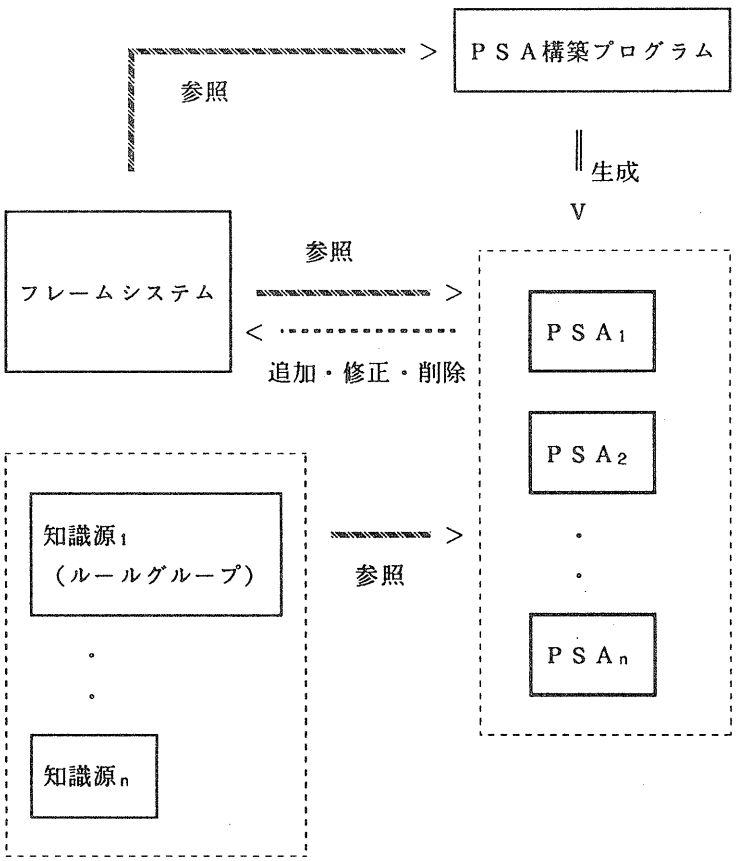

図1．　ＰＳＡシステムの構成．

ＰＳＡシステムは図１に示されるようにつぎの４つから構成される．

ａ．複数のＰＳＡ（Problem Solving Agent）：問題解決機．
ｂ．知識源：ルールグループ．
ｃ．フレームシステム：フレーム形式の知識を扱う．
ｄ．ＰＳＡ構築プログラム：フレームシステムを用いてＰＳＡを構築する．

ＰＳＡの構造は図２に示されるように，推論制御部とデータを蓄積するデータ部からなる．推論制御部には，メッセージ処理や，知識源と推論方法の指定，推論結果のモニタ，データ部変化のモニタを行なうルールが記述される．推論制御部中のルールをメタルールと呼ぶ．データ部には，対象問題に関する情報，および，仮定や目標，推論による途中結果等のデータが蓄積される．

ＰＳＡはメッセージを受理することにより起動される．メッセージに対応する動作は推論制御部で定義される．すなわち，データの登録，推論の実行，推論結果のＣＲＴへの表示，他のＰＳＡへのメッセージ発信等である．推論は知識源と推論方法を指定することにより実行される．１つの知識源は複数のＰＳＡによって使用できる．推論実行時の副作用（取り消しができない実行）はデータ部に対してのデータの断言と削除のみである．

