

Title	発話時における調音運動に基づいた調音結合の分析
Author(s)	鈴木, 丈晴
Citation	
Issue Date	2005-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1939
Rights	
Description	Supervisor: 党 建武, 情報科学研究科, 修士

修 士 論 文

発話時における調音運動に基づいた調音結合の分析

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

鈴木 文晴

2005年3月

修士論文

発話時における調音運動に基づいた調音結合の分析

指導教官 党 建武 教授

審査委員主査 党 建武 教授
審査委員 赤木 正人 教授
審査委員 小谷 一孔 助教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

310054 鈴木 丈晴

提出年月: 2005 年 2 月

概要

調音結合は人間の音声生成過程における自然な調音現象である。また音声情報処理上の重要な課題のひとつである。しかしながら調音結合に関する分析と定式化は十分とは言えない。

本研究では発話時における調音結合の分析を調音運動を分析することによって定式化されたキャリアモデルの概念を確かめさらに子音と母音の調音運動の分離を試みた。キャリアモデルの未決定のパラメータを調音データから求め現在のキャリアモデルの向上を計った。さらにそのパラメータが正しい値であるかを模擬実験を通じて確かめ、さらにキャリアモデルの正当性を確かめた。

目次

第1章	序論	1
1.1	研究の背景と目的	1
1.2	本論文の構成	2
1.3	本研究のアプローチ	2
第2章	発話時の調音運動の分析	4
2.1	調音結合における調音器官の運動	4
2.1.1	調音位置	4
2.1.2	スペクトルグラム	6
2.2	EMMA データ	7
第3章	音素別における調音位置の分析	9
3.1	母音に着目した調音位置の分析	9
3.2	歯茎音に着目した調音位置の分析	10
第4章	子音と母音の調音運動の分離	12
4.1	調音特徴を考慮したスペクトル	12
4.2	調音運動の再構成	13
4.3	中心点 T2 に対する相対的な調音運動	13
第5章	調音結合を考慮した時における舌の動きの模擬	16
5.1	Carrier Model	16
5.2	母音についての拘束の度合い	17
5.3	歯茎子音について拘束の度合い	18
5.4	パラメータの評価	23
第6章	全体の考察	28
第7章	結論	29
7.1	本論文で明らかにされたことの要約	29
7.2	今後の課題	29

謝辭	30
参考文献	30

目次

1.1	本研究の位置付け	2
2.1	人間の発声器官の構造	5
2.2	音素環境が異なる音素/a/の調音位置	5
2.3	調音データと同期した音声のスペクトルグラム	6
2.4	日本語 5 母音のフォルマント分布	7
2.5	観測点	8
3.1	最適化を行ったパラメータの調音位置での 5 母音の分布、等高線の概形	9
3.2	最適化を行ったパラメータの調音位置での 5 母音の分布、斜めから見た分布の概形	10
3.3	舌尖付近のパラメータの調音位置での 5 母音の分布、等高線の概形	11
3.4	舌尖付近のパラメータの調音位置での 5 母音の分布、斜めから見た分布の概形	11
4.1	周波数加算平均したスペクトル	12
4.2	調音運動の T1, T2, T3 における再構成した波形	14
4.3	再構成した波形 T1, T3 の速度	14
4.4	中心点 T2 に対する相対的な調音運動	15
5.1	キャリアモデルの概念	17
5.2	舌背付近の調音位置での 5 母音の平均	19
5.3	舌尖付近の調音位置での母音/a/の平均値の分布	20
5.4	舌尖付近の調音位置での母音/i/の平均値の分布	20
5.5	舌尖付近の調音位置での母音/u/の平均値の分布	21
5.6	舌尖付近の調音位置での母音/e/の平均値の分布	21
5.7	舌尖付近の調音位置での母音/o/の平均値の分布	22
5.8	評価の手順流れ図	24
5.9	VCVCV の音素列を与え中心の母音/a/についての舌尖付近の調音位置	25
5.10	VCVCV の音素列を与え中心の母音/i/についての舌尖付近の調音位置	25
5.11	VCVCV の音素列を与え中心の母音/u/についての舌尖付近の調音位置	26
5.12	VCVCV の音素列を与え中心の母音/e/についての舌尖付近の調音位置	26

5.13 VCVCV の音素列を与え中心の母音/o/についての舌尖付近の調音位置 . . . 27

表 目 次

5.1	舌背付近のパラメータを用いて求めた平均値間の距離	18
5.2	母音についての拘束の度合い	18
5.3	舌尖付近のパラメータを用いて求めた平均値間の距離	23
5.4	歯茎子音についての拘束の度合い	23

第1章 序論

1.1 研究の背景と目的

調音結合は人間の音声生成過程における自然な調音現象で、音声情報処理上の重要な課題の一つである、しかしながら、調音結合に関する分析は十分とはいえない。近年、MRIデータに基づいて喉頭や声道形状などの発話器官を表し、調音目標に基づく筋収縮運動を考慮する生理学的モデルを用いた党らの音声合成法では、実際の発話者の音声生成過程を模擬するので、その発話者の個人性を表現することができた。

一方、ヒトが発話する際、調音結合が起きるため、文の音素を先読みしながら脳の中で調音位置を設定する。また調和運動により、実際の調音目標が隣接する音素の影響を受け変動する現象がある。しかし、先述の生理学的発話モデルはこの調音結合を考慮していないため、自然性の良い音声の生成が十分ではなかった。

調音結合に関する研究として、Ohman は音声波形から求めたスペクトルから調音結合の調査を行い、音素列 VCV において母音 (V) と子音 (C) の間の音声には、わたり部分が存在するとしてモデル化をはかった [6][11]。しかしこの分析には調音データを用いていないため調音運動においてもわたり部分が存在するか不明であるためそのままの適応はできない。また Recasens らが舌背の動きと第2フォルマントを分析し、母音 (V) と子音 (C) からなる VCV 音素系列における子音の調音位置が前後の母音の調音位置と関係があることを見出し、その関係の度合いを DAC (調音位置の拘束の度合) のパラメータとして記述した [8]。しかしこの拘束の度合いは1次元のパラメータであり、現在の生理学的発話モデルは2次元のパラメータを必要とするためそのままの適応はできない。

また、党らによって先ほど述べた過去の先行研究を元に母音の調音運動と子音の調音運動をそれぞれ”搬送波”と”調波”とみなして調音結合モデル(キャリアモデル)を提案した [7]。しかしこのモデルにはいくつかの決定されていないパラメータがありまだ未完成である。またこのモデルの評価についても実際の調音データとの比較がされていないため実際の調音位置に即しているか不明である。

そこで本研究では調音結合モデルにある未定であるパラメータを、舌の動きの観測データから隣接する音素との関係の度合いとして2次元のパラメータで求め、キャリアモデルを用いて調音結合を考慮した調音位置の再構成を目的とする。そのため調音データを用いて調音結合による調音目標の変化を分析して定量化する。さらに定量化したデータから調音位置の拘束の度合のパラメータを推定し、得られたパラメータからキャリアモデルを用いて調音位置を計算し調音結合を考慮した調音位置を実現する。この結果はヒトの脳の中

での発話計画のメカニズムの解明にも応用できる。

1.2 本論文の構成

本論文では、第2章に発話時の分析と題して調音結合の分析にあたり用いたデータと調音結合の一般的な説明を行う。

第3章では音素環境が異なる音素の調音位置の分析とその結果を母音と子音(歯茎音)別で示す。

第4章では、多数の音素環境を考慮したときの舌の動きを周波数加算平均の手法により調音運動を再構成した説明を行う。

第5章では、拘束の度合いを求めその値をキャリアモデルに用いて観測データと同じ音素列から調音位置を計算してその調音位置の分布の評価を示す。また、模擬した調音位置の分布がより観測値に従った分布になるように新たにパラメータ追加したことを示す。

1.3 本研究のアプローチ

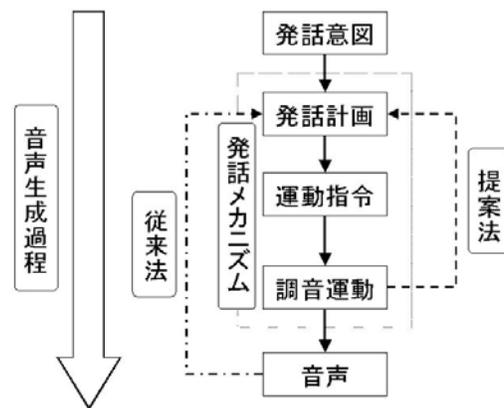


図 1.1: 本研究の位置付け

本研究と従来の研究との比較と概念を示したものを図 1.1 に示す。この図の概念にしたがって調音運動の分析を行って脳の中での発話計画はどのように行っているかを調査する。

1. 意識が情報源となり、意図した言語が形成され、その言語を音素系列に変換する。
2. 変換を行い音素系列に基づいて脳の中で発話するために調音位置、音源などの設定を行う。
3. 対応する運動指令が脳から音声器官の筋肉に送出される。
4. 運動神経指令に従って、それぞれの音声器官が活動する。

5. 音声器官の運動によって、意図した言語情報を持った音声波形が発生する。

本研究では、調音結合を考慮した調音目標の設定を行う。これまでの研究では調音運動の波形から、安定の位置を調音目標に設定している。そこで観測データの安定点を測定し、安定点を分析することで調音結合を考慮した調音位置を調音目標に設定する。それには隣接する音素の影響を分析し隣接する音素が定量的にどれだけ調音目標をずらすかを分析しなくてはならない。本研究の手法としてはまず調音位置の影響を観測データから安定点における調音位置の分布を求めた。さらにその分布から隣接する音素を選択し各音素環境で平均の値を求めた。その平均値の値が他の音素環境とどれだけ異なるかを分析することによって影響度を求めた。この影響度を音素の拘束の度合いとして先行研究で行われているモデルに適応した。

