

TR - H - 166

統計的時系列モデルを用いた変換聴覚  
フィードバック法の検討

An investigation of Transformed Auditory Feedback  
method by a statistical time series model

加藤 比呂子

河原 英紀

1995. 9. 22

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2 ☎ 0774-95-1011

ATR Human Information Processing Research Laboratories

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-774-95-1011

Facsimile: +81-774-95-1008

©(株)ATR人間情報通信研究所

統計的時系列モデルを用いた変換聴覚フィードバック法の検討  
An investigation of Transformed Auditory Feedback method by a  
statistical time series model

加藤比呂子 河原英紀

Hiroko Kato Hideki Kawahara

ATR 人間情報通信研究所

ATR Human Information Processing Research Laboratories

概要

統計的時系列モデルを音声時系列データから抽出された基本周波数と振幅二乗平均値にあてはめることにより、音声発声生成機構に対する聴覚からの影響について解析した。その結果、1. それぞれのゆらぎの原因となるノイズは互いに影響を及ぼしあっている 2. 変換聴覚フィードバック法は自然な発声状態を若干変化させているかもしれないということがわかった。

この解析により、変換聴覚フィードバック法から得られた知見を裏付けることができ、また、その手法では出来なかった音声パラメータの相互の関係を探ることができた。

We analyzed the effects to the mechanism of speech utterance and production from auditory system by fitting a stochastic time series model to pitch and power data extracting from recorded speech time series. It includes that 1. The noise contributing to each fluctuation of pitch and power influences each other, 2. Transformed Auditory Feedback (TAF) method might make a little bit changing the normal speech condition.

In this analysis, we can support the results obtained from TAF method. Furthermore, we can investigate mutual relationship between speech parameters which the TAF method could not find.

## 1 はじめに

音声生成機構に聴覚がどのように関わるかという研究は、1950年代から遅延聴覚フィードバックに伴う研究 (Fairbanks (1995), Lee (1950)) 等多数報告されていた。しかし、これらの結果は正常な聴覚のフィードバックの役割の上に成り立っている検討ではなかった。

通常の発声時において、聴覚情報がどのように音声の生成を制御しているかについて定量的評価をする場合、問題は、いかにして正常な発声過程を妨害せずに聴覚フィードバックの役割を見い出さかであった。

その要求に答えるために、河原 (1993) により開発されたのが変換聴覚フィードバック (Transformed Auditory Feedback 以下 TAF) 手法であった。これは音声生成機構と聴覚のシステムを介したフィードバック経路に微小な既知の摂動信号を入力し、それに対する出力の応答をみることによって信号伝送路の動特性をみるという手法である。ここでいう動特性とは、音声発生に際しどのような聴覚系の影響があるかという特性を指す。この実験法から得られた知見は、

1. 入力した摂動信号として、Maximum period sequence を用いていて、その摂動信号と耳に入る基本周波数との相関のピークに対し、摂動信号と発声された基本周波数の相関のピークが 100 ~ 200ms 遅れと負の相関を呈している。これらのピークの位置の間隔を聴覚フィードバック経路からの応答時間とすると、聴覚系から及んでいる摂動を補正する応答があるということを定量的に示すことができた (Kawahara (1993))。またこの結果を全極モデルに基づく方法 (Hirayama and Kawahara (1994)) により検証をおこなった
2. 摂動を加えない自然発声の状態で作成した音声データから抽出した基本周波数の系列に対し、単変量 AR モデルをあてはめ、特性根の配置と、そのインパルス応答を調べ、TAF より得られた応答遅延時間と対応がつくことを確認した (Hirayama and Kawahara (1994)、Kawahara (1994))
3. 伝達特性推定法で、閉ループと開ループ特性を調べることにより、音声生成相互作用にはそれまで得られていた 100 ~ 200ms の応答時間よりさらに遅い時定数を含む系が存在することが確認された (Kawahara 1995)

であった。

TAF でおこなわれた手法では正常の発声を妨げないとはいえ、人工的にデータに信号を加えて変調したデータに対し解析を施している。しかし伝達関数により聴覚系の伝達特性を調べるためには、システムに何らかの入力信号をいれないと解析は難しい。また、TAF の場合、単変量の信号解析をおこなっているため、音声を構成する音声パラメータ間の相互の関係を調べるのは難しい。

そこで本稿では、自然発声、マスキングそして TAF の条件下のもとで得られた生の音声実験データから音声パラメータを抽出し、統計的時系列モデルをあてはめることによるシステム解析を紹介する。具体的には、

1. 単変量 AR モデルによる TAF の解析結果の検証、
2. 音声信号から抽出した基本周波数 (以下  $F_0$ ) と音声の振幅の自乗平均値 (以下 rms) との関係についての多変量時系列モデルによる解析

をおこなう。

特に項目 2 については、データの中には、非定常性を含む信号もあり得ることから、Kato, Wada and Ishiguro (1994) と Kato (1995) により開発された、平均非定常多変量フィードバックモデルの procedure を適用した。

本稿の構成は、2章に本稿で取り扱うデータの説明を、3章では TAF の検証を、4章では平均非定常多変量時系列解析による結果を、そして最後に5章にはまとめ、と成っている。また解析に必要な計算方法についての簡単な説明は付録にまとめた。