ＰＳＡは論理型プログラム言語Prologにより実現されているので，変数を英大文字を先頭に持つ語で表わし，その他は英小文字の語で表わす．

４．連続音声合成システム

４．１　システム全体の構成

開発したシステム全体の構成図を図３に示す．なお，矢印で示しているメッセージの内容については，４．２で述べる．

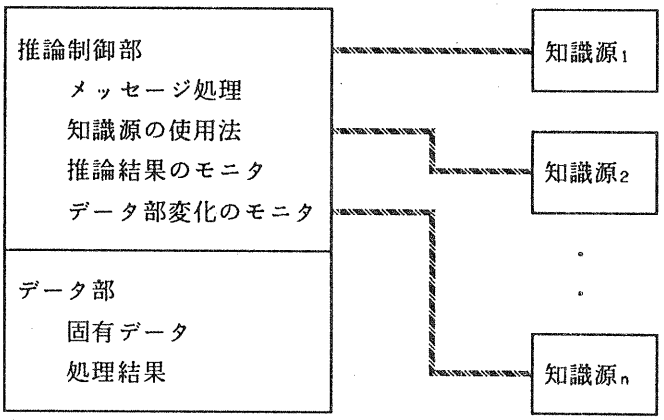

図２．　ＰＳＡの構成と知識源．

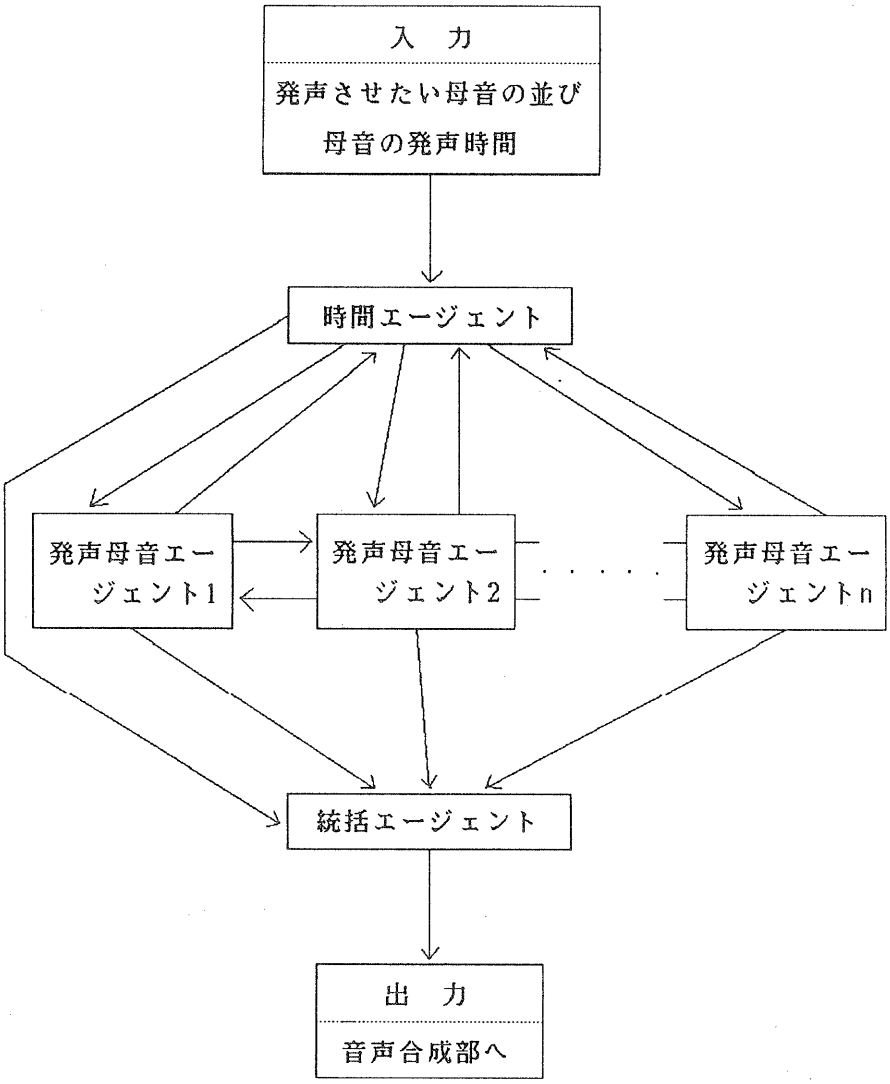

図3　システム構成図

図3に示したように，本システムは大きく分けて以下の3種類のエージェントに分けられる．

（1）時間エージェント：　発声母音エージェントと統合エージェントの生成および起動を行い，問題が生じれば時間の調節を行なう．

（2）発声母音エージェント：　母音が発声できる舌・口蓋の位置が確保できるように隣接発声母音エージェント，および，時間エージェントと交渉する．

（3）統合エージェント：　発声母音エージェントから送られたデータをまとめ，D／A変換部に送る．

以下にシステムの全体の流れについて述べる．

最初に，時間エージェントに発声させたい母音の並びと発声時間が送られると，必要な発声母音エージェント及び統合エージェントを作成し，必要なメッセージを送る．

各発声母音エージェントは，基本的には，発声順序に起動する．前の発声母音エージェントにおける舌・口唇の位置に対して発声しようとする母音の舌・口唇の位置が可能なように調節する．できなければ，舌・口唇の位置に関して前の発声母音エージェントと交渉する．それでダメな場合は時間エージェントに発声時間の変更を要求する．

統合エージェントでは，すべての発声母音エージェントからメッセージを受理すれば，すべてのメッセージをまとめD／A変換部へ送る．

4．2　各エージェントの動作

時間エージェント，発声母音エージェント，統合エージェントのそれぞれの動作について以下に述べる．

4．2．1　時間エージェント

時間エージェントは，発声母音エージェントと統合エージェントの作成と時間に関する調整を行なう．母音の発声時における舌や口唇の動きによって生ずる時間的拘束の知識を持つ．

最初，時間エージェントは，入力データである「発声させたい母音の並び」・「母音・母音間の発声時間」を受けると，発声母音エージェントを発声させたい母音の総数分作成・起動する．また，統合エージェントの起動も行い，発声母音の総数がいくらであるかを通知する．その後，入力された母音の発声時間が発声するための時間的最低条件を満たしているかを調べ，満たしていない場合には，再度母音の発声時間を入力するように求める．満たしている場合には，各発声母音エージェントに，その母音の発声時間・前後の母音との間の時間・前後の母音が何であるかを通知する．