第2章 発話時の調音運動の分析

2.1 調音結合における調音器官の運動

2.1.1 調音位置

人間が音声を生成するプロセスの第一段階は、相手に伝えたい内容を表現する言語を選択し、それを文法に合う言語形式に変えることである。次にこれに従って脳から発声器官に運動神経指令が出され、発話器官の種々の筋肉が動いて、空気振動としての音声がつくられる。人間の発声器官の構造は図 2.1[1] に示す。全体としては一つの連続した官を成している。腹筋が横隔膜を押し上げることによって、肺が押し出された空気は気管を通った後、喉頭の声門 (glottis) すなわち左右の声帯 (vocal cords) の間を通る。通常の呼吸の時は声門は大きく開いているが声を出そうとすると声帯が接近する。この間を肺からの空気を通り抜けようとするために、空気流と声帯との相互作用により、声門が周期的に開閉し、規則的な空気を断続が生じる。これは非対称三角波で近似でき、これが音声の音源となる。これを喉頭原音または声帯音源 (glottal source) と呼ぶ。声帯の緊張が大きく、かつ肺からの空気圧が高いと、声帯の振動周期 (基本周期, fundamental period) が短くなって、音源の音の高さが高くなり、逆の時は低くなる。基本周期の逆数を基本周波数 (fundamental frequency; F_0) と呼び、声の高さ (ピッチ, pitch) に対応する。喉頭より上の部分は声道 (vocal tract) と呼ばれ、成人では約 15-17[cm] の長さがあり、顎、舌、口唇などを動かすことによって、音源波に音色が付与される。鼻腔は、軟口蓋 (口蓋帆) を持ちあげることにより、声道から遮断される。

声道の形を調節することを調音 (articulation) と呼び、発話のための各発話器官の動きを調音運動と呼ぶ。調音に用いられる各部分を調音器官 (articulatory organ) と呼び、その中で、舌、口唇、口蓋帆のように自由に動けるものを特に調音器 (articulator)、調音によって生ずる声道の狭めの位置を調音点または調音位置 (place of articulation) と呼ぶ。

また、会話音声のように連続して発話された音声では、人間の発声器官の動作には慣性がありその速さは制限されるので前後の音素の影響を受けて、音素の音響的変動が起きる。音素と音素の中間部において音響的性質が連続的に推移して、いわゆるわたり部が生じる現象を調音結合 (coarticulation) とする。観測データから異なる音素環境で同じ音素の調音位置を図 2.2 に示す。このように隣接した音素の特徴が調音器官上で相互にオーバーラップし、時間的に広がりを持った変動性として現れている。連続音声の特徴である調音結合には、協調動作、先行性調音、キャリーオーバー、なまけなどの現象が知られて

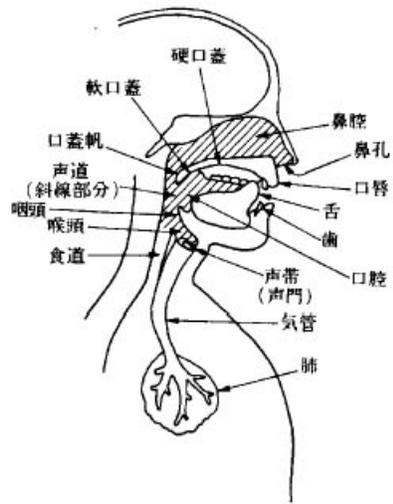


図 2.1: 人間の発声器官の構造

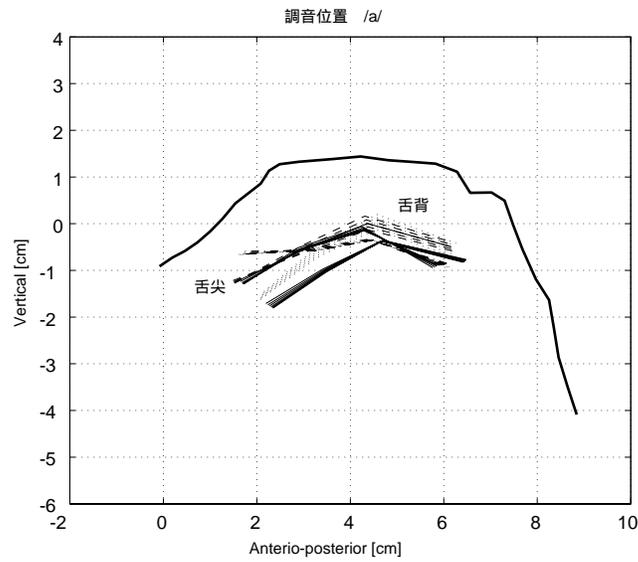


図 2.2: 音素環境が異なる音素/a/の調音位置

いる。これらの現象の背景には、顎、唇、舌などの調音器官が個々に運動学上の自由度を持つと同時に、全体として多自由度の力学系を構成していることや、調音器官の運動がその力学的構造が持つ動特性に拘束されることなどが関係している。

2.1.2 スペクトルグラム

音素を特徴付ける優勢な周波数成分は、声道の共振周波数に対応し、フォルマント (formant) と呼ばれる。有声音には通常3個程度の特徴的なフォルマントがあり、周波数の低い方から、第1、第2、第3フォルマントと呼ばれる。音韻性の点から特に重要なのは、第1と第2フォルマントである。フォルマントのピークの周波数をフォルマント周波数と呼ぶ。フォルマント周波数もまた調音結合の影響を受けて異なる音素環境で変化してしまう。先ほどあげた理由のため、観測される音響的現象もかなり複雑になる。観測データからフォルマントの移動をみるために異なる音素環境で同じ音素のスペクトルグラムを示した。先ほどあげた影響をスペクトルグラムから見れる。音素環境が変化すると同じ音素で

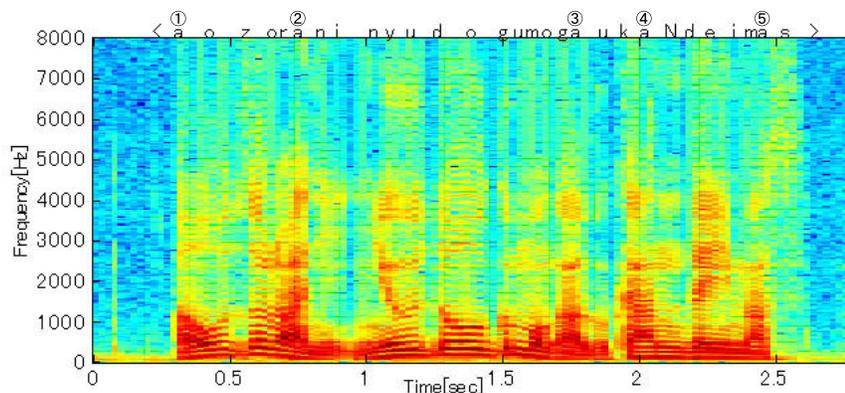


図 2.3: 調音データと同期した音声のスペクトルグラム

も周波数の性質が変化していることがわかる。このことから調音結合を分析する上で連続的なデータから分析する必要がある。また本研究では調音運動から調音結合の分析を行うので音声に対応した連続的な調音運動のデータが必要になる。図 2.4[2] は、第1フォルマント (F_1) を横軸に第2フォルマント (F_2) を縦軸にとった F_1 - F_2 平面に、日本語5母音の分布を示している。図 2.4 のように調音結合などの影響により音声の特徴の分布が重なり合っている部分は両方の音素の特徴を含んでいる。このようにどちらの音素の特徴であるか判断することが困難になってしまう。

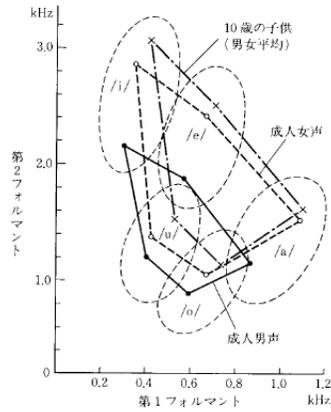


図 2.4: 日本語 5 母音のフォルマント分布

2.2 EMMA データ

本研究で用いた調音運動の観測データはNTTの磁気センサーシステムEMMA(electromagnetic midsagittal articulographic)[12]により計測したものである。計測点としては、正中矢状断面上における下顎、上唇、下唇、軟口蓋、喉頭にそれぞれ1点、舌尖から舌背部までの4点をそれぞれ T_1 、 T_2 、 T_3 、 T_4 の計8点とした。サンプリングレートは250[Hz]であった。上顎に取り付けたコイルの位置を原点として、前部から後部への水平方向をX軸とし、下部から上部への垂直方向をY軸とした。合計16点の位置データを収録した。音声資料は、3名の成人男性が通常の発話速度で朗読した日本語360文である。音声信号と調音運動データに基づいて、音素ごとに中心位置を求め発音記号が付けられている。

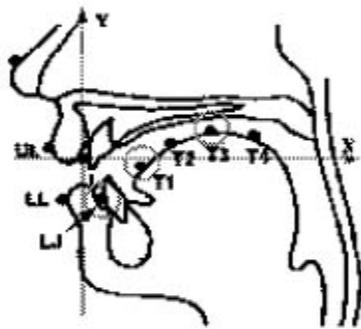


图 2.5: 観測点

第3章 音素別における調音位置の分析

3.1 母音に着目した調音位置の分析

各分布で $V_bCV_cCV_b$ (V は音素の母音、C は音素の子音) の音素列に注目し前後の母音の影響だけを考える。ここでは $V_bV_cV_b$ の分布を求めることによって母音が母音にどれだけ影響しているかという分布を求めた。連続的な調音データの特に関舌の動きを見て調音結合の分析を行う。調音データから、音素列から見たい各音素の前後の音素を含めて調音位置の分布をみた。約 360 文を用いて母音のセグメント各 6 点より、1600 から 3600 個のデータで構成した分布である。この分布はより舌背を捕らえるために最適化を行ったパラメータからの調音位置を用いている。図 3.1 と図 3.2 を見ると各音素の分布の概形はいくつかのピークがありいくつかの分布が重なりあってできていると考えられる。これは前後の音素に影響しピークの数や分布が重なりあっていくつものピークを持つような概形になっていると考えられる。また音素 /i/ の分布が 5 母音の中で一番小さな幅で動いていてかつ頻度が高いことから他の音素から一番影響を受けないと考えることができる。図に載せていない発話者のデータにもこれらと同じ傾向がみられる。