## 2 データ

被験者は二人である (以後 subject1、subject2 とする)。subject1 が G $\sharp$ (207.58Hz) 音、subject2 が C(130.75Hz) 音を聞いた後で、その tone を保つように母音「ア」を持続発声した。発声をするときの条件として、自然な発声状態 (以下 状態 WO)、ヘッドホンによりピンク雑音と発声する前に聞いた pure tone でマスクした状態 (以下 状態 WM)、また TAF 条件下の状態 (以下 状態 T) とした。図 1 に、これらの状態をブロック線図に簡単に示す。

ここで、 $P$  は音声生成機構のシステムの特性を、 $A$  は聴覚システムの特性を示す伝達関数と定義する。また、 $n$  は神経パルスの束から発せられる信号、 $v$  は声帯の振動から得られる揺らぎ等を示すものとする。TAF 実験法で考えている伝達特性は、 $P$  と  $A$  を別々に測定することはできないので、

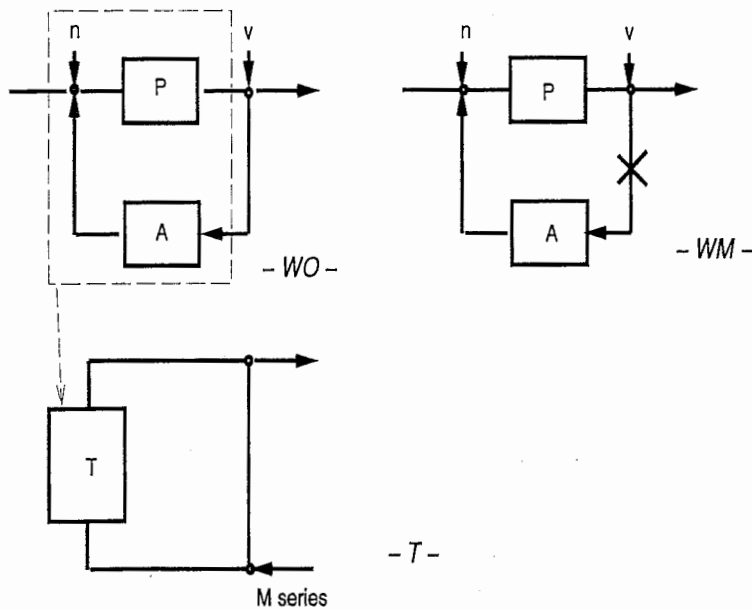


図 1: 状態 WO、WM、T を表すブロック線図

状態 WO の図の点線で囲った部分をひとまとめとした、聴覚系 A から音声生成機構 P への開ループの特性

$$nP + PA \quad (1)$$

を求めること相当する。TAF では、得られた音声データを、入力信号として使う Maximum period sequence と同じ周期分で平均化しているため、神経系から発生すると予想されている信号を平均値 0 のノイズと仮定して、その伝達特性は PA 自身とおくことができる (-T- 図)。また、状態 WM は、自分の声がマスキングされるため、音声生成機構から聴覚系の経路はカットされることになる (-WM- 図 × 印)。

データの収集については、Kawahara(1993、1994)と同様の実験系を用いておこなった。発声の際マイクから唇までの距離を約 20cm の状態でおこない、発声された音声信号は、8kHz 16bit で量子化され、計算機に取り込まれた。取り込まれた音声データは、音声信号処理ソフト EPSF の formant プログラムを用いて、50ms の長さを  $\cos^4$  の窓を利用し分析開始位置を 10ms ずらせながら切り出して F0 を求め、またその長さ分の rms も抽出した。図 2 に抽出されたデータ例を示す。横軸は時間 (× 50 msec) を示す。2 分間の持続母音発声のため、途中息継ぎの箇所がみられる。また、縦軸

は生データを常用対数変換した値で、近似的に線形に取り扱うことにする。上から、状態 WO、WM  
そして TAF 実験法の条件下で収集された F0 と rms である。縦軸の値 2.6 ~ 3.2 に観測される変動幅  
の大きい方が rms 系列であり、2.2 ~ 2.4 の幅を rms よりも小さく変動しているのが F0 系列である。

### 3 単変量 AR モデルによる TAF 実験結果の検証

TAF では聴覚を介した経路の特性を解析することを狙ったが、聴覚が実際の音声の生成において、  
どの程度の役割を果たしているかを明らかにする手段として、状態 WO と WM の揺らぎの性質を調  
べることも必要と考えられる。図 1 状態 WO における出力の伝達特性を示す関数は、

$$v + P \frac{n + vA}{1 - PA} \quad (2)$$

となる。これに対して図 1 状態 WM における出力の伝達特性は、フィードバック経路が切断される  
ので、仮に聴覚系に何らかの入力信号  $\alpha$  が入ったとすると、

$$nP + \alpha AP + v \quad (3)$$

となる。しかし、状態 WM の実験では、 $\alpha$  に相当する気導音や骨導音の雑音がフィードバックへ影  
響を及ぼさないようにするために約 80dB のピンク雑音を用いたので、その部分が消去されて (3) 式  
は

$$nP + v \quad (4)$$

と示すことにする。TAF 実験では主に F0 について解析していたので、その比較をするために状態  
WO と WM の F0 の時系列データに対し、単変量 AR モデルをあてはめ、その係数と分散からパワー  
スペクトルを求めた。その計算方法については付録 A 参照。