また，発声母音エージェントから時間の変更が必要であるという要求を受けた場合には，再度母音の発声時間を入力するように要求をする．

4．2．2　発声母音エージェント

発声母音エージェントの作成は時間エージェントによって行われ，時間エージェントから母音の発声時間・前後の母音との間の時間・前後の母音が何であるかの情報を受け取った後に起動する．起動後，初めに発声する母音の声道の形が基準になることから，2番目に発声する発声母音エージェント以後は，1つ前の発声母音エージェントから母音の発声時間・前後の母音との間の時間・舌及び口唇の位置データをメッセージとして受け取った後に推論を開始する．

発声母音エージェントは，各母音の舌・口唇の基準位置に関する知識，位置の制限に関する知識を持っている．前の母音の舌・口唇の位置からその母音の舌・口唇の基準位置までの移動が，要求された時間内に移動可能であれば推論は成功する．また基準位置までの移動が不可能である場合も，その母音であることを聞き取れる限界の位置まで妥協し，そのことよって要求された時

間内での移動が可能になれば，推論は成功となる．推論が失敗した場合には，前の母音に対して，舌・口唇の位置データを変更するように要求する．しかし，その変更が不可能であった場合には，時間エージェントにその母音の時間を変更するように要求を行う．要求に対して変更があれば再度推論を行なう．

最終的に，推論が成功した場合は，その結果である舌・口唇の位置データ，母音の発声時間をメッセージとして統合エージェント及び次の発声母音エージェントに送る．

4．2．3　統合エージェント

統合エージェントは，D／A変換部に送るためのデータをまとめるためのエージェントである．したがって，連続母音発声のための知識はもっておらず，またそのための推論も行わない．

統合エージェントは作成されたとき，同時に作成されたすべての発声母音エージェント名を時間エージェントから送られる．したがって，各発声母音エージェントからのメッセージの有無がチェックできる．基本的に，発声母音エージェントは，発声順序に従って起動されるので，すべての発声が旨く計画されるまで最後の発声母音エージェントからメッセージが送られることはない．ある発声母音エージェントから複数回メッセージが送られることもあるが，最新のセッセージのみを有効にすればよい．

