

規則音声合成のためのパワー制御モデルの検討

濱上 知樹^{†‡}

[†] 千葉大学大学院自然科学研究科

〒263 千葉市 稲毛区 弥生町 1-33

[‡] セコム IS 研究所 音声情報処理研究室

〒181 三鷹市 下連雀 8-10-16

hamagami@sp.isl.secom.co.jp

あらまし VCV を基本単位とした規則音声合成方式において、様々な発話速度に対応することのできる、パワー制御モデルの検討を行なった。パワー制御を行なうにあたり、自然発声におけるパワーのばらつきと、発話速度の変化に伴うパワー時間変化パタンの変動特性について分析した。これらの実験結果をもとに、パワーパタンを、VCV 音節単位で滑らかに変化するマクロパタンと、子音区間において微細な変化をするミクロパタンに分離する方法を提案し、制御パラメータの抽出とデータベース化の方法について検討を行なった。本パワー制御を用いた規則合成音の評価結果から、本手法が、自然なパワー時間変化パタンを合成する上で有効であることが明らかとなった。

キーワード 規則合成, パワーパタン, パワー制御, 発話速度

A Study of Speech Power Control Model for Synthesis by Rule

Tomoki HAMAGAMI^{†‡}

[†] Graduate School of Science and Technology, CHIBA Univ.,

1-33 Yayoi-cho, Inage-ku, Chiba-shi, 263

[‡] SECOM Intelligent Systems Lab., Speech Processing Dept.,

8-10-16 Shimorenjaku, Mitaka-shi, Tokyo, 181

hamagami@sp.isl.secom.co.jp

Abstract This report describes a VCV power pattern control method which can correspond to a variety of speech rates for speech synthesis by rule. An analysis of the relationship between the speech rate and power pattern variations revealed a) a significant correlation between the speech rate and dynamic range of the VCV power pattern and b) both the VC (release segment) and CV (attack segment) power slope value are not affected by variations in the speech rate. Based on these results, a hybrid macro and micro power pattern control model is proposed. To construct this proposed system, the extraction analysis of the macro pattern parameter and the construction of the micro pattern database were examined. From the quality evaluation experiment, we have confirmed that this power control model is effective in improving the quality of synthesized speech.

key words speech synthesis by rule, power pattern, power control, speech rate

1 まえがき

高品質で自然な合成音声を作るためには、音韻継続時間長、ピッチパターン、スペクトルパタンのそれぞれにおいて、バランスのとれた制御を行う必要がある。このうち、音韻継続時間長、ピッチパターン、スペクトルパターンを制御するパラメータについては、古くより多くの検討がなされ、音声合成に必要な様々な知見が得られている。一方、近年の規則合成の技術は、大量の合成単位を用い、単位の選択基準に制御パラメータのバランスを保つ機構を含めたことにより、自然性を向上させてきた^{1),2),3)}。これらの新技術による品質向上を背景として、これまで、あまり検討が進んでいなかったパワー制御に関しても、いくつかの研究成果が報告されている。例えば、母音部中心パワーの値を、基本周波数、隣接音韻、文・呼吸段落位置などとの相関関係をもとに定量的にモデル化したもの⁴⁾、パワー変動に対する音声知覚の許容源を調べ、その精度の中で推定を試みたもの⁵⁾ などがある。これらの試みは、母音部の中心位置におけるパワーを規則から導出することを目的としていた。しかし、パワーの時間変化パターン（以下、パワーパターン）については、合成単位から抽出されたパターンを適当に伸縮して用いる程度の処理しか行なわれておらず、パワーパターンの変化が合成音声の自然性に与える影響については明らかとなっていない。

このような背景に基づき、本報告では、これまで検討が遅れていたパワーパターンについてモデル化を試み、実際にVCVタイプの規則合成システムに適用して得られた結果について述べる。

2 パワーパタンの定義

パワー制御に関しては、音声生成の立場から、音源パワーを扱うものと、出力される音声パワーを扱うものの2通りの考え方が存在する。本稿では、後者の音声パワーを扱うこととし、合成音として実現されるべき単位時間あたりのエネルギーをデシベルの単位で表したものを「パワー」と呼ぶことにする。また、パワーの時系列によって作られるパターンを「パワーパターン」と呼ぶことにする。

母音部パワーの代表値として、母音部のエネルギー重心位置 (CEGV: the Center of Energy Gravity of Vowels)⁶⁾ におけるパワーを用い、これをCEGVパワーと呼ぶことにする。また、子音部において最小のパワーとなる点を、子音ターゲット (TC: Target of Consonant)⁷⁾、最小のパワーの値をTCパワーと呼ぶことにする。図-1にCEGVパワーとTCパワーの分析例を示す。