TAF 実験法で得られた解析結果を検証するために、聴覚系のみからの影響の特性を調べる。その  
ために、単純に状態 WO の特性を WM の特性で割ることによって得られる

$$\frac{1}{1 - PA} \quad (5)$$

の  $PA$  が、TAF で得られる特性に相当していることに着目する。WO のパワースペクトルを WM  
のパワースペクトルで割算した値の平均値を二人の被験者について計算した結果を図 3 に示す。上段  
が subject1、下段が subject2 の推定値である。また、TAF で得られた  $PA$  に相当する伝達特性から

得られたパワースペクトルを図4と5に示す。これらの結果と比較すると、WOをWMで割った特性よりもTAFによる特性におけるパワースペクトラムのピークが周波数の低いほうに約1Hzずれている。これは波長にすると50msecである。このずれの原因として考えられるのは、

1. 状態WMではピンク雑音挿入により気導音等の入力を妨害していることを仮定しているが、実はその雑音に対する応答の機構は複雑であると想定されるため、聴覚系からの入力をマスクするのは pure tone を用いたものだけを扱った方がいいのかもしれない
2. TAF 実験法では自然発声を妨害しないことを前提にしているが、振動信号挿入の影響や、フィードバックで返している音量が自然発声のときよりも非常に大きいことによる影響が、発声に及んでいるかもしれない

ということが考えられる。

## 4 F0 と rms の相互の関係について

### 4.1 平均非定常多変量フィードバックモデル

F0 と rms の相互の関係を調べるために、平均非定常多変量フィードバックモデルを用いたシステム解析をおこなう。

観測されるデータはF0とrmsなので、2変量時系列データ  $\mathbf{y}(n) = (y_1(n), y_2(n))^t$  ( $n = 1, \dots, N$ ) が、以下のように各成分に分解されると仮定する。

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{t}(n) + \mathbf{i}(n), \quad (6)$$

ここで、 $\mathbf{x}(n)$ 、 $\mathbf{t}(n)$ 、 $\mathbf{i}(n)$  はそれぞれ、循環変動成分、トレンド(平均非定常)成分、観測ノイズ系列とする。

トレンドまわりの循環変動成分には、2変量ARモデル

$$\mathbf{x}(n) = \sum_{m=1}^M \mathbf{A}(m)\mathbf{x}(n-m) + \mathbf{v}_x(n), \quad \mathbf{v}_x(n) \sim N(\mathbf{0}, \Sigma_x) \quad (7)$$

を仮定する。ここで、 $\mathbf{v}_x(n) \sim N(\mathbf{0}, \Sigma_x)$  とは、ノイズ系列が平均ベクトル  $\mathbf{0}$ 、分散共分散行列  $\Sigma_x$  の2次元正規分布に従うという意味である。(7)式で示した2変量ARモデルの係数行列と、分散共分散行列がわかれば、F0とrmsの相互の関係を探ることができる(付録C参照)。

トレンド成分については、図 2 により F0 はある一定の平均値のまわりに変動する定常過程、また、rms は非定常過程と扱うことにし、F0 については平均値を下げ、rms には 2 階差のトレンド、

$$\nabla^2 t_2(n) = v_{t_2}(n), \quad v_{t_2}(n) \sim N(0, \sigma_t) \quad (8)$$

を仮定した。ここで、 $\nabla$  は階差オペレータを示す。F0 の平均値はそれぞれの被験者が発声するとき聞いた tone の周波数に近いものと推察される。

また、観測ノイズ系列には

$$i(n) \sim N(0, \Sigma_i) \quad (9)$$

を仮定する。

この 2 変量ベイズ型時系列モデルを状態空間表現で表すことで、カルマンフィルタによりモデルの尤度の計算が可能になる ( 付録 B 参照 )。最適なモデル選択には情報量基準 AIC を用いる。そして数値的最適化により未知ハイパーパラメータに相当する、式 (7) の AR の係数行列、それぞれの系列に仮定されている分散の値を求めることが出来る。

#### 4.2 状態 WO と TAF 条件下における F0 と rms の関係について

図 2 には、息継ぎの箇所が数ヶ所観測されていて、息継ぎの度に発声機構の状態が変化していくと考えられたので、本稿では最初のサンプルを用いて解析した。

モデル (6) 式により分解された成分の一結果を図 6 に示す。図上段は原系列を点線で、推定された平均値 (約 210Hz) と滑らかなトレンド成分は実線で示されている。また下段は実線が F0、点線が rms の循環変動成分 (左図) と残差の系列 (右図) を示している。

3章の解析で、状態 WM のデータ収集条件について今後検討する必要があるために、ここでは状態 WO と T 条件下についての解析結果についてのみ議論することにする。

まず WO の状態の場合における F0 と rms の循環変動成分についての対数変換したパワースペクトルを図 7 に示す。この結果は F0 と rms が相互に影響を及ぼし合っている上での、それぞれ自身の特性を示している。パワーのピークに位置については、subject1 は F0 が 3.25Hz、rms は特にピークはない。両系列ともに長い周期成分にもパワーが存在している。また、subject2 については、F0 は 3.75Hz、rms は 3.5Hz 付近と長い周期成分にパワーを強くもつ。両被験者の F0 のピークは、3章で示した TAF の結果に近い値を示しているようだ。