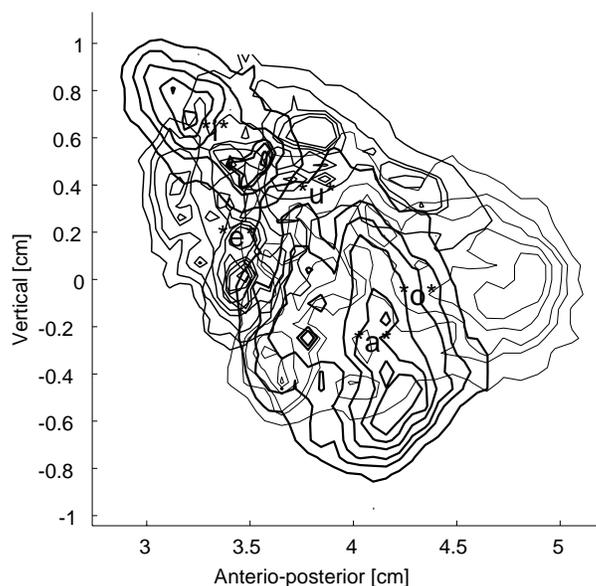


図 3.1: 最適化を行ったパラメータの調音位置での 5 母音の分布、等高線の概形

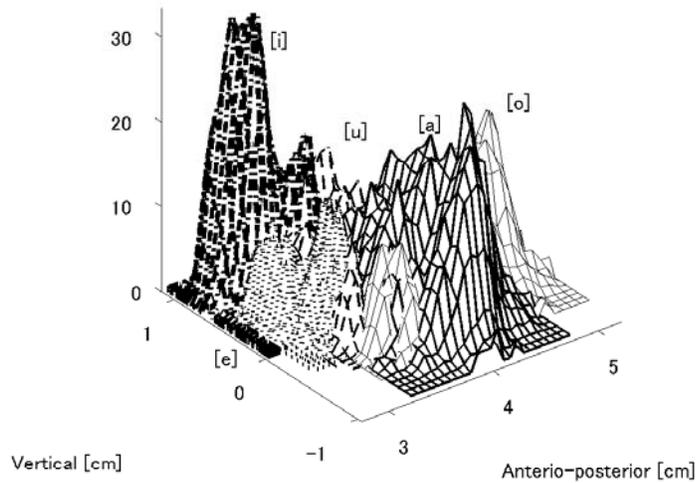


図 3.2: 最適化を行ったパラメータの調音位置での 5 母音の分布、斜めから見た分布の概形

3.2 歯茎音に着目した調音位置の分析

各分布で $V_bCV_cCV_b$ (V は音素の母音、 C は音素の子音) の音素列に注目し前後の歯茎音子音の影響だけを考える。ここでは CV_cC の分布を求めることによって歯茎音子音が母音にどれだけ影響しているかという分布を求めた。両唇音、歯茎音、硬口蓋音、軟口蓋音、と子音は大きく分類すると以上四つの種類の調音位置から音声を生成している。子音の多くは舌尖で構音するので発話する時に主に舌尖部分の動きを目標として発話運動を制御する。さらに先ほどあげた 4 つの種類の子音から最も舌尖を引っ張る用に制御する歯茎音を分析する。そこで舌尖付近で歯茎子音が母音に対してどれほど影響しているかを母音について行った分析方法で行った。このデータの数は先ほどと同じ文章の数の約 360 文から母音のセグメント各 6 点より 300 から 550 個のデータで構成した分布である。両端を歯茎子音のみで分布を求めたそのため、両端が母音の分布よりサンプル数が少ない分布になってしまった。さらに分布を見ると音素 /u/ と /o/ の分布が二つに別れていることがわかる。これは構音する際、唇を使うためであり、主な調音位置が二つ以上の発生器官の部位からなり、部位が変動するため分布が別れたと考えられる。また両端が母音の時の分布と同じように母音 /i/ の分布の分散は小さく、他の母音より拘束の度合いが大きいと考えることができる。

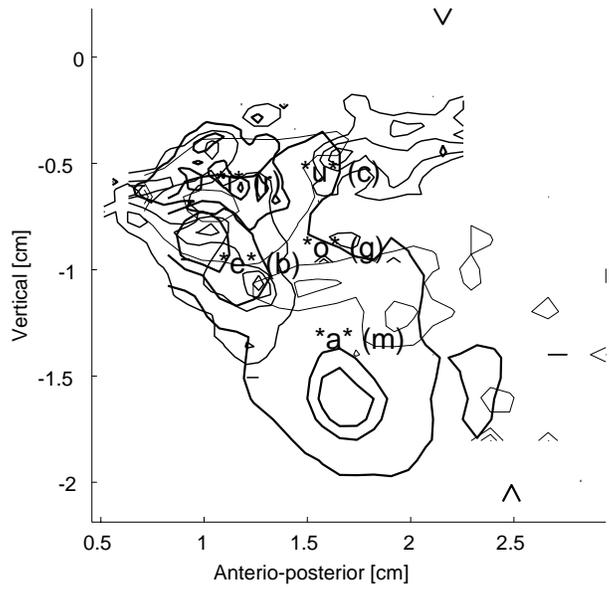


図 3.3: 舌尖付近のパラメータの調音位置での 5 母音の分布、等高線の概形

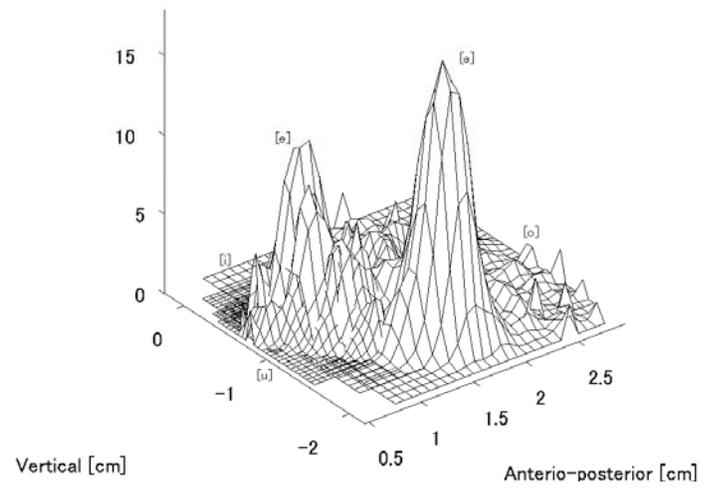


図 3.4: 舌尖付近のパラメータの調音位置での 5 母音の分布、斜めから見た分布の概形

第4章 子音と母音の調音運動の分離

4.1 調音特徴を考慮したスペクトル

党らにより舌における母音の調音運動と子音の調音運動をそれぞれ“搬送波”、“調波”と見なして調音結合のモデル(キャリアモデル)を提案した。このモデルの検証として母音と子音の運動成分を分離しそれぞれの調音運動を再構成することによってモデルの検証と母音と子音の相互作用を検討する。

上記の分析を行う際、特定の発話文を用いた場合、音素のバランスの問題は回避できない。そのため、すべての音響環境での舌の調音運動を考察すべきである。本研究では周波数加算平均の手法により平均的な音素環境を作成し調音運動の波形を再構成することによって多数の音素環境を考慮に入れた。そこで、舌の動きに注目して舌の観測点 T_1, T_2, T_3, T_4 を分析に用いた。日本語 352 文の音声発話に対して約 2 秒間のデータを抽出し、観測点における調音運動の波形にフーリエ変換によりスペクトルを求め、周波数上で加算平均を行い平均的なスペクトルを求める。このスペクトルにはすべての音素環境を考慮した総合的な音響環境とみなすことができる。

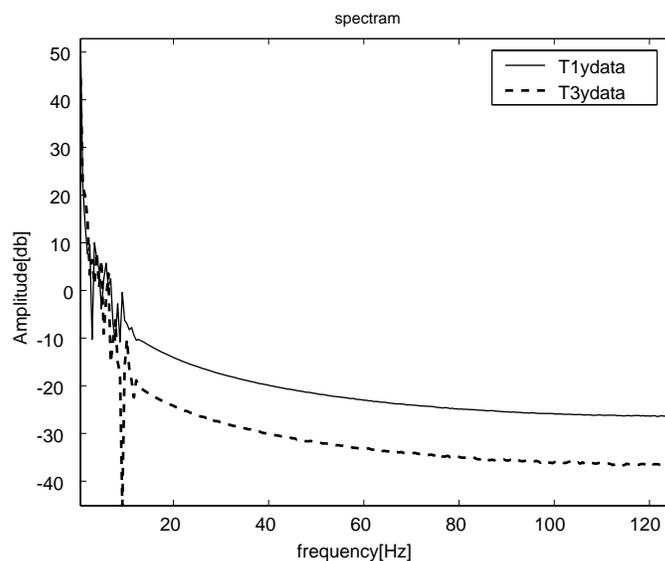


図 4.1: 周波数加算平均したスペクトル

4.2 調音運動の再構成

子音の多くは舌尖で構音するので発話する時に主に舌尖部分の動きを目標として発話運動を制御する。それに対して母音の場合、発話運動が舌全体の動きに関わっている。このことを念頭に置き、観測点 T_1 により子音の調音運動、 T_3 は母音の調音運動を表す。これらのデータを用いてそれぞれ子音の発話運動と母音の発話運動の分析を行う。周波数加算平均により求めたスペクトルから複素フーリエ級数展開を施すことによって周波数領域から時間領域での調音運動の波形を再構成した。