パワーパターンは、CEGVパワーと、TCパワーを滑らかにつなぐことによって得られる大まかなパターンと、子音

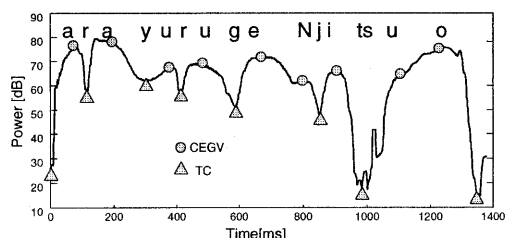


図 1: CEGV パワーと TC パワーの分析例

区間などで瞬時的に変化する細密なパターンにわけて考えることができる。大まかなパターンとは、声帯波を音源とする有声母音、撥音、定常的な有声子音などから作られる、滑らかなパワーパターンである。これをマクロパターンと呼ぶことにする。細密なパターンは、子音部における破裂、摩擦、無声化母音など、声帯波以外の音源から作られている。これをミクロパターンと呼ぶことにする。

3 自然音声のパワーパターン変動について

既に伊藤らにより、母音部中心位置におけるパワー変動が音声知覚に及ぼす影響について、検討が行なわれている。その結果、自然発声の平均的なパワーパターンに対して、4.1dB 程度のばらつきまでが、パワー変動が許容される限界であることが示されている⁵⁾。

一方、実際の自然音声中のパワーパターンが、発話速度や音韻環境などの、発話の条件に依存して、どの程度変動しているかは明らかではない。また、自然性に与えている影響も不明である。そこで、同じ文章を発話速度を変化させて発声した場合のパワーパターンの変動を分析し、自然発声中のパワーパターンの変動傾向について調べることにした。

3.1 実験 1: 自然音声のパワーパターンのばらつき

数文節よりなる文を低速 (5.0mora/s)、中速 (7.5mora/s)、高速 (10.0mora/s) で発声した音声データを収録した。それぞれの発話速度で、10 回づつ連続発声を行なったものを 1 セットとし、各セット間の影響をできるだけ少なくするために、1 セット毎に半日以上の間隔をおいて収録を行なった。1 つの文につき 5 セットの発声データ収録し、合計 10 文章分についてパワーパターンの分析を行なった。各パワーパターンから各母音部の CEGV パワーと、各子音部の TC パワーを目視にて抽出した。これらのパワー値を結んで作られるパターンを、「近似パワーパターン」と呼ぶことにした。各近似パワーパターン中の CEGV パワー、TC パワーのそれぞれについて、(1) 各発話速度データ毎の平均値と変動範囲 (変動 A)、(2) すべての発話速度データの平均値と変動範囲 (変動 B)、を求めた。

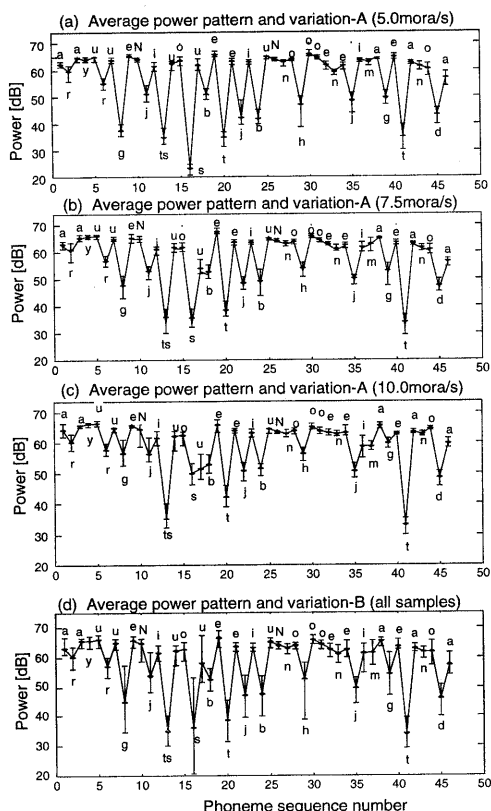


図 2: (a)(b)(c) 各発話速度における近似パワーパタンの平均パターンと変動 A の例, (d) 全発話速度における近似パワーパタンの平均パターンと変動 B の例 (「あらゆる現実をすべて自分の方へねじまげたのだ」)

3.2 結果 1

図-2(a),(b),(c) に, 「あらゆる現実をすべて自分の方へねじまげたのだ」と発声したときの, 近似パワーパタンの各発話速度毎の平均パターン, 及び, 各 CEGV と TC における変動 A を示す。また, 図-2(d) に, 同文発声の全発話速度データから分析された近似パワーパタンの平均パターン, 及び, 各 CEGV と TC における変動 B を示す。10 文章すべてについて分析された変動 A, 変動 B の平均値を図-3 に示す。