次に、F0 と rms のノイズがシステムにどのような影響を示すかというのを図 8 に示す。ここで、上段が subject1、下段が subject2 についての結果である。左側の 2 つの図では、推定値の曲線を境に下方が F0 のノイズから rms への影響、上方が rms のノイズから rms 自身への影響を、また、右側の 2 つの図では推定値の上方が rms のノイズから F0 への影響、下方が F0 のノイズから F0 自身への影響を表す。両被験者に共通していえることは、rms から F0 への寄与が大きいことである。これは、音量が声の高さに寄与しているということを示している。また、subject2 は subject1 に比べて声の高さに対して音量のコントロールが旨く出来ているであろうことも表している。これらの相互の関係は、聴覚を介したものだけでなく、さらに声帯の振動機構によるものでもあることが示唆される。それを直接的に検証するためには、例えば声門下圧を変化させたり声帯に外部から機械的操作を加える等の方法によることが必要と考えられる。TAF 実験では聴覚を介した情報の影響を検出することのみに限られるので、この検証はそれらの他に新たな影響を補足できる可能性を生むことになるかもしれない。

次に F0 と rms をサブシステムとしたフィードバックループに対して、それぞれのチャンネルにインパルスを入れたときの、もう一方のチャンネルの応答を推定した結果を図 9、10 に示す。上段の推定値は WO の状態で得られたものである。この場合、被験者に共通して言えることは、F0 にインパルスを入れたときの rms の応答より、rms にインパルスを入れたときの F0 の応答がより長く応答している。これは、ノイズ寄与率で rms のノイズから F0 への寄与の方が高いことと一致している。

下段は TAF 条件下で得られた結果だが、状態 WO よりも長く応答している。これでは TAF と WO の伝達特性が異なって見えるが、3 章での解析で議論した摂動信号に応答していることによるものかと考察されるが、これだけの結果からは未だ断定できない。

## 5 まとめ

以上の解析結果をまとめると、

1. 人工的に変調しない音声データから得られた基本周波数に対し、単変量 AR モデルから推定された聴覚系による結果と TAF 実験法による推定結果には約 1Hz のずれがある。
2. 平均非定常多変量フィードバックモデルにより、基本周波数と音声の振幅自乗平均値との間にフィードバックの関係が認められた。基本周波数のゆらぎには、声の大きさのゆらぎによる影響が強いことが明らかになった。

3. 平均非定常多変量フィードバックモデルにより分解された循環変動成分から求めた個々の系列のパワースペクトルのピークは TAF から推定されたパワーのピークとほぼ同じ位置にある。TAF では聴覚からの影響を示す位置と推察され、分解された系列では、相互の関係を含めたモデルから得られた推定値である。両者の結果を照らし合わせて考えると、音声生成に聴覚からの影響を含んでいることが推察される。

今後の課題としては、WM の状態でのマスキングの条件を検討して、その時の基本周波数と振幅自乗平均値との関係を、平均非定常多変量フィードバックモデルで明らかにすることにより、音声生成機構に対する聴覚系からの役割を明らかにすることである。

#### 参考文献

- [1] G. Fairbanks, 1955, Selective vocal effects of delayed auditory feedback, *Journal of Speech and Hearing Disorders*, 20(4), pp.333-346.
- [2] B. S. Lee, 1950, Effects of delayed speech feedback, *The Journal of The Acoustical Society of America*, 22(6), pp.824-826.
- [3] Kato, H., 1995, A study of multivariate mean-nonstationary time series model for estimating mutual relationships, unpublished Ph.D. dissertation, The Graduate University for Advanced Studies, Department of Statistical Science, Tokyo.
- [4] Kato, H., Wada, T., Ishiguro, M., 1994, A study of human body balance by new multivariate feedback models with common low frequency components, *Japanese Journal of Biometrics* 2.
- [5] 河原英紀, 1993, 変換聴覚フィードバックによる音声生成・知覚相互作用の検討, 音響学会聴覚研資, H-93-24.

- [6] 平山和彦, 河原英紀, 1994, 音声基本周波数の揺らぎに対する聴覚フィードバック条件の影響について, 信学技報, SP94-48.
- [7] Kawahara, H., 1994, Effects of Natural Auditory Feedback on Fundamental Frequency Control, IC-SLP'94, 24.2.
- [8] 河原英紀, 1995, フィードバック音声の基本周波数への摂動に対する応答の成分への分解について, 信学技報, SP94-105, pp.53-60.