$$T_k y(t) = a_0 + \sum_{i=1}^{N/2} a_i \cos(2\pi f_0 i t) + b_i \sin(2\pi f_0 i t) \quad (4.1)$$

$$a_i = 2\text{real}(C_i)/N \quad b_i = -2\text{imag}(C_i)/N \quad (4.2)$$

$$a_0 = C_0/N \quad k = 1, 2, 3 \quad i = 1, \dots, N/2 - 1 \quad (4.3)$$

C_n は周波数の加算平均により求めた値、 N はフーリエ変換のポイント数、 f_0 は周波数分解能である。周波数上の加算平均した複素数を用いて複素フーリエ級数展開した観測点 T_1, T_2, T_3 の波形を図 4.2 に示す。さらに再構成した波形から各観測点の速度を求めた波形を図 4.3 に示す。図 4.2 の波形はすべての音素環境を考慮した調音運動を示している。図 4.3 は再構成した波形の速度の変化を示している。発話運動の速度がゼロとなった時、発話器官は安定な状態となっているので、その部分はそれぞれ母音と子音の中心となっている。このことから図 4.2 での波形のピークの位置はなんらかの音素の中心であるといえる。垂直方向での舌尖（観測点 T_1 ）と舌背（観測点 T_3 ）の動きは多くの場合逆位相となっている。また舌尖は舌背より垂直方向で運動が若干速くなっていることがわかる。

4.3 中心点 T_2 に対する相対的な調音運動

多数の音素環境を考慮した場合再構成した観測点 T_1, T_2, T_3 の振る舞いを考察すると、図 4.2 より観測点 T_3 と観測点 T_1 を比較すると垂直方向に逆位相で動いている。また観測点 T_2 の振る舞いについては、 T_3 と T_1 との中心点として振る舞っているようである。そこで T_3 と T_1 との関係をあきらかにするため、 T_3 と T_1 からそれぞれ T_2 を引いた波形を求めた。その関係は次式のようにになっている。

$$T_k y(t) = T_k y - T_2 y \quad k = 1, 3 \quad (4.4)$$

先行研究で明らかになったように、舌尖と舌背とは水平の前後運動においては相関が高い。そのため本研究では、垂直方向の運動成分のみを取り出して考察した。多数の音素環境を考慮した平均環境では、舌尖と舌背の動き幅はほぼ同じである。多くの場合、舌尖と

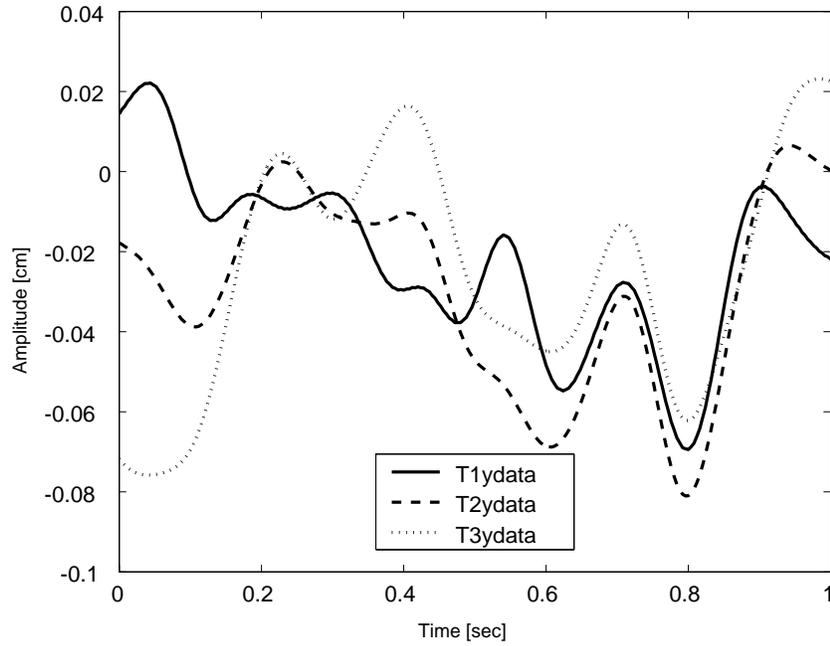


図 4.2: 調音運動の T1, T2, T3 における再構成した波形

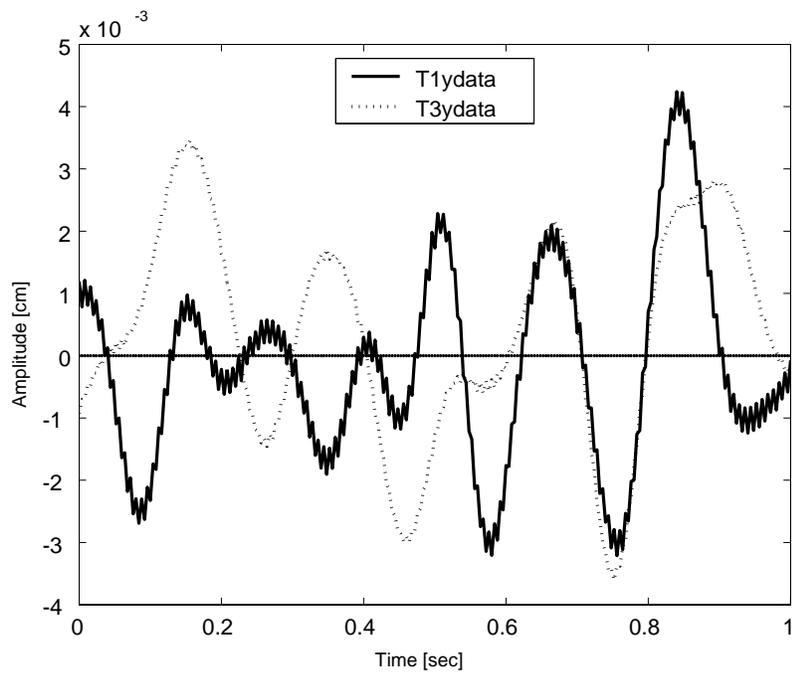


図 4.3: 再構成した波形 T1, T3 の速度

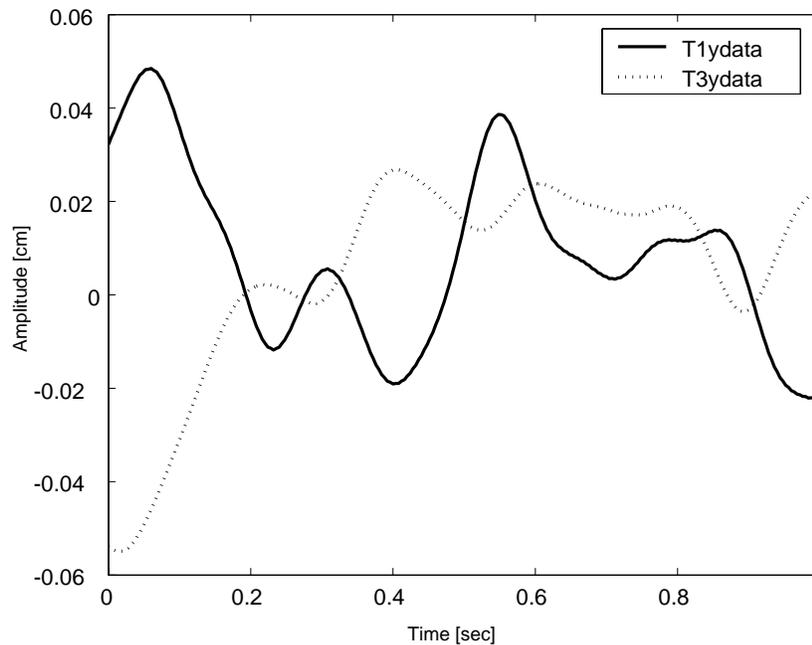


図 4.4: 中心点 T2 に対する相対的な調音運動

舌背は逆位相となっている。それは主に調音結合の影響と考えられる。同位相になっている区間では舌尖子音を含めていない音節の調音運動にあたる可能性がある。運動の速度を考察したところ、垂直方向で舌尖は舌背より運動が若干速くなっていることがわかった。舌尖と舌背の共通の成分を取り除いた場合、つまり T_1 と T_3 から T_2 を引いた場合、母音の調音運動 (T_3) は穏やかに動いて、舌尖子音の運動 (T_1) は母音の運動より速やかに動いていることを明らかにした。図 4.4 に示した波形から、音声の発話運動は速やかな子音の調音運動は穏やかな母音の調音運動の上に重畳して統合したものであることが容易に理解できる。党らが提案した調音のキャリアモデルでは母音の調音運動を緩やかに変化する搬送波、子音の調音運動を速やかに変化する調波として考えていた。よってこの分析とほぼ一致していることがわかった。

第5章 調音結合を考慮した時における舌の動きの模擬

5.1 Carrier Model

音声には2種類の調音結合があげられる。1つは左から右の(LR:キャリーオーバー)もう1つは右から左の(RL:先行調音)である。キャリーオーバーとは、先行音素の特徴が後続音の調音に影響を与える現象を示す。また先行調音とは音素の部分的な調音特徴が時間的に先行して達成される現象のことである。連続音素列を発話を行う際、舌、下顎および口唇などにおいて前音素の調音位置が前後の音素の影響を受ける。影響を受けるために本来の調音目標と異なった位置に調音器官が到達し発話する。これは話し手が”look-ahead”で脳の中で発話計画を行っているために起きる。この過程を、Henkeは音素セグメントのモデルとして提案した[5]。またOhmanは、調音結合を音声のスペクトルからVCV音素列において母音の調音運動とその母音の調音運動の影響を受けた子音の調音運動が重畳しているとしてモデル化を行った[6][11]。