5．実験結果

ここでは，本システムの動作を具体的に示す．動作例として図4の様に母音の発声を行ってみる．

図5に本システムを使用した場合と使用しなかった場合の前舌部の2点の動作状況をグラフにして示す．ただし，X座標は顔の前方向であり，Z座標は上方向を示している．本システムを使用せずに単に母音の位置をつなげた例に比べ，無理のない動作となっている．特に，第2音素”う”における，1点目のX座標やZ座標に違いが見られる．

合成音声は，本システムを使用しなかった場合には「いうおう」となってしまうが，本システムを使用すれば「ゆおう」でもなく「いうおう」でもない「言おう」という微妙な連続母音の発声が実現された．また，「や」「ゆ」「よ」「わ」の4つの音の発声も実現された．

母音名	い	いう間	う	うお間	お	おう間	う
時間（sec）	0.050	0.100	0.020	0.220	0.170	0.085	0.010

図4　動作例の母音及びその発声時間

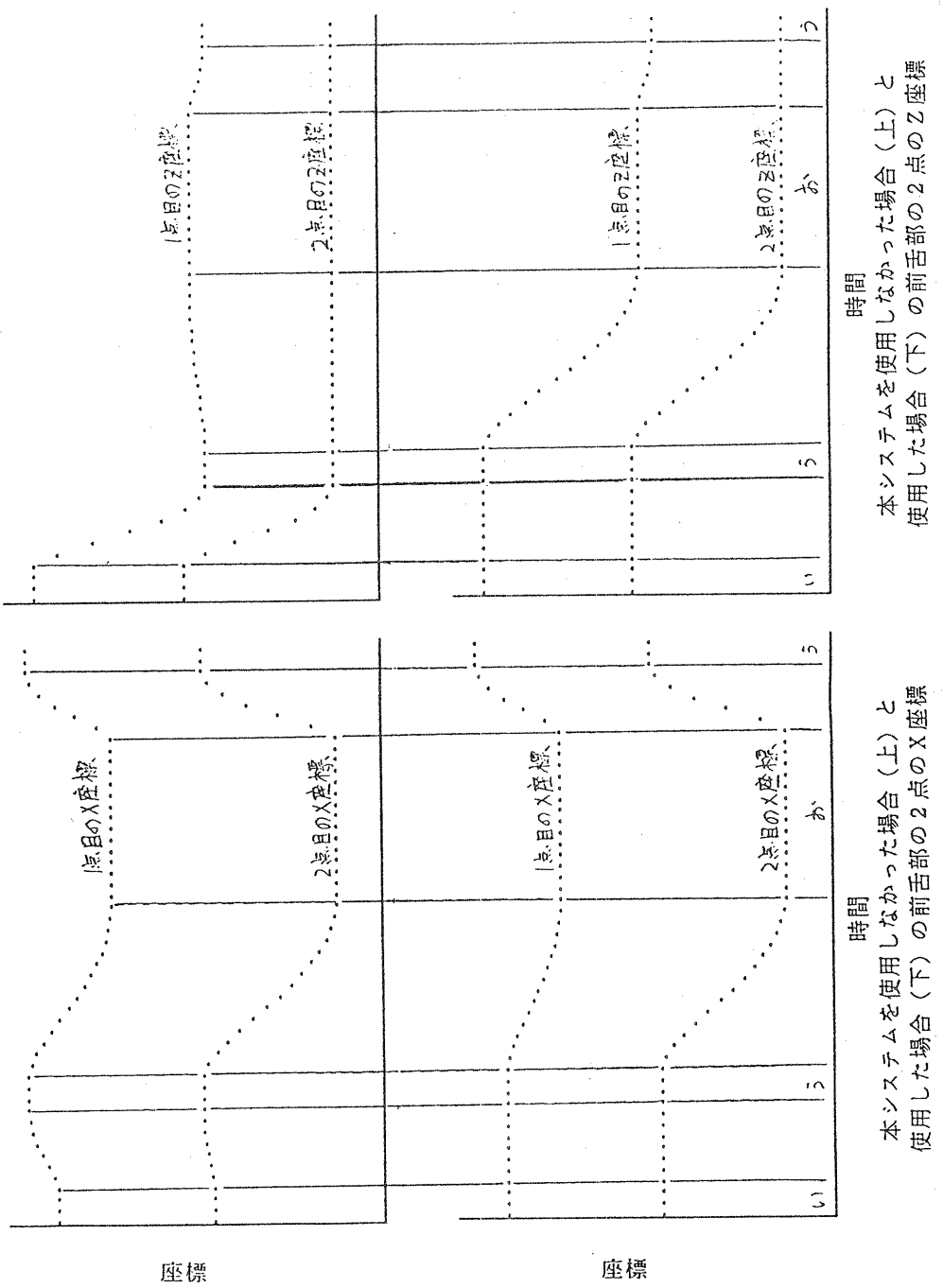

図5　本システムを使用した場合と使用しなかった場合の
前舌部の2点の動作状況

6．むすび

本論文では，自然な連続音声合成を各母音発声時の声道形状間の干渉と捉え，協調問題解決を用いた．実験では，無理の少ない舌や口唇の動きが得られた．さらに，声道の変化により発声できる「や」「ゆ」「よ」「わ」の４つの音も実現できた．入力データは特別なものではなく，発声母音に対して発声時間と振幅（声の大きさ），ピッチ幅（声の高さ）を与えるだけでよい．そして，発声させたい母音の流れを考慮した発声が実現された．したがって，人間の発声により近い発声が得られたものと考える．

個性的な発声という点については，舌・口唇の動きに関する知識を変更することにより様々な実験が可能である．声門のデータや声道のモデルを改良すれば，より自然な音声が合成できるであろう．