## 付録

### A 単変量 AR モデルからパワースペクトルの計算

単変量 AR(M) モデル

$$x(t) = \sum_{m=1}^M a(m)x(t-m) + \epsilon(t) \quad \epsilon(t) \sim N(0, \sigma^2) \quad (10)$$

の係数と分散を、最小 2 乗法等から求めると、周波数応答関数

$$A(f) = \frac{1}{1 - \sum_{m=1}^M a(m)e^{-i2\pi fm}} \quad (11)$$

が得られるのでそのパワースペクトルは、

$$p(f) = \frac{\sigma^2}{|1 - \sum_{m=1}^M a(m)e^{-i2\pi fm}|^2} \quad (12)$$

で求められる。

### B 平均非定常多変量時系列モデルの同定

4.1 で示されたモデル (6) 式は状態空間表現

$$\mathbf{z}(n) = \mathbf{F}\mathbf{z}(n-1) + \mathbf{G}\mathbf{v}(n) \quad (13)$$

$$\mathbf{y}(n) = \mathbf{H}\mathbf{z}(n) + \mathbf{w}(n), \quad (14)$$

で表すことができる。ここで  $\mathbf{z}(n)$  は状態ベクトル、 $\mathbf{v}(n)$  はシステムノイズを表す。

基本的に、このモデルはベイズモデルなので、そのハイパーパラメータベクトルを示すと

$$\Theta = (a_{11}(1), \dots, a_{11}(M), a_{12}(1), \dots, a_{12}(M), a_{21}(1), \dots, a_{21}(M), a_{22}(1), \dots, a_{22}(M), \quad (15) \\ \sigma_{x_1}^2, \sigma_{x_2}^2, \sigma_{t_2}^2, \sigma_{w_1}^2, \sigma_{w_2}^2).$$

というようになる。式 (13) は、このハイパーパラメータベクトルと、状態の平均ベクトルの初期値より与えられた事前分布を定義するものである。また、式 (14) は、ベイズモデルのパラメータに相当する  $\mathbf{z}(n)$  を与えられたときの観測値  $\mathbf{y}(n)$  の条件付き分布を定義付ける。

式 (13) と (14) から、 $\pi(\mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(n) | \Theta, \bar{\mathbf{z}}(0))$  と  $f(\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(n) | \mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(n), \Theta)$  が以下によって定義される:

$$\pi(\mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(n) | \Theta, \bar{\mathbf{z}}(0)) = \prod_{i=1}^N \phi_p(\mathbf{z}(i) | \mathbf{F}\mathbf{z}(i-1), \Sigma_{\mathbf{z}}) \quad (16)$$

$$f(\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(n) | \mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(n), \Theta) = \prod_{i=1}^N \phi_2(\mathbf{y}(i) | \mathbf{H}\mathbf{z}(i), \Sigma_{\mathbf{y}}). \quad (17)$$

ここで、 $\phi_2(x | \mu, \Sigma)$  は平均  $\mu$ 、分散  $\Sigma$  の 2 次元正規分布を示すとすると、

$$\Sigma_{\mathbf{z}} = E\{\mathbf{G}\mathbf{v}(n)\mathbf{v}(n)^t \mathbf{G}^t\},$$

$$\Sigma_{\mathbf{y}} = E\{\mathbf{w}(n)\mathbf{w}(n)^t\}.$$

$p$  は状態ベクトルの次元である.

$(\mathbf{y}(1), \dots, \mathbf{y}(n))$  の分布は

$$f(\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(n) | \Theta, \bar{\mathbf{z}}(0)) = \prod_{i=1}^N f(\mathbf{y}(i) | \mathbf{Y}(i-1), \Theta, \bar{\mathbf{z}}(0)), \quad (18)$$

で表される

$$f(\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(n) | \Theta, \bar{\mathbf{z}}(0)) = \int \int \dots \int f(\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(n) | \mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(n), \Theta) \\ \times \pi(\mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(n) | \Theta, \bar{\mathbf{z}}(0)) dz(1) dz(2) \dots dz(n),$$

によって定義される. ここで、 $\mathbf{Y}(i-1) \equiv (y(1), y(2), \dots, y(i-1))$  と

$$f(\mathbf{y}(i) | \mathbf{Y}(i-1), \Theta, \bar{\mathbf{z}}(0)) = \frac{f(\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(i) | \Theta, \bar{\mathbf{z}}(0))}{f(\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(i-1) | \Theta, \bar{\mathbf{z}}(0))}, \quad (19)$$

という関係を示す.

データ  $(\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(n))$  が与えられたとき、モデル (6) の尤度

$$L(\Theta, \bar{\mathbf{z}}(0)) = \prod_{i=1}^N f(\mathbf{y}(i) | \mathbf{Y}(i-1), \Theta, \bar{\mathbf{z}}(0)), \quad (20)$$

は、以下に示すカルマンフィルタのアルゴリズムで計算することができる. 初期値  $\mathbf{z}(0|0)$  と  $\mathbf{V}(0|0)$  が与えられると、

予測

$$\mathbf{z}(n|n-1) = \mathbf{F}\mathbf{z}(n-1|n-1),$$

$$\mathbf{V}(n|n-1) = \mathbf{F}\mathbf{V}(n-1|n-1)\mathbf{F}^t + \mathbf{G}\mathbf{Q}\mathbf{G}^t,$$

フィルタリング

$$\begin{aligned}\mathbf{K}(n) &= \mathbf{V}(n|n-1)\mathbf{H}^t(\mathbf{H}\mathbf{V}(n|n-1)\mathbf{H}^t + \Sigma_{\mathbf{w}})^{-1}, \\ \mathbf{z}(n|n) &= \mathbf{z}(n|n-1) + \mathbf{K}(n)(\mathbf{y}(n) - \mathbf{H}\mathbf{z}(n|n-1)), \\ \mathbf{V}(n|n) &= (\mathbf{I} - \mathbf{K}(n)\mathbf{H})\mathbf{V}(n|n-1),\end{aligned}$$