3.3 考察 1

実験 1 の結果から, TC, CEGV それぞれの点における変動 A, 変動 B の傾向として, 以下のことが明らかとなった。

1. TC パワーにおいては, 発話速度毎の変動 A の最大値が 5.1dB であるのに対し, 全発話速度における変

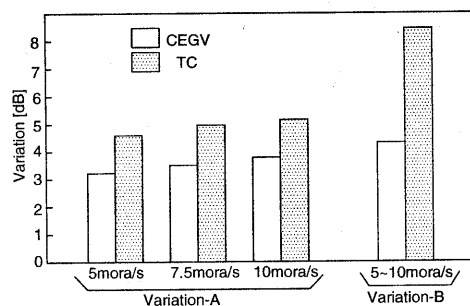


図 3: すべての文章についての変動 A, および変動 B の平均値

動 B は 8.4dB と大きなばらつきが生じていた。これより, TC パワーは, 発話速度に大きく影響を受けており, 発話速度に応じて TC パワーが変化することが予想された。発話速度の上昇にともない, 平均近似パワーパターンにおける「谷」の部分が浅くなる現象が認められ, これが平均近似パワーパタンの形状を大きく変えていると考えられる。

2. CEGV パワーにおいては, 発話速度毎の変動 A の最大値が 3.7dB であるのに対し, 全発話速度における変動 B も 4.2dB と遜色ない安定性が得られている。これより CEGV パワーは発話速度の影響を大きくは受けないことが明らかとなった。
3. TC パワーにおける発話速度毎の変動 A は, 4.8 から 5.1dB の間に収まっているが, 発話速度が速い発声ほど, やや大きくなる傾向がある。音韻種別にみると有声子音のみの変動が 4dB 以下の変動に収まるのに対し, 無声子音の変動は 6dB を越え, やや不安定になる傾向が認められた。
4. CEGV パワーにおける発話速度毎の変動 A は, 3.2dB から 3.7dB の間に収まっているが, 発話速度が速くなるとやや大きくなる傾向がみられる。また, 文献 5) に基づく変動の許容範囲 (4.1dB) 内に CEGV パワーが実現されていることから, 同一発話速度内での CEGV パワーの安定性が確認された。例外として, 無声子音に後続する狭母音 (tsu, su) では, 発話速度が速い場合に, 大きな変動が生じる事がある。

以上の結果から, CEGV パワーからみた TC パワーの落ち込み方に, 発話速度の影響が強く現れる傾向が認められ, 規則合成のパワーパターン制御規則として, 発話速度に応じた CEGV パワーと TC パワーのバランスを保つしくみが必要であると考えられた。また, 文献 5) で示唆されているパワー変動の許容値は, 発話速度の変化に対応した

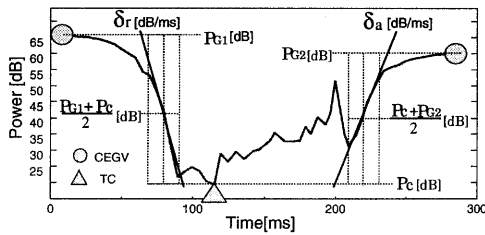


図 4: VCV パタンにおける、立ち下がり角度 δ_r と立ち上がり角度 δ_a

CEGV パワーの不安定性によって生じる自然のばらつき (変動 B) にはほぼ一致していることが確かめられた。

3.4 実験 2：パワーパタンの立ち下がり角度、立ち上がり角度の変化について

文章中の出現頻度が比較的高い 100 種の VCV による無意味単語 (/arVCVra/) を用い、低速 (5.5mora/s)、中速 (7.5mora/s)、高速 (10.0mora/s) の 3 種類の発話速度で各 5 回づつ発声したデータのパワーパタンを分析した。分析されたパワーパタンから、以下の 2 つのパラメータを分析した。

1. 立ち下がり角度： δ_r [dB/ms]

パワーパタンが、先行母音の CEGV パワー (P_{G1} [dB]) と TC パワー (P_C [dB]) の平均値 ($(P_{G1}+P_C)/2$ [dB]) を通過する時刻を中心として、30ms の重み付き窓内 (6 フレーム分) でパワーパタンを 1 次式で近似した場合の傾き。

2. 立ち上がり角度： δ_a [dB/ms]