また我々は調音結合の分析を行い、舌尖の動きは、水平の方向では舌背部の動きに高い相関をもっていたが、垂直の方向では独立であった[9]。この分析結果はOhmanの分析結果と同じ結果であり、正しいと考えられる。また隣接した音素で調音結合の影響を定量的に分析するためにRecasensらは舌背の動きと第2フォルマントを分析し、母音(V)と子音(C)からなるVCV音素系列における子音の調音位置が前後の母音の調音位置と関係があることを見出し、その関係の度合いをDAC(調音位置の拘束の度合)のパラメータとして記述した[8]。この研究では7つの子音と母音/a/と/i/からなる音素列の分析を行った。その結果隣接した音素からの度合いを3つのレベルに定義した場合、この拘束の度合いに依存することが解明された。さらにこの結果から拘束の度合いの概念を用いる調音結合のモデル化を試みた。

一般に、音声は、子音と母音から成る流れと見なすことができる。また母音と子音の調音結合の影響は一般的に母音の方が子音より影響がある。したがって速やかな子音の調音運動に対して相対的に影響力が強い穏やかな母音の調音運動がある。これは子音と母音の調音運動の分離の結果からも穏やかな母音の調音運動と速やかな子音の調音運動という結果と同じである。

このモデルは従来から設定していた調音位置を脳の発話計画に従って調音結合を考慮した調音位置に再計画するというものである。一般的に母音と子音の調音結合の影響は母音の方が子音より強い。さらに再構成の結果から、速やかな子音の調音運動に対して相対的

に影響力が強い穏やかな母音の調音運動と考えられる。このようなメカニズムから、キャリアモデルは速やかな子音の調音運動と穏やかな母音の調音運動とを別々に考え、この2つの波形を足し合わせることによって母音と子音を含めた調音運動を再構成する。子音の

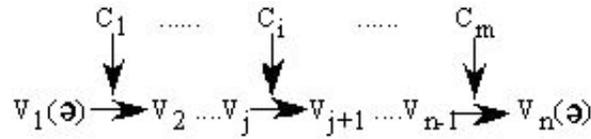


図 5.1: キャリアモデルの概念

調音位置は “tug-of-war” の関係から隣接する音素から影響を受ける。そして母音についての仮定の調音位置を子音の調音位置の中に設定する。また、脳の中で音素の先読みが行われていると考えられるので後続音素に重みがかかった調音位置になると考えられる。(V は母音の調音位置、C は子音の調音位置)

$$G_i = \alpha V_j + \beta V_{j+1} \quad \alpha \leq \beta \quad (5.1)$$

通常、子音は母音とともに発話するので主に前後の母音の影響を受けて発話されると考えられるので先ほど求めた仮定の調音位置の影響を受けた調音位置になる。また、音素に対しての拘束の度合いを考慮した調音位置を考慮にいれ、子音についての調音結合を考慮した調音位置を考える。次のような式が考えられる。

$$C'_i = \frac{(d_{ci}C_i + G_i)}{d_{ci} + 1} \quad (5.2)$$

次に母音の調音結合を考慮した調音位置を考える。子音の影響を受けて前後の母音の調音位置も変化する。このことから母音に拘束の度合いをかけたものと調音結合を考慮した子音の調音位置の影響を受けて以下のような式を考えることができる。以上の手順で調音結合を考慮した調音位置を合成する。さらに $\tau = 0.5$ 、 $\gamma = 0.5$ として本研究を行った。

$$G'_i = \frac{\gamma d_{ci} C'_i + \tau d_{vj} v_j}{(\gamma d_{ci} + \tau d_{vj})} \quad \tau \approx \gamma \quad (5.3)$$

以上このようなモデルを用い調音結合を考慮した調音位置を合成する [3]。しかしこのキャリアモデルを生理学的モデルに適用するには各音素に対しての拘束の度合いを求めなくてはならない。

5.2 母音についての拘束の度合い

調音波形から母音の調音波形と子音の調音波形、またモデルの概念にしたがって母音と子音の拘束の度合いを分けて考える。今回は母音についての拘束の度合いに焦点を当てて

調査を行った。人が発話する際音素の先読みが行われているので、前の音素と後ろの音素では影響の度合いが変わってくる。このことを考慮に入れ今回は前後同じ音素にして音素間の拘束の度合いだけに着目した。 $V_bCV_cCV_b$ (Vは音素の母音、Cは音素の子音)の音素列に注目し前後の母音の影響だけを考える。両端の母音を同じ音素にし調音位置の平均の位置を求めた。図 5.2 から、両端の母音が中央の母音を本来の中心の位置から引っ張るような影響を与えていると考えられる。

そこで表 5.1 のようなマトリックスを求めた。これは各 5 母音間での音素列から両端を選択した平均値を求めその距離をからなるものである。これにより中央の母音が両端の母音にどれだけ影響されてるかを定量的にみた。結果から音素 /i/ については値が 0.13cm より小さい母音が影響して 0.23cm より大きく動いている。このマトリックスは行で見ると中心の母音が両端の母音にどれだけ影響されているかが定量的に見れる。また、列で見ると両端の母音が中心の音素にどれだけ影響しているかということがわかる。このことから列の値が拘束の度合いを意味している。この値が大きければ隣接する音素に強く影響する。そこでこの列の値を正規化して DAC(拘束の度合い) を求めた。値を表 5.2 に示す [10]。

表 5.1: 舌背付近のパラメータを用いて求めた平均値間の距離

$V_b \backslash V_c$	/a/	/i/	/u/	/e/	/o/
/a/	0.152	0.058	0.106	0.124	0.107
/i/	0.288	0.059	0.236	0.281	0.389
/u/	0.238	0.122	0.196	0.180	0.184
/e/	0.085	0.130	0.117	0.029	0.101
/o/	0.226	0.016	0.143	0.092	0.201

表 5.2: 母音についての拘束の度合い

	/a/	/i/	/u/	/e/	/o/
DAC	1.2	2.7	2.0	1.0	1.5

5.3 歯茎子音について拘束の度合い

各分布で $V_bCV_cCV_b$ (Vは音素の母音、Cは音素の子音)の音素列に注目し CV_cC の前後の歯茎子音の影響だけを考える。歯茎子音についての拘束の度合いに焦点を当てて調査を行った。サンプル数が少なかつたため各母音ごとに歯茎子音の距離を求めた。求めたものを図 5.3 から図 5.7 に示す。さらに選択した音素の組によってサンプル数が異なるため、