が逐次的に計算される。ここで  $\mathbf{z}(n|j)$  と  $\mathbf{V}(n|j)$  は、

$$\begin{aligned}\mathbf{z}(n|j) &\equiv E\{\mathbf{z}(n)|\mathbf{Y}(j)\}, \\ \mathbf{V}(n|j) &\equiv E\{\mathbf{z}(n) - \mathbf{z}(n|j)(\mathbf{z}(n) - \mathbf{z}(n|j))^t | \mathbf{Y}(j)\}\end{aligned}$$

により定義される条件付き平均ベクトルと、分散共分散行列を示す。

そして、20 によるデータの条件付き分布は、 $n = 1, \dots, N$  とすると、

$$\begin{aligned}f(\mathbf{y}(n)|\mathbf{Y}(n-1), \Theta, \bar{\mathbf{z}}(0)) &= \{(2\pi)^2 \det \Sigma(n)\}^{-1/2} \exp\{-1/2 \Delta \mathbf{y}^t(n) \Sigma(n)^{-1} \Delta \mathbf{y}(n)\} \\ \Delta \mathbf{y}(n) &\equiv \mathbf{y}(n) - \mathbf{H}\mathbf{z}(n|n-1), \\ \Sigma(n) &\equiv \mathbf{H}(n)\mathbf{V}(n|n-1)\mathbf{H}^t(n) + \Sigma_{\mathbf{w}}.\end{aligned}$$

によって定義される。

ハイパーパラメータベクトルは、数値的最適化で推定される。

$\Theta$  の最大対数尤度推定値を  $\hat{\Theta}$  とし、ハイパーパラメータの数を  $q$  とすると、モデルの AIC 値は、

$$AIC(q) = -2l(\hat{\Theta}, \bar{\mathbf{z}}(0)) + 2q, \quad (21)$$

により得られる。

観測値に対して、最適なモデルが選ばれたら、それぞれの成分の平滑値をフィルタリング値を用いて固定区間平滑化アルゴリズム

平滑化 ( $n = N-1, N-2, \dots, 1$ )

$$\begin{aligned}\mathbf{A}(n) &= \mathbf{V}(n|n)\mathbf{F}(n+1)^t\mathbf{V}(n+1|n)^{-1} \\ \mathbf{z}(n|N) &= \mathbf{z}(n|n) + \mathbf{A}(n)\mathbf{z}(n+1|N) - \mathbf{z}(n+1|n) \\ \mathbf{V}(n|n) &= \mathbf{V}(n) + \mathbf{A}(n)\mathbf{V}(n+1|N) - \mathbf{V}(n+1|n)\mathbf{A}(n)^t.\end{aligned}$$

により、求めることができる。

### C 循環変動成分によるシステム解析

モデルの循環変動成分 (7) 式に仮定された多変量 AR モデルの係数行列と分散共分散行列が推定されると、変量間の関係を知るためのシステム解析が可能になる。

まず、周波数応答関数  $F(f)$

$$F(f) = [I - \sum_{m=1}^M A(m) \exp(-i2\pi fm)]^{-1},$$

がただちに計算される。ここで  $I$  は単位行列を、また  $*$  は共役転置とする。この周波数応答関数を用いて、循環変動成分のクロスパワースペクトル  $P(f)$

$$P(f) = F(f) \Sigma_x F^*(f) \quad (0 \leq f \leq 0.5),$$

が計算できる。  $P(f)$  により循環変動成分の物理的な解釈を得ることができる。また、  $i$  番目の対角成分  $P_{ii}(f)$  は、第  $i$  チャンネルのパワースペクトルを表す。

また、ある系列のイノベーションのパワーからの他の系列に対する寄与の割合は、ノイズ寄与率

$$r_{ji}(f) = \frac{|A_{ji}(f)|^2 \sigma_i^2}{\sum_{i=1}^k |A_{ji}(f)|^2 \sigma_j^2}, \quad (22)$$

により得られる。ここで  $\sigma_i^2$  は、分散共分散行列  $\Sigma_x$  の第  $i$  番目の対角成分である。

第  $i$  番目のチャンネルへインパルスを入れたときのシステムの応答は、式 (7) における  $x_i$  ( $i = 0, 1, 2, \dots$ ) を計算することにより求められる。そのためには、

$$x_i(j) = 0, \quad v_{x_i}(j) = 0 \quad (i = 1, 2, j = \dots, -2, -1) \quad (23)$$

