

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

発話時における調音運動に基づいた調音結合の分析

鈴木 丈晴 (310054)

北陸先端科学技術大学院大学 情報科学研究科

2005 年 2 月 10 日

キーワード: 調音結合、調音運動、音声生成、調音モデル.

1 はじめに

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。

本研究では、我々は調音結合を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。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る。また調音データから調音結合モデルの確かめを行っていない。調音結合モデルの未定であるパラメータを、調音データから隣接する音素との関係の度合いとして求め、調音結合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調音結合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。

2 子音と母音の調音運動の分離

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調音結合モデルの検証と母音と子音の相互作用を検討する。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。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。

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

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

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

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。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。さらに再構成した調音運動の波形の速度と観測点 T_1 と T_3 の中心点 T_2 から相対的に見た調音運動の波形より、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することによるものとみなすことができる。

3 キャリアモデル

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

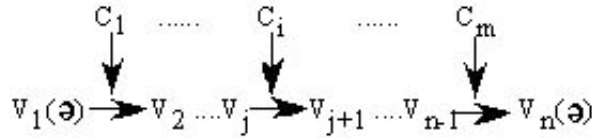


図 1: キャリアモデルの概念

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想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)

$$G_i = \alpha V_j + \beta V_{j+1} \quad \alpha \leq \beta \quad (4)$$

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と考えられるので先ほど求めた仮想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。

$$C'_i = \frac{(d_{ci}C_i + G_i)}{d_{ci} + 1} \quad (5)$$

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。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。

$$V'_i = \frac{\gamma d_{ci} C'_i + \tau d_{vj} V_j}{(\gamma d_{ci} + \tau d_{vj})} \quad \tau \approx \gamma \quad (6)$$

3.1 母音に着目した調音位置の分布を用いた距離

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、脳の中での発話計画から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。 $V_b C V_c C V_b$ (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。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両端を全ての音素で求めた平均値との距離を求めた。この値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。この結果、表を列でみると拘束の度合いと考えられるので正規化を行い、拘束の度合いを求めた。

表 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 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

3.2 歯茎音に着目した調音位置の分布を用いた距離

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ったため母音ごとに歯茎子音の距離を求めた。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、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。

$$d_i : \text{距離} \times \text{選択した音素のサンプル数} \quad (7)$$

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。

$$N : \text{全歯茎子音のサンプル数} \quad n : \text{歯茎子音のサンプルの組の数} \quad (8)$$

$$D = \frac{1}{N} \sum_{i=1}^n d_i \quad (9)$$

母音での距離と比較しその値に近ければ同じ拘束の度合いとした。また値が小さいもの

表 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

に対しては拘束の度合いは1とした。この結果は歯茎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。

表 4: 歯茎子音についての拘束の度合い

DAC \ V	/a/	/i/	/u/	/e/	/o/
DAC	2.7	1	2.7	1	2.7

3.3 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の分析に用いたデータが異なる。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。キャリアモデルに用いた値は今回求めた拘束の度合いと先行研究で行われていた値を使って模擬を行った。

$$\lambda = \overline{T_1} - \mu \quad (10)$$

$$\mu : \text{模擬の結果の平均} \quad \overline{T_1} : \text{観測データの値の平均値} \quad (11)$$

$$\hat{T}_1 = T'_1 + \lambda \quad (12)$$

$$T'_1 : \text{模擬を通して得た値} \quad \hat{T}_1 : \text{補正をかけた模擬の値} \quad (13)$$

模擬の結果から /a/、/i/、/e/ に対しては妥当であるような振る舞いをするのがわかったが /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。また今回求めた拘束の度合いはある程度正しいという結果になった。さらに舌尖付近では子音の影響をあまり受けないことがこの模擬によってわかった。

4 まとめ

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模擬を通じて再現できることを明らかにした。