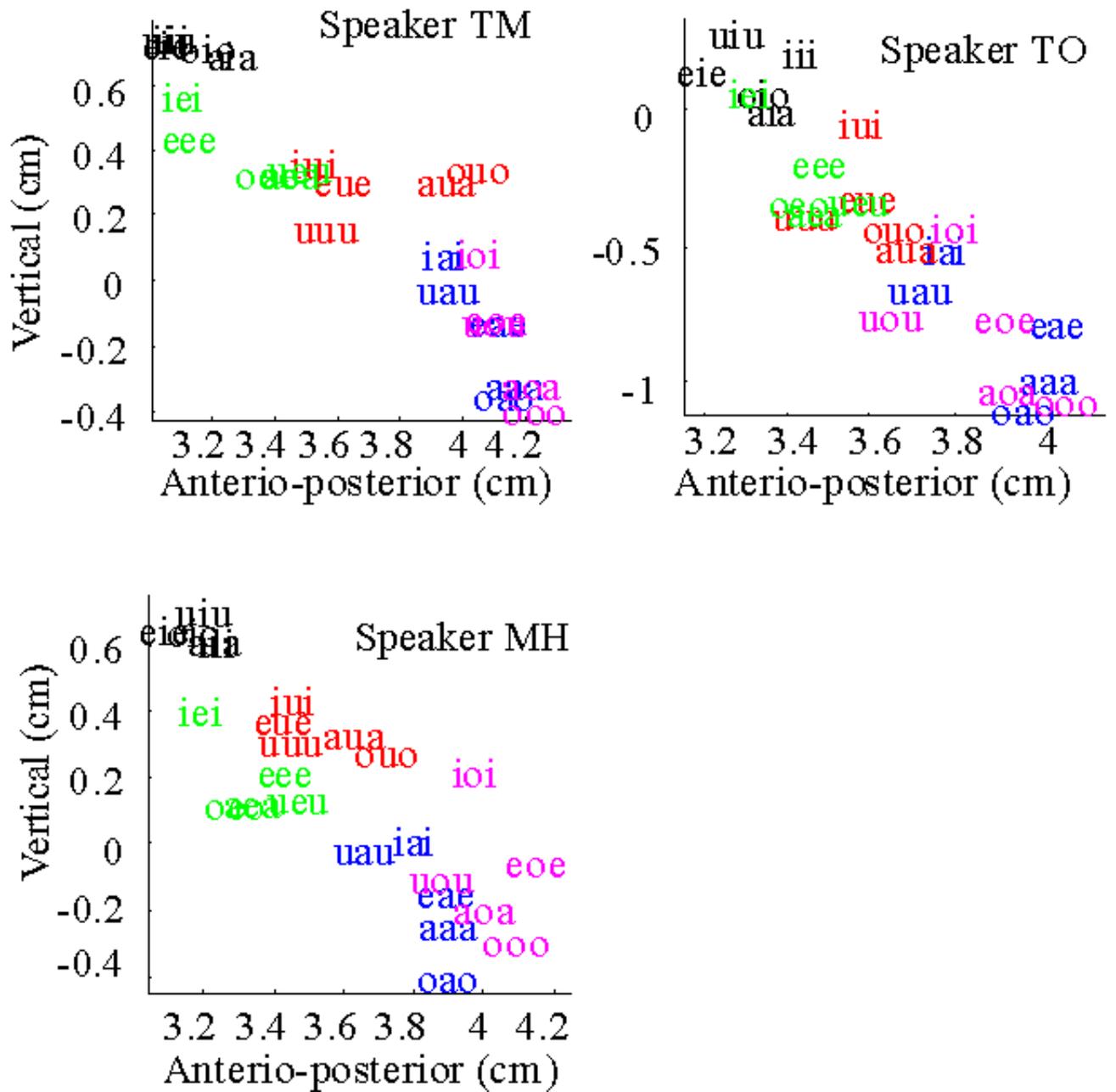


図 5.2: 舌背付近の調音位置での 5 母音の平均

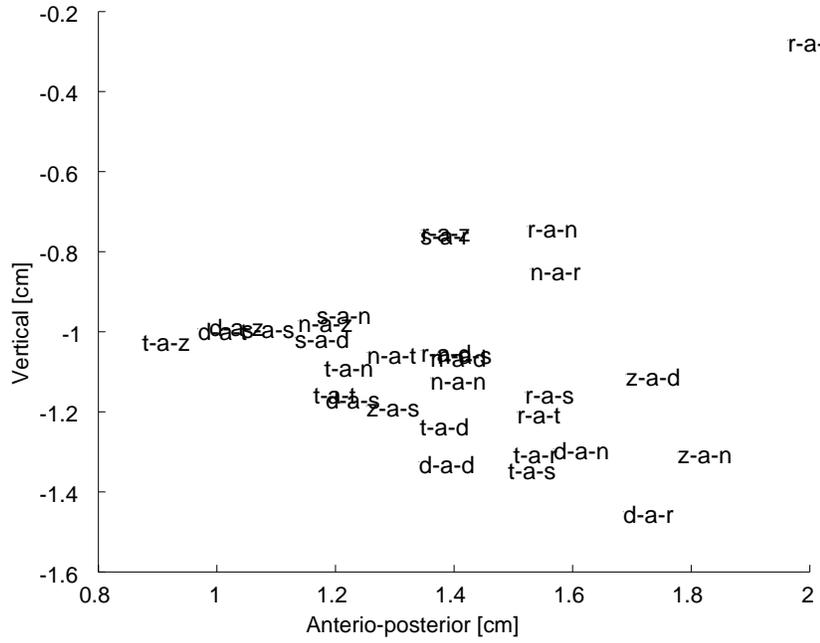


図 5.3: 舌尖付近の調音位置での母音/a/の平均値の分布

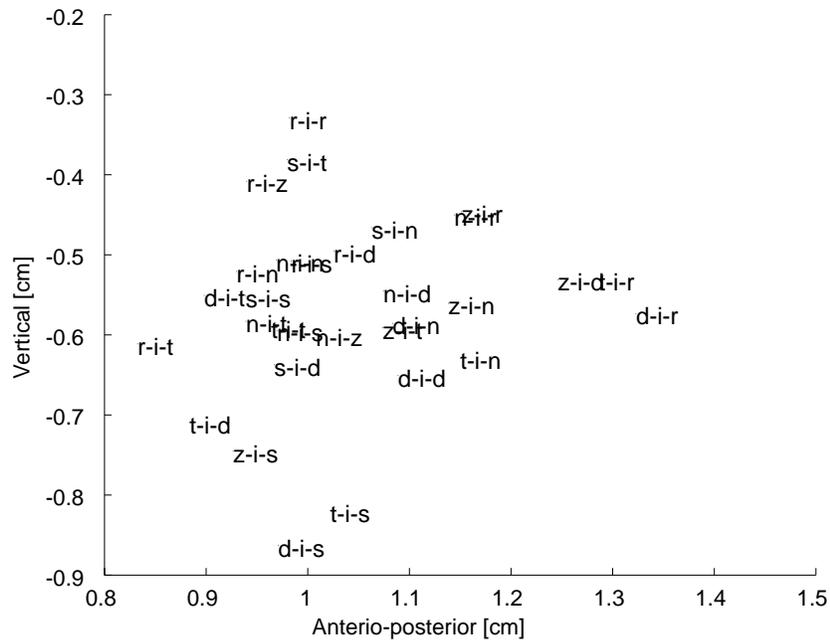


図 5.4: 舌尖付近の調音位置での母音/i/の平均値の分布

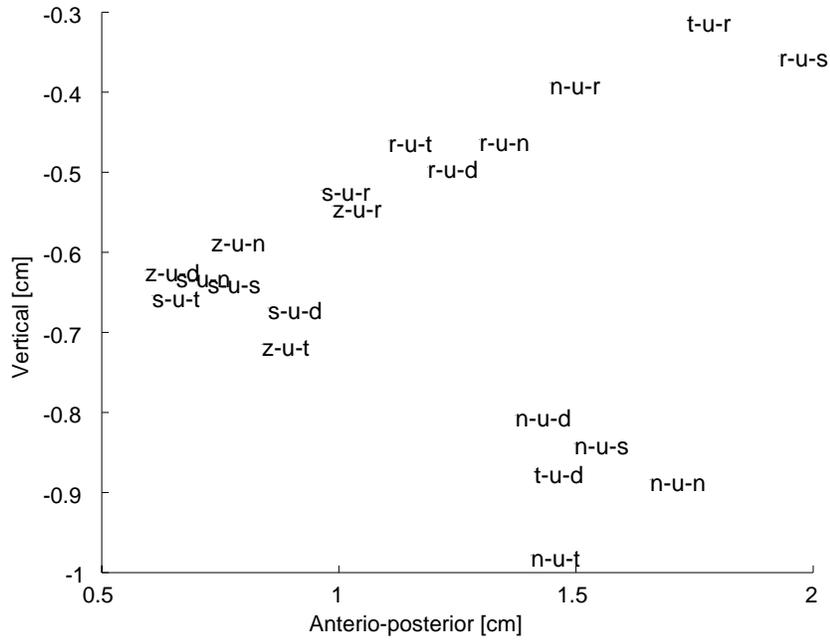


図 5.5: 舌尖付近の調音位置での母音/u/の平均値の分布

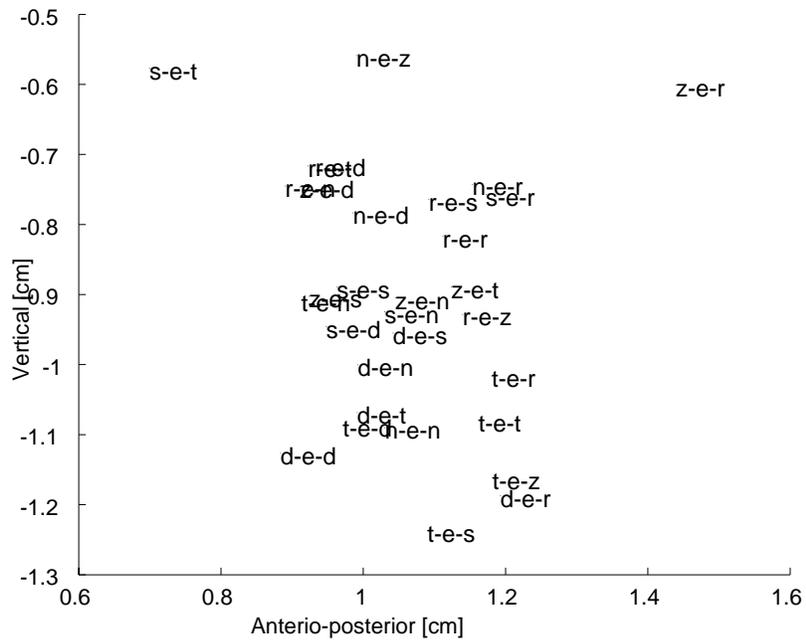


図 5.6: 舌尖付近の調音位置での母音/e/の平均値の分布

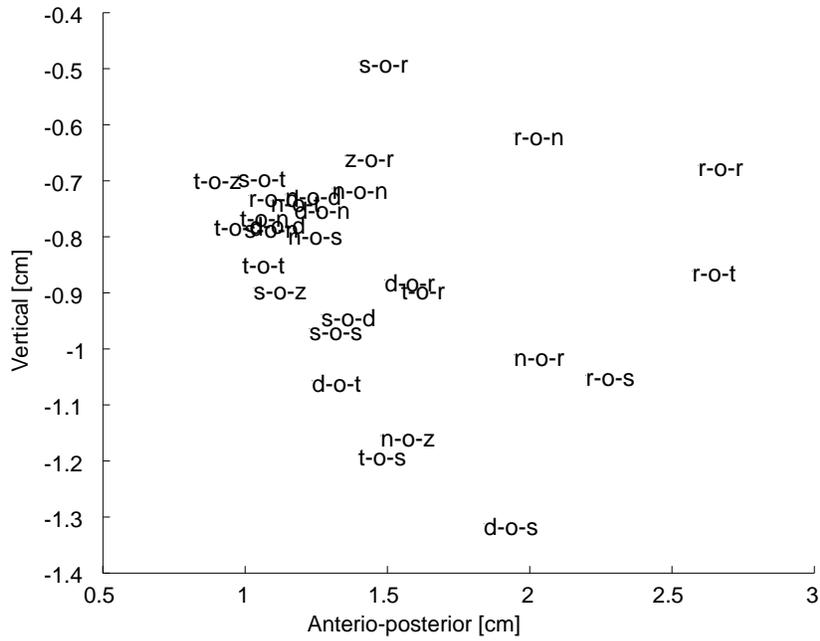


図 5.7: 舌尖付近の調音位置での母音/o/の平均値の分布

両端の子音を決めた分布の個数を均等に扱うために両端の子音を決めた組全てにサンプル数を掛け合わせその値を足し、さらに全サンプル数で割ってやることによって平均の値を均等にした。選択したサンプルの組に対しての距離を求めるために平均値間の距離と選択したサンプル数を掛け合わせてサンプルの組に対しての距離を求めた。

$$d_i : \text{距離} \times \text{選択した音素のサンプル数} \quad (5.4)$$

距離は先ほど母音について求めた方法と同じ方法で求めたものでサンプル数は前後の歯茎子音を選択しそのサンプル数とする。

$$N : \text{全歯茎子音のサンプル数} \quad (5.5)$$

$$n : \text{歯茎子音のサンプルの組の数} \quad (5.6)$$

$$D = \frac{1}{N} \sum_{i=1}^n d_i \quad (5.7)$$

さらに歯茎子音全体に対して1つの母音に影響しているかを調べるためにサンプルの組に対して求めた平均間の距離を全サンプルの組で足し合わせて歯茎子音全サンプルの組に対しての距離になる。その距離を全サンプル数で割ってやることによって全サンプルを含めた歯茎子音を均等に考慮した平均間の距離にした。この結果を表 5.3 に示す。この値