と

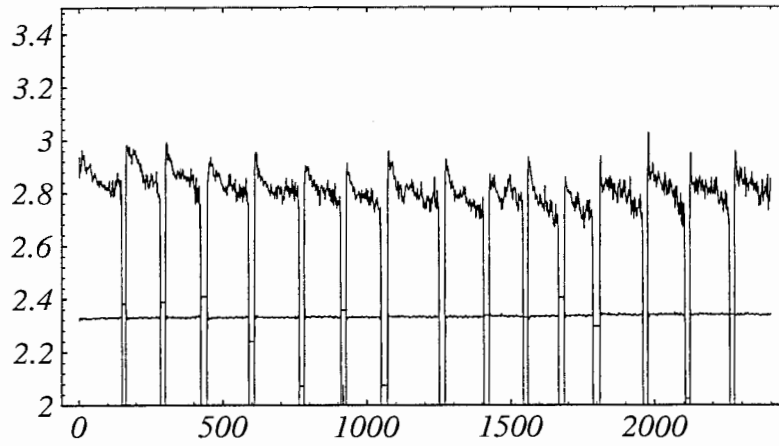
$$u_i(j) = \begin{cases} 1, & j = 0, \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

とおいたときに、

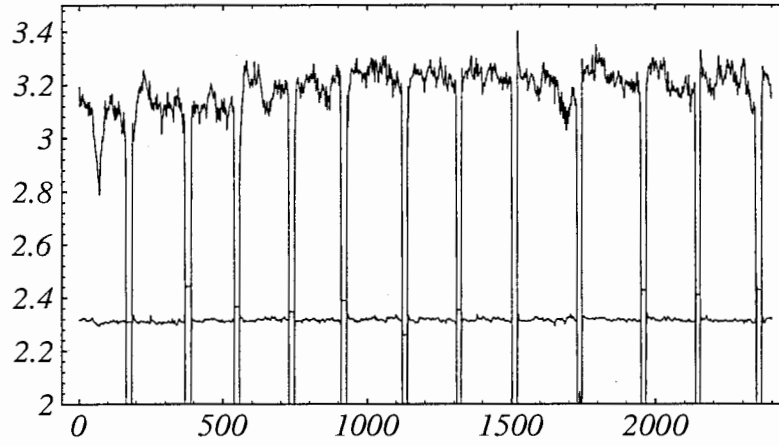
$$v_{x_i}(j) = u_i(j) - \sum_{m=1}^M a_{ii}(n) u_{i-m}(j) \quad (j = 0, 1, 2, \dots) \quad (25)$$

と変換しなければいけない。ここで、  $v_{x_i}(n)$  は  $x_i$  に加わるノイズ系列を示す。

*WO subject1.rms and F0*



*WM subject1.rms and F0*



*T subject1.rms and F0*

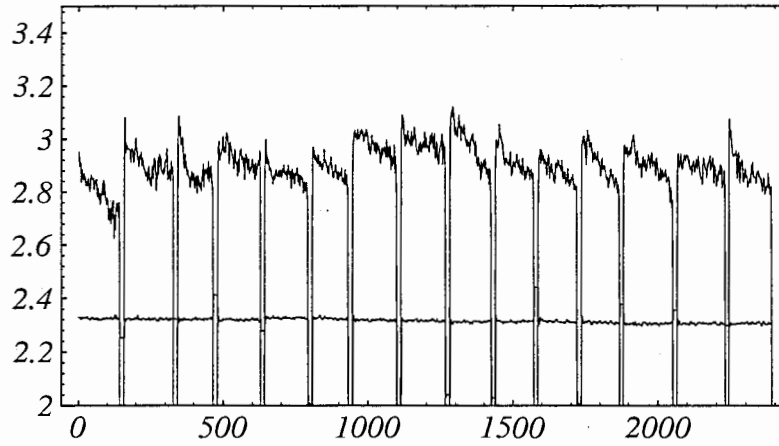


図 2: 状態 WO(上段)、WM(中段)、T(下段)のときの音声データから抽出した基本周波数  $F_0$  と振幅二乗平均値 rms の常用対数変換値. 横軸は  $\times 50$ msec を示す.