パワーパタンが、TC パワー (P_C [dB]) と後続母音の CEGV パワー (P_{G2} [dB]) との平均値 ($(P_C+P_{G2})/2$ [dB]) を通過する時刻を中心として、30ms の重み付き窓内 (6 フレーム分) でパワーパタンを 1 次式で近似した場合の傾き。

δ_r および、 δ_a の分析の例を図-4 に示す。前述の 100 種の各 VCV について、同じ発話速度で発声されたパワーパタン毎に、 δ_r と δ_a の平均値を求め、発話速度の変化による δ_r 、 δ_a の変化を分析した。

3.5 結果 2

図-5 に、100 種の VCV について求めた δ_r と δ_a の平均値、及び、標準偏差を発話速度毎に示す。

3.6 考察 2

以上の結果より、以下のことが明らかとなった。

1. 発話速度の変化に伴う δ_r 、 δ_a の変化は小さく、立ち下がり角度、立ち上がり角度は発話速度の影響を大

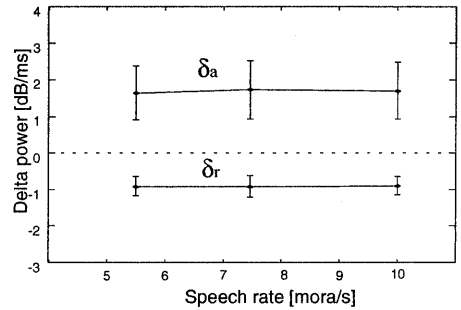


図 5: 発話速度の異なる VCV パタンから分析された立ち下がり角度と立ち上がり角度

くは受けていないことがわかった。

2. δ_r と δ_a の標準偏差は、 δ_r の方が δ_a に比べて小さくなり、立ち上がり角度の方が立ち下がり角度より音韻の影響を受け易いことがわかった。

以上の分析結果から、パワーパタンの概形 (マクロパタン) に関するパラメータのうち、(1) 「CEGV パワー」と「立ち下がり角度、立ち上がり角度」は、発話速度変化に対して変化が少ないこと、(2) 「TC パワー」は発話速度の変化に影響を受けること、が明らかとなった。

4 VCV パワーパタンの立ち下がり角度、立ち上がり角度と、自然性について

前節までの結果から、発話速度に対して、立ち下がり角度、立ち上がり角度を不変量として扱える可能性を示した。これらのパラメータをパワーパタン制御に用いるにあたり、パラメータを人工的に変化させた場合の合成音の変化について検証する必要がある。これについては、既に単独母音の立ち上がり角度を変形させることにより、母音の継続時間長感覚に無視できない変化が生じることを示している⁸⁾。しかし、VCV 音節以上の合成音に関して、立ち下がり角度、立ち上がり角度を変形させた場合に、合成音の自然性にどのような影響が生じるかは明らかではない。ここでは、VCV パワーパタンの立ち下がり角度、立ち上がり角度を独立に変化させた合成音を作成し、これらの弁別限を評価することにより、VCV 音節の立ち下がり角度、立ち上がり角度の変化が自然性に及ぼす影響について調べることにした。

4.1 立ち下がり角度、立ち上がり角度の変形方法

VCV パワーパタンの立ち下がり角度、立ち上がり角度を変化させる方法として、以下のような方法を用いた。

まず、目視にて、自然発声の VCV パワーパタンを、立ち下がり (RL) 区間、定常 (ST) 区間、立ち上がり (AT)

区間の3つに分離した。RL区間は、先行母音のCEGVパワーが維持されている点から母音の終了点までの区間である。ST区間は、母音においてCEGVパワー以上のパワーが持続する区間である。AT区間は、後続母音の開始点からCEGVパワーに到達するまでの区間とした。RL区間のパワーパターンに対し、 w 倍の時間長伸縮を加えることで、RL区間の立ち下がり角度(δ_r [dB/ms])を $1/w$ 倍に変形することができる。同様に、AT区間のパワーパターンに対し、 w 倍の時間長伸縮を加えることで、AT区間の立ち上がり角度(δ_a [dB/ms])を $1/w$ 倍に変形することができる。伸縮に伴う時間長の補償は、それぞれの母音のST区間で行なう。RL区間、AT区間のパターン変形の例を図-6に示す。