は各母音が歯茎子音に影響を受けてどれだけ遠くの平均値になっているかを示している。この遠さつまり距離の長さが大きければ大きいほど歯茎子音が各母音に影響しているかを示すことになる。また値が小さければ歯茎子音に影響をあまり受けないということになる。さらに音素/u/と/o/ に対しては調音を行う際、一般的に舌以外にも口唇も調音に使うので分布が割れてしまった。また、歯茎子音の子音において、舌尖付近では子音はあまり母音に影響を与えないという数字になった。しかし母音/a/については大きな変動を示す結果だった。これらの距離を母音についての拘束の度合いを求めたとの同様に子音についても拘束の度合いを求めなくてはならない。

表 5.3: 舌尖付近のパラメータを用いて求めた平均値間の距離

C \ V	/a/	/i/	/u/	/e/	/o/
C:歯茎子音	0.289	0.020	0.407	0.007	0.246

そこで母音で求めた拘束の度合いと同じ用に扱うために母音で求めた拘束の度合いを見て同じような振れ幅であれば拘束度合いを同じとして今回はこの値を扱った。また母音/u/と/o/のような主な調音に口唇も使うようなものに関してはその値を除いて他の音素で最大になっているものを使った。以上で母音と母音の拘束の度合いと、子音と母音の

表 5.4: 歯茎子音についての拘束の度合い

C \ V	/a/	/i/	/u/	/e/	/o/
C:歯茎子音	2.7	1	2.7	1	2.7

拘束の度合いが求まったのでキャリアモデルを用いて模擬実験を行い調音結合を考慮した調音位置になるかを調査した。

5.4 パラメータの評価

ここではキャリアモデルで求めた調音目標から実際の生理学的発話モデルで求めた調音位置を使って分布を求めた。さらにこの求めた分布の整合性を確かめるために分布と重ね合わせてどのくらい重なり合うかを調べた。また発話モデルを構築する際のデータと今回分析のデータが異なる。発話者が異なることから声道における個人差を考慮に入れる必要がある。そこである程度の補正が必要であると考えた。今回、模擬値の分布と観測データの分布の中心点をそろえた。

$$\mu : \text{シュミレーション結果の平均} \quad (5.8)$$

$$\bar{T}_1 : \text{観測データの値の平均値} \quad (5.9)$$

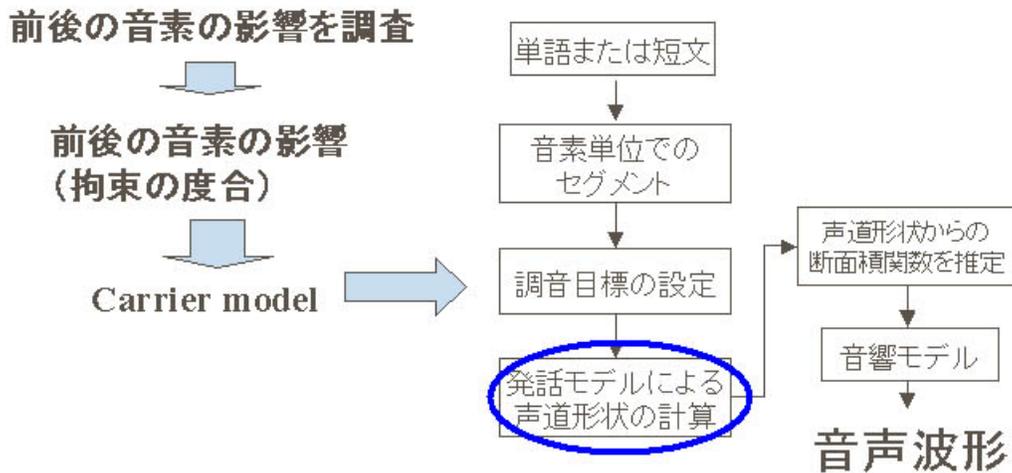


図 5.8: 評価の手順流れ図

$$\lambda = \bar{T}_1 - \mu \quad (5.10)$$

$$\lambda = \bar{T}'_1 + \lambda \quad (5.11)$$

$$\bar{T}'_1 : \text{シュミレーションを行った値} \quad (5.12)$$

このようにして求めた値を実際の観測値とシュミレーションを行った値とを重ねて表示することによってシュミレーションの値の整合性を確かめた。丸でプロットした点がシュミレーションの値で等高線が観測データから求めたものである。このシュミレーションでは発話の際の音素の長さを一定にして発話の際の音素の中心から2ステップ分をプロットした。これは発話の際文章によって音素の長さが変化してしまうためである。シュミレーションで与えた音素は本研究で分析を続けてきた VCVCV 音素列を与えた。たとえば音素/a/ならば aCaCa となるような音素列で、子音には実際の観測データにあった歯茎子音の組を用いてシュミレーションを行った。観測データは VCVCV 音素列から中心の CVC、歯茎子音を求める時に用いた観測データを用いて行った。

シュミレーションの結果は図 5.19 から 5.13 に示す。以上のような結果になったシュミレーションの値が分布と完全に当てはまらなかった理由としてはモデル構築の際のデータの発話者と今回用いたデータの発話者が異なるため筋肉また骨が異なる。値としては初期値を今回用いたデータに合わせたが、人によって発話に個人性があることから完全な一致は困難である。また、音素によって発話を行う際に主に動かす部位が舌以外にも存在するが、現在の発話モデルには発話器官すべてを同時に制御するようなモデルはないため音素/u/、/o/の用に口唇によって分布が変動してしまうものに対応できない。また実際の観測データの音素の長さは音素環境によって変化する。また話すごとに長さが変化してしまうために観測データと現在のモデルのシュミレーションでは全く同じデータをつくるの