分散型の解決方法を用いたことから，このシステムは拡張性が非常に高く，将来ピッチ・声の大きさ・子音等を加えていく場合に非常に有利である．

本システム構築に関して，立命館大学理工学部情報工学科卒業生鮫島秀治氏，倉内靖之氏，および，同４回生竹原直美さん，田中雄一君に協力していただいた．感謝します．

本研究は，平成元，２年度科学研究費補助金（一般研究（Ｃ））研究課題番号０１５８００３６による．

［参考文献］

１）杉浦淳：３次元声道像に基づく自然音声合成に関する研究，大阪大学大学院工　学研究科修士論文（１９９０）．

２）杉浦淳，松村雅史：磁気共鳴映像法による３次元声道形状の計測，電子情報通　信学会，音声研究会，SP89-118（１９９０）．

３）城戸健一著：音声合成と認識，オーム社（１９８６）．

４）鶴田節夫，鬼塚武郎：協調推論型知識情報処理の一方式，情報処理学会論文誌，　Vol.30, No.4, pp.427--438（１９８９）．

５）塚本克治他：ＡＩ～情報処理から知識処理へ～，アスキー（１９８８）．

６）小川均，田村進一：混成型問題解決機ＰＳＡ2.0について，電子情報通信学会技　術研究報告，AI87-20，（１９８７）．

７）システム総合開発：ＥＳＰＡＲＯＮユーザーズ・マニュアル（１９８７）．

８）David Chapman：Penguin Can Make Cake, AI MAGAZINE, Vol.10, No.4, pp.45--50（１９８９）．

９）矢田光治：ＡＩ入門，オーム社（１９８７）．

１０）LIFEBOAT：Arity Prolog V5 リファレンスマニュアル（１９８８）．

１１）Ｂ．Ｗ．カーニハン，Ｄ．Ｍ．リッチー（石田晴久訳）：プログラミング言　語Ｃ，共立出版（１９８１）．

１２）日本放送協会編：ＮＨＫアナウンス・セミナー，日本放送出版協会（１９８５）．