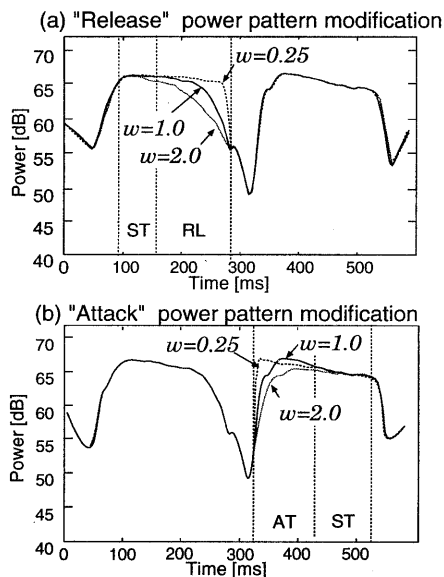


図 6: 母音のパワーパターン立ち下がり角度、立ち上がり角度の変形操作

4.2 実験 3

無意味単語 (/araCara/) の発声を、発話速度 7.0 mora/s に固定して収録した。子音 (C) は /b/, /g/, /p/, /s/, /m/, /r/ の 6 種類に限定した。これらの組み合わせによる合計 6 種類の単語について、パワーパターンを分析した。さらに分析されたパワーパターンの VCV の立ち下がり角度、立ち上がり角度を先に示した方法によって変形し、各変形パターン毎に分析合成音を作成した。ここで、伸縮率 w は、2.0, 1.5, 1.0, 0.5, 0.25 の 5 種類とした。同一単語中の立ち下がり角度に変形を加えたもの同士 (グループ A) と、同一単語中の立ち上がり角度に変形を加えたもの同士 (グ

ループ B) のすべての組み合わせについて、「どちらがより自然か」という尺度において対比較試験を行なった。被験者は 3 名であり、1 つの組み合わせに対し、1 人あたり 3 回の判定を行なった。

4.3 結果 3

図-7 に各サンプル同士の対比較試験の結果をもとに、自然性における Thurston Case IV の心理的距離を求めた結果を示す。ここで、パワーパターンの変形を行っていないオリジナルの音声 ($w = 1.0$) を 0 としている。

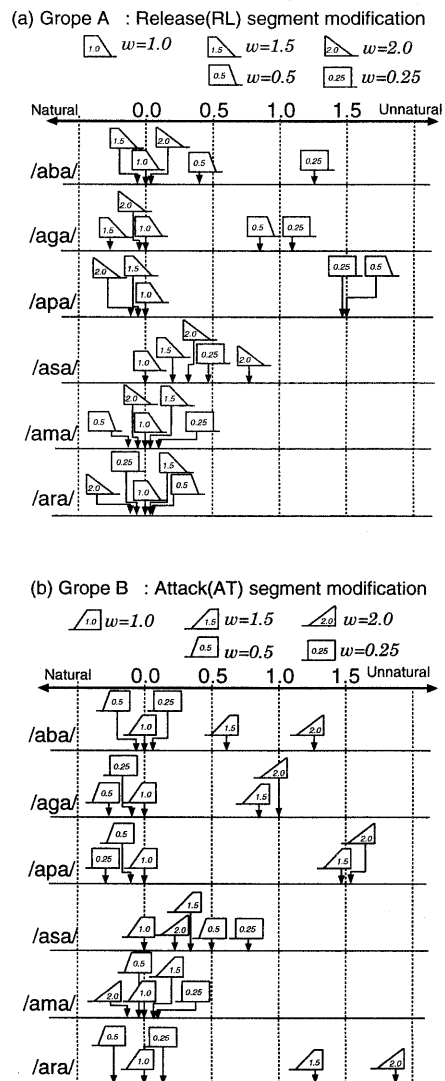


図 7: VCV パワーパターンの立ち下がり角度、立ち上がり角度の変化と弁別傾向

4.4 考察 3

結果 3 より以下のことが明らかとなった。

1. どの VCV についても、パワーパターンを変形させないオリジナルのパワーパターン ($w = 1.0$) を用いた合成音は、安定した高い自然性が実現されている。
2. グループ A (RL 区間の変形) においては、パワーの角度が急になった場合 ($w < 1.0$) には、自然性が劣化する傾向がみられるが、角度が緩やかになる場合 ($w > 1.0$) には自然性は殆んど変化しない。
3. グループ B (AT 区間の変形) においては、パワーの角度が急になった場合 ($w < 1.0$) には、自然性の変化はみられないが、角度が緩やかになった場合 ($w > 1.0$) には自然性が劣化する。これはグループ A の結果と対比的である。
4. これらの傾向は、破裂音を挟む VCV では顕著であるが、ある程度子音のパワーが大きな鼻音では目立った傾向はみられなかった。これより、VCV パワーパタンのダイナミックレンジが大きな音節ほど、立ち下がり角度、立ち上がり角度が自然性に与える影響が大きいと考えられた。