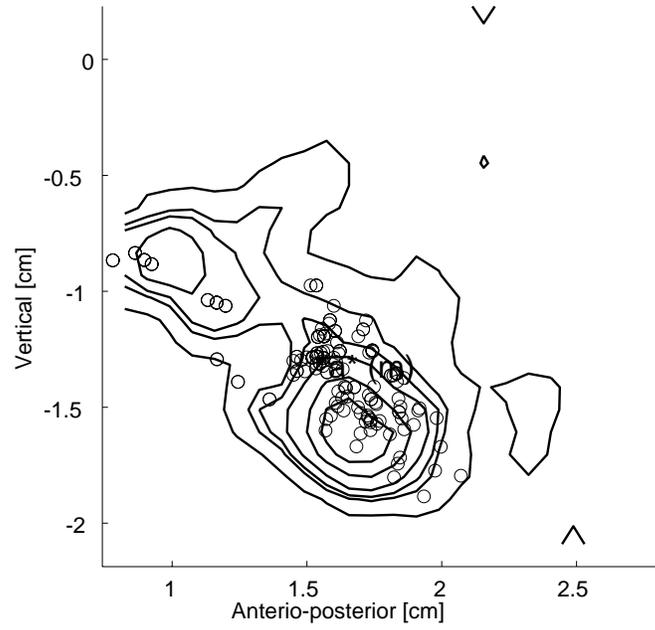


図 5.9: VCVCV の音素列を与え中心の母音/a/についての舌尖付近の調音位置

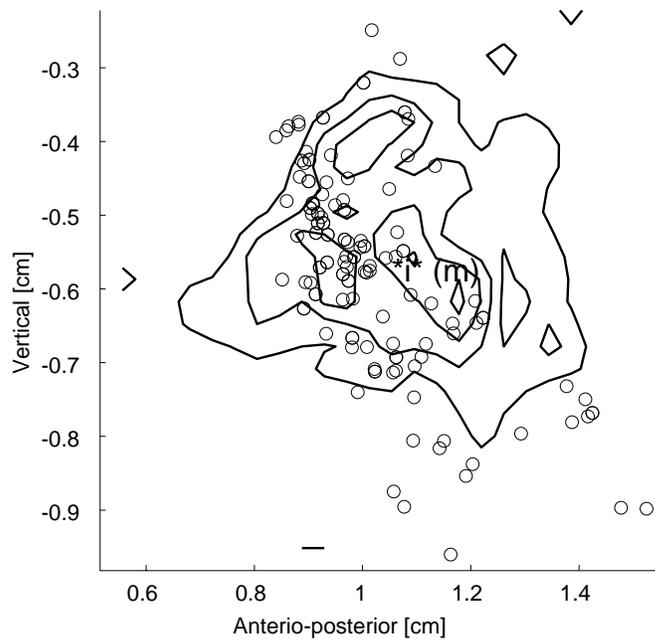


図 5.10: VCVCV の音素列を与え中心の母音/i/についての舌尖付近の調音位置

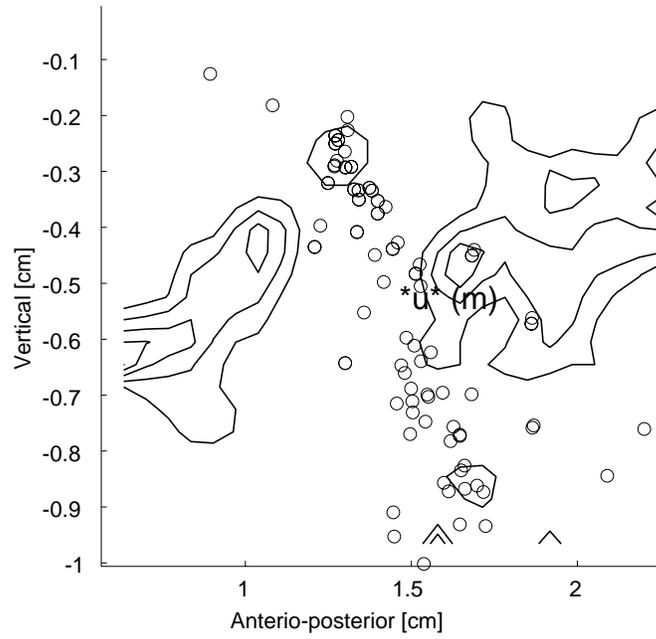


図 5.11: VCVCV の音素列を与え中心の母音/u/についての舌尖付近の調音位置

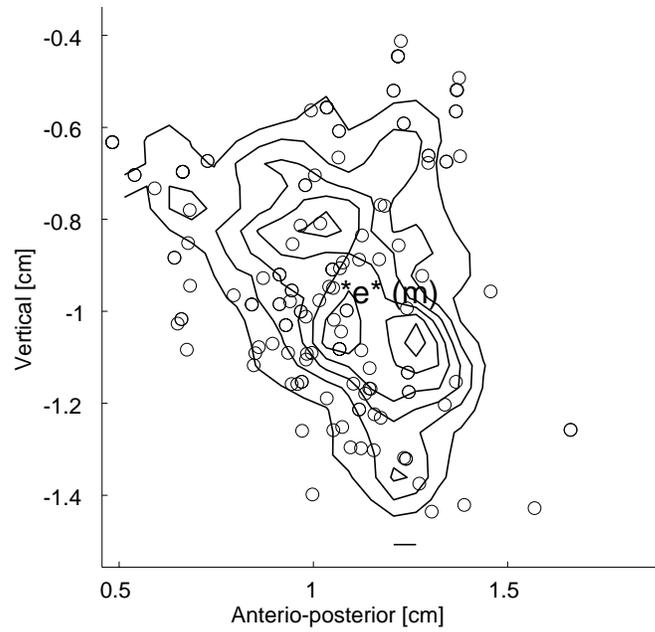


図 5.12: VCVCV の音素列を与え中心の母音/e/についての舌尖付近の調音位置

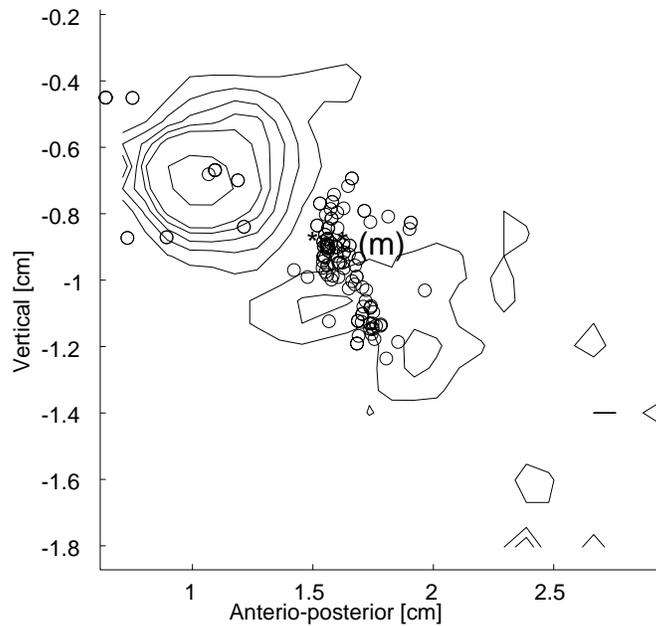


図 5.13: VCVCV の音素列を与え中心の母音/o/についての舌尖付近の調音位置

は困難である。これらの原因で完全にマッチングが行えなかったと考えられる。 以上のような結果になったシュミレーションの値が分布と完全に当てはまらなかった理由としてはモデル構築の際のデータの発話者と今回用いたデータの発話者が異なるため筋肉また骨が異なる。値としては初期値を今回用いたデータに合わせたが、人によって発話に個性があることから完全な一致は困難である。また、音素によって発話を行う際主に動かす部位が舌以外にも存在するが、現在の発話モデルには発話器官すべてを同時に制御するようなモデルはないため音素/u/、/o/の用に口唇によって分布が変動してしまうものに対応できない。また実際の観測データの音素の長さは音素環境によって変化する。また話すごとに長さが変化してしまうために観測データと現在のモデルのシュミレーションでは全く同じデータをつくるのは困難である。これらの原因で完全にマッチングが行えなかったと考えられる。

第6章 全体の考察

本研究ではシュミレーションで用いたキャリアモデルの概念が実際の観測データに正しいかを確認した。手法としては、調音運動の波形を周波数加算平均により多くの音素環境を考慮して代表的な調音運動を再構成した。その代表的な調音運動の波形を用いて母音と子音の調音運動の分離を図った。その結果、音声の発話運動はより、速やかな子音の調音運動は穏やかな母音の調音運動の上に重畳することによるものとみなすことができる。その結果、調音のキャリアモデルの考えはほぼ正しいことがわかった。しかし今回の分析では実際の調音データにより子音の調音運動と母音の調音運動の完全な分離ができなかった。また VCVCV 音素列を分析することによって母音が母音に及ぼす影響を拘束の度合いとして定量的分析を行いまた、子音についても拘束の度合いを求めた。その求めた値をキャリアモデルを用いて調音目標を設定しその調音目標に対してモデルがどのように振る舞うかをシュミレーションを行い確かめた。シュミレーションの結果から /a/、/i/、/e/ に対しては妥当であるような振る舞いをする事がわかったが /u/、/o/ のような、口唇も調音を行う際に変動してしまうような1つの発話器官だけで主に発話するものには生理学的発話モデルが対応していなかったために確かめることができなかったと考えられる。

第7章 結論

7.1 本論文で明らかにされたことの要約

本研究では調音運動に基づいた調音結合を分析を行った。その結果過去の先行研究と同じ結果になり、またキャリアモデルの概念の、子音は速やかな調音運動で母音は穏やかではあるが相対的に影響力の強い調音運動として考えることは正しいと考えられることが明らかにされた。さらにキャリアモデルを用いることによって調音結合を考慮した調音位置を音素別ではあるがシュミレーションを通じて再現できることを明らかにした。

7.2 今後の課題

本研究で用いたデータの数は十分とは言えず特に子音についての拘束の度合いを求めるとき、音素別での評価が困難になった。このため、EMMA を使ってより多くの音素環境を考慮したデータの採取を行いより正確な調音位置の分布を求める必要がある。また子音の調音運動が”調波”で母音の調音運動が”搬送波”として構成され、重畳されることによって観測される調音運動の波形になることを本研究では完全に行えなかったので調音運動の完全な分離を行うことにキャリアモデルの定式化の再考を行う必要がある。さらに生理学的発話モデルには発話器官すべてを同時に制御することができない。このため/u/、/o/のような、口唇も調音を行う際に変動してしまうようなものに対応できない。そこで口唇も制御できるような生理学的発話モデルに改良する必要がある。

謝辞

本研究を行うにあたり、多大なるご指導、御鞭撻を頂きました党 建武 教授に深く感謝の意を表します。また、本研究を進める過程において、有益な助言を頂きました赤木 正人 教授、鵜木 裕史 助手に心より感謝します。また、多大なるアドバイスを下さった石本 裕一 氏、羽二生 篤 氏、西本 博則 氏をはじめ、議論に御協力いただいた赤木研究室の皆様にご感謝致します。最後に、本研究に対する議論とアドバイスを頂いた、全ての皆様に厚く感謝いたします。

参考文献

- [1] 古井貞熙, デジタル音声処理, 東海大学出版会, 1985.
- [2] 田淵行則, 前川喜久雄, 窪園晴夫, 本多清志, 白井克彦, 中川 聖一 音声, 岩波講座 言語の科学. 2, 岩波書店,1998
- [3] 日本音響学会, 音響用語辞典, コロナ社, 1988
- [4] 今井 聖, 音声信号処理, 森北出版株式会社, 1996
- [5] Henke, L, “Dynamic articulatory model of speech production using computer simulation.”, MIT doctor thesis, 1996
- [6] Ohman, S, “Coarticulation in VCV utterance Spectrographic measurements”, Acoust. Soc. Am.39, 151-168 1966
- [7] Dang, J. Honda, K. and Pascal, P “Implement of Coarticulation in Physiological Articulatory Model”, ICA.25, 5 2004
- [8] Recasens,D. Pallares, M. and Fontdevila, J. “A Model of lingual coarticulation based on articulatory constraints”, J. Acoust. Soc. Am.102, 544-561 1997
- [9] Dang, J. Honda, M. and Honda, K “Investigation of coarticulation in continuous speech”, Acoustical Science and Technology 25,5, 318-329 2004.
- [10] Dang, J. Wei, J. Suzuki, T. Honda, K. Pascal, P. and Honda, K “Investigation and Modeling of coarticulation in speech productin”, ICSLP,51 2004
- [11] Ohman, S “Numerical models of coarticulation” J. Acoust. Soc. Am. 40, 310-320 1967.
- [12] Okadome, T. and Honda, M. “Generation of articulatory movements by using a kinematic triphone model” JASA, 453-463 2001.