以上の結果より、VCV パワーパターンにおける立ち下がり角度 (δ_r)、立ち上がり角度 (δ_a) の変化は、子音による影響度の差はあるものの、合成音声の自然性に無視できない影響を及ぼすことが明らかとなった。また、実験 2 の結果と合わせることで、発話速度を変化させた場合でも、通常の発話速度程度 (7.0mora/s) の VCV パワーパタンの立ち下がり角度、立ち上がり角度を、変化後のパターンに用いることが可能であると結論づけられた。

5 パワーパタンのマクロ／マイクロモデル

3, 4 章の実験結果をふまえ、マクロパターンとマイクロパターンによるパワーパターン制御モデルを提案する。まず、マクロパターンを形作るパラメータの分析方法について述べ、これを使ったマクロパターン合成方法を示す。次にマイクロパタンの分析方法と、データベース化の手順を示す。さらに、分析合成に適用した実験と、品質評価を行なった結果について述べる。

5.1 マクロパタンの分析

マクロパターンを形作るパラメータ (マクロパラメータ) として、3.4 節で述べた VCV 単位の立ち下がり角度 (δ_r) と、立ち上がり角度 (δ_a) に加え、発話速度に応じた TC 位置を決定するための、以下の 2 つのパラメータを導入する。

1. 分割係数: α

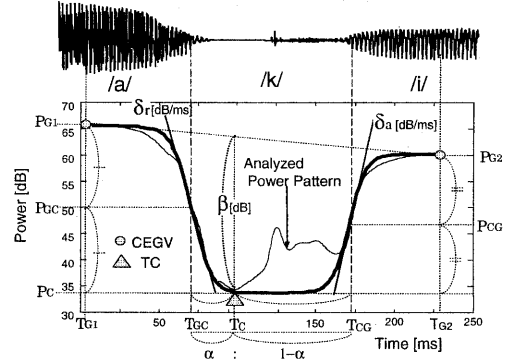


図 8: マクロパラメータの分析

表 1: マクロパラメータテーブルの要素

VCV	δ_r δ_a	D_s	α_s	β_s
		D_m	α_m	β_m
		D_f	α_f	β_f

子音パワー区間における TC の位置を、区間を 1 とした場合の内分比で表したものの。ここで、子音パワー区間とは、先行母音の CEGV パワー ($P_{G1}[\text{dB}]$) と TC パワー ($P_C[\text{dB}]$) との平均パワー ($P_{GC} = (P_{G1} + P_C)/2[\text{dB}]$) を通過する時刻 ($T_{GC}[\text{ms}]$) から、TC パワー ($P_C[\text{dB}]$) と後続母音の CEGV パワー ($P_{G2}[\text{dB}]$) との平均パワー ($P_{CG} = (P_C + P_{G2})/2[\text{dB}]$) を通過する時刻 ($T_{CG}[\text{ms}]$) までとした。

2. ディップ係数: $\beta[\text{dB}]$

先行母音の CEGV パワー ($P_{G1}[\text{dB}]$) に到達した点と、後続母音の CEGV パワー ($P_{G2}[\text{dB}]$) に到達した点を結んだ線分からみた、TC パワーの落ち込み具合を dB 単位で表したものの。

マクロパタンの分析の際は、ピッチの影響を受けないような比較的広い幅 (30ms 以上) の窓を用いてパワーパターンを分析する。このパワーパターンに先行母音の CEGV、子音の TC、後続母音の CEGV をそれぞれマークし、 α 、 β を求める。3 種類の異なる速度 (5.0mora/s, 7.5mora/s, 10.0mora/s) で発声された自然音声の VCV それぞれに対しパラメータの分析を行ない、中速度の時の δ_r 、 δ_a と共に、表-1 に示すパラメータのテーブルを作成した。ここで作成されるパラメータを、マクロパラメータと呼ぶことにする。

5.2 マクロパタンの合成

ある発話速度から、リズム規則によって導出された当該 VCV 時間長 (D_g) をもとに、以下の手順に従ってマクロパタンを合成する。

- スペクトルパタンデータベース中のラベル情報をもとに、CEGV 位置 (T_{G1}, T_{G2})、および、子音パワー区間の境界位置 (T_{GC}, T_{CG}) が定まる。当該母音の種類とピッチ周波数、フレーズ内位置によって先行母音、後続母音の CEGV パワー (P_{G1}, P_{G2}) を決定する。
- 当該 VCV のマクロパラメータテーブルより、マクロパラメータを取得し、目的継続時間長 D_g と、 D_s, D_m, D_f との比率をもとに、発話速度に応じた α, β を、それぞれ $\alpha_s, \alpha_m, \alpha_f$ 、と $\beta_s, \beta_m, \beta_f$ からの内挿によって求める。CEGV 位置 (T_{G1}, T_{G2}) と CEGV パワー (P_{G1}, P_{G2})、および、 α, β より、発話速度に対応した TC 位置 (T_C) と TC パワー (P_C) が決定される。
- 当該 VCV 固有の立ち下がり角度 (δ_r)、立ち上がり角度 (δ_a) をテーブルから取得し、パタンの枠組をつくる。
- 最後に、以下の式に基づいたマクロパタンを作成する。

$$M_c(t) = \begin{cases} \frac{P_{G1} + P_C}{2} + \frac{P_C - P_{G1}}{2} \frac{\exp(\delta_r(t - T_{GC})) - 1}{\exp(\delta_r(t - T_{GC})) + 1} & \text{for } t \leq T_C \text{ (立ち下がり)} \\ \frac{P_C + P_{G2}}{2} + \frac{P_{G2} - P_C}{2} \frac{\exp(\delta_a(t - T_{CG})) - 1}{\exp(\delta_a(t - T_{CG})) + 1} & \text{for } t > T_C \text{ (立ち上がり)} \end{cases} \quad (1)$$

5.3 ミクロパタンの分析

ミクロパタンの分析においては、子音部における比較的パワー変化の早い特徴を得るために、無声子音では 10ms 程度の短い窓を掛けたパワー分析を行ない、有声子音は、マクロパタンと同様の 30ms の窓を用いる。このようにして得られた細密なパワーパタンから、前節で述べたマクロパタン $M_c(t)$ との差分をとり、ミクロパタン $M_i(t)$ を作成する。

図-10 に、例として /aki/ について分析されたミクロパタンを示す。このようにして得られたミクロパタンのうち、子音区間に相当する箇所のみをデータベースに登録する。これらのミクロパタンの各点は、時間的に対応するスペクトルデータと時間同期がとられており、スペクトルパタンとの同期関係を保つようになっている。

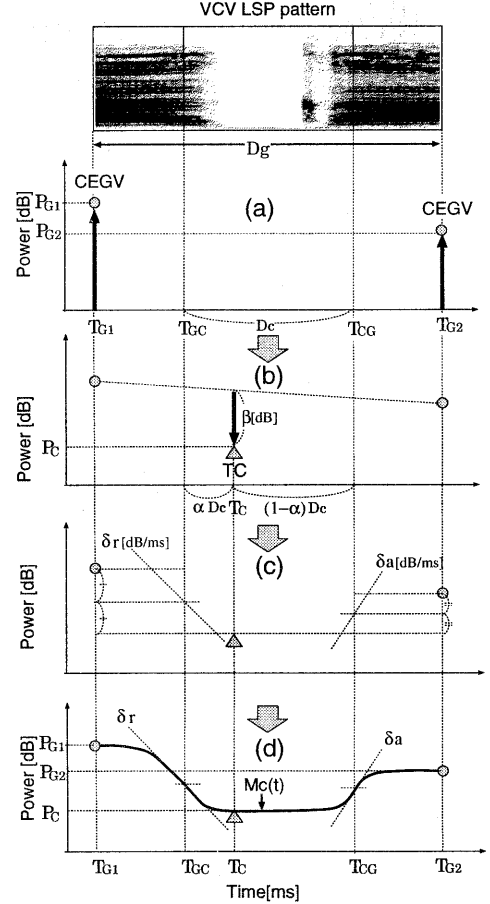


図 9: マクロパラメータによるマクロパタンの合成方法

6 規則合成への適用

6.1 実験 4

本パワーパタン制御方法の規則合成システムへの応用をはかった。本規則合成システムは 920 単位の VCV データベースを用いている。このうち、文章中の出現確率が上位 300 単位までの VCV についてマクロパラメータテーブルとミクロパタン DB を作成した。ここで選択された 300 単位は、新聞記事などから抽出した 812 文章を分析した結果であり、全文章の 90% までの VCV 単位を含んでいる。これらの VCV パタンと、マクロパラメータテーブル、ミクロパタン DB を持つ規則合成システムを構築した。

10 個の単語について、3 種類の発話速度 (5.0mora/s, 7.5mora/s, 10.0mora/s) による、規則合成音を作成した。それぞれのサンプルについて、(1) 元の VCV パワーパタンを VCV 継続時間長 (D_g) に合わせて線形にパワーパタン

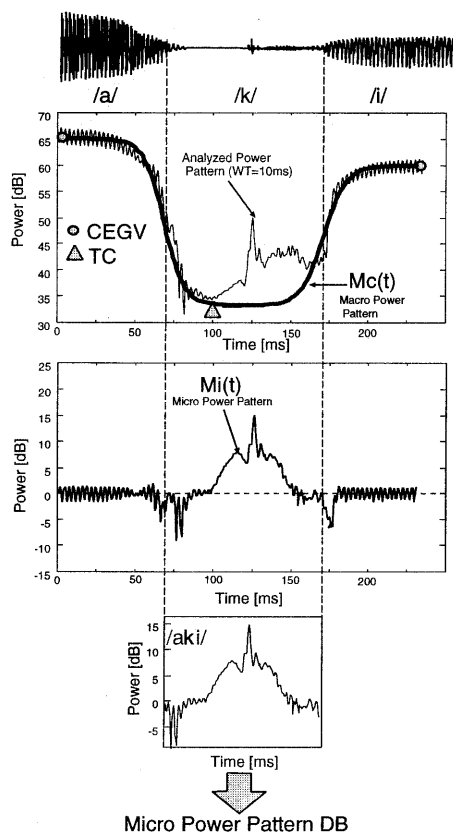


図 10: ミクロパワーパターン

ンを伸縮させた合成音, (2) 本手法を用いてパワーパターンを作成した合成音, の2種類の規則合成音を作成した。3人の被験者に対し, 同じ単語, 同じ発話速度の合成音(1),(2)を対比較し, どちらが自然であるかを強制選択させた。1人あたりの提示数は, 1つの組み合わせにつき, 4回とした。

6.2 結果 4

2種類の合成音を提示した場合に, より自然であると判定された合成音の選択率を, 図-11に示す。

以上の結果より, 本手法によって作られる規則合成音声は, 発話速度が変化した場合においても, 自然性を劣化させないことが明らかとなった。特に, 発話速度が速くなった場合でも, 音韻性を落さずに自然なパワーパターンを合成できることがわかった。

7 まとめ

VCVを基本単位とした規則音声合成方式において, 発話速度の変化に対応することのできる, パワーパターン制

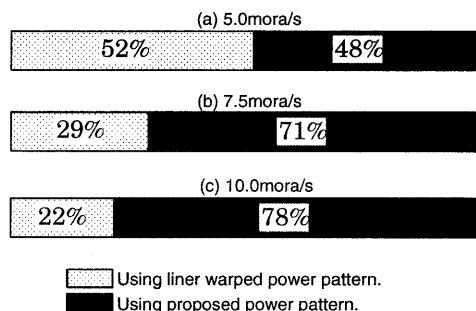


図 11: 線形伸縮によるパワーパターンと, 本手法によるパワーパターの規則合成音における評価結果

御モデルを提案した。本手法は, 自然音声の発話速度変化に対するパワーパターン変動をモデル化し, マクロなパターンを表すパラメータテーブルと, ミクロなパターンを保存するデータベースを用いている。規則合成を用いた評価実験より, 本パワー制御を用いた合成音は, 発話速度の変化に対しても, 音韻性, 自然性を劣化させることなく, 自然なパワーパターンを合成するのに有効であることがわかった。本手法は, 規則合成音の自然性向上に役立つだけでなく, 話速変換や音声情報の圧縮にも応用可能である。

謝辞

研究の機会を与えていただいた, セコム(株)常務取締役 橋本新一郎 IS 研究所長, 並びに 日頃御指導頂く 音声情報処理研究室, 古村光夫室長に感謝いたします。また, 日頃御教示頂く, 千葉大学大学院自然科学研究課 平田廣則教授に感謝申し上げます。

参考文献

- (1) 中島信弥, 浜田洋, “音韻環境に基づくクラスタリングによる規則合成法,” 信学論 **J72-D-II**, 1174-1179 (1989).
- (2) 武田一哉, 安部勝雄, 匂坂芳典, “選択的に合成単位を用いる規則合成,” 信学論 **J73-D-II**, 1945-1951 (1990).
- (3) 広瀬啓吉, “音声の出力に関する研究の現状と将来,” 音響学会誌 **52**, 857-861 (1996).
- (4) 三村克彦, 海木延佳, 匂坂芳典, “統計的手法を用いた音声パワーの分析と動特性の制御,” 音声研究会資料 SP91-4, (1991).
- (5) 伊藤憲三, 広川智久, 佐藤大和, “音声合成のための音韻セグメントパワー制御の検討,” 音声研究会資料 SP92-12, (1992).
- (6) 加藤雅代, 古村光夫, 橋本新一郎 “母音部エネルギー重心点に着目した日本語リズム規則,” 音響学会誌 **50**, 888-896 (1994).
- (7) 濱上知樹, 古村光夫, “規則合成のためのパワー制御モデル,” 音講論集 **2-8-11**, 241-242 (1993.10).
- (8) 濱上知樹, 橋本新一郎, “パワー包絡が聴覚に与える影響について,” 音講論集 **1-5-21**, 257-258 (1992.10).