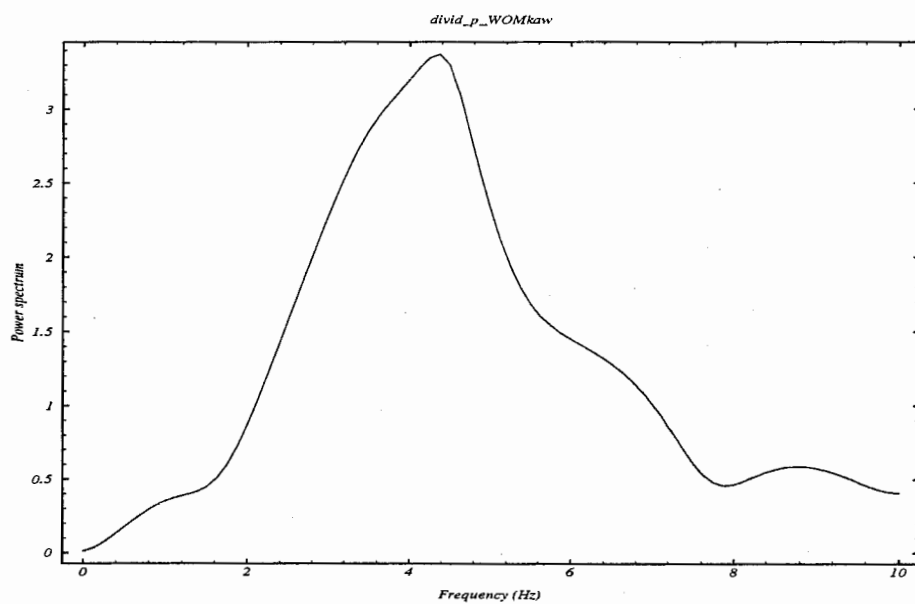
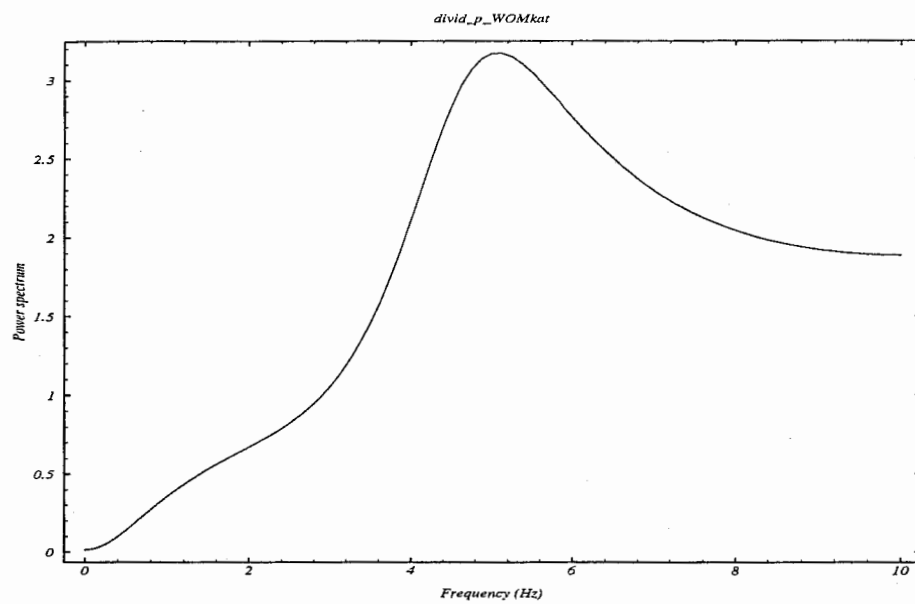


図 3: WO 状態のパワースペクトルを WM 状態のパワースペクトルで割算した値の平均値. 上段: subject1. 下段: subject2.

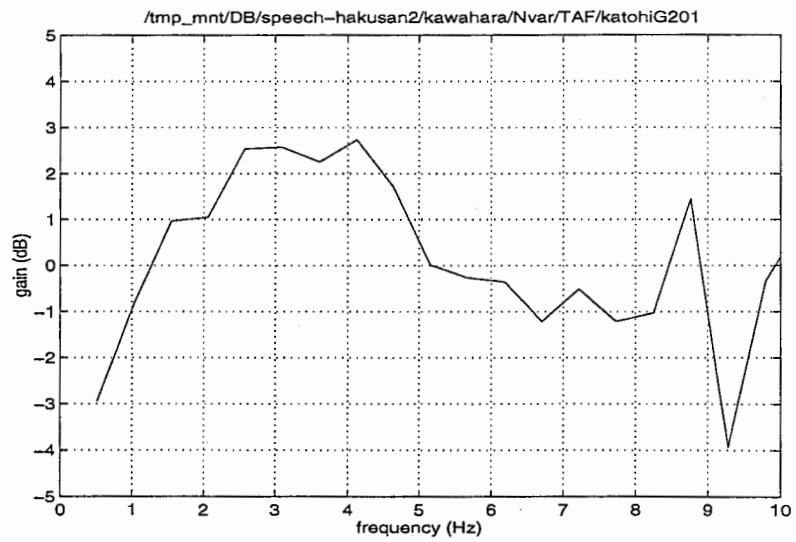


図 4: TAF 実験法で推定された subject 1 の PA の特性

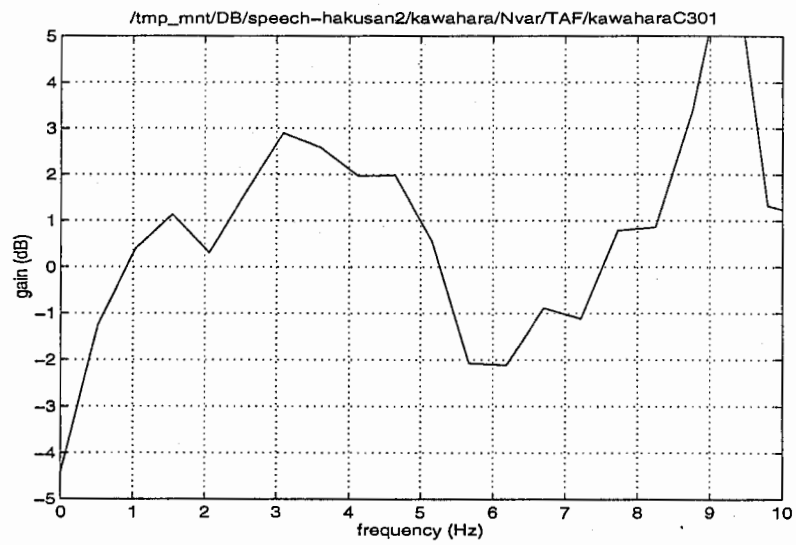


図 5: TAF 実験法で推定された subject 2 の PA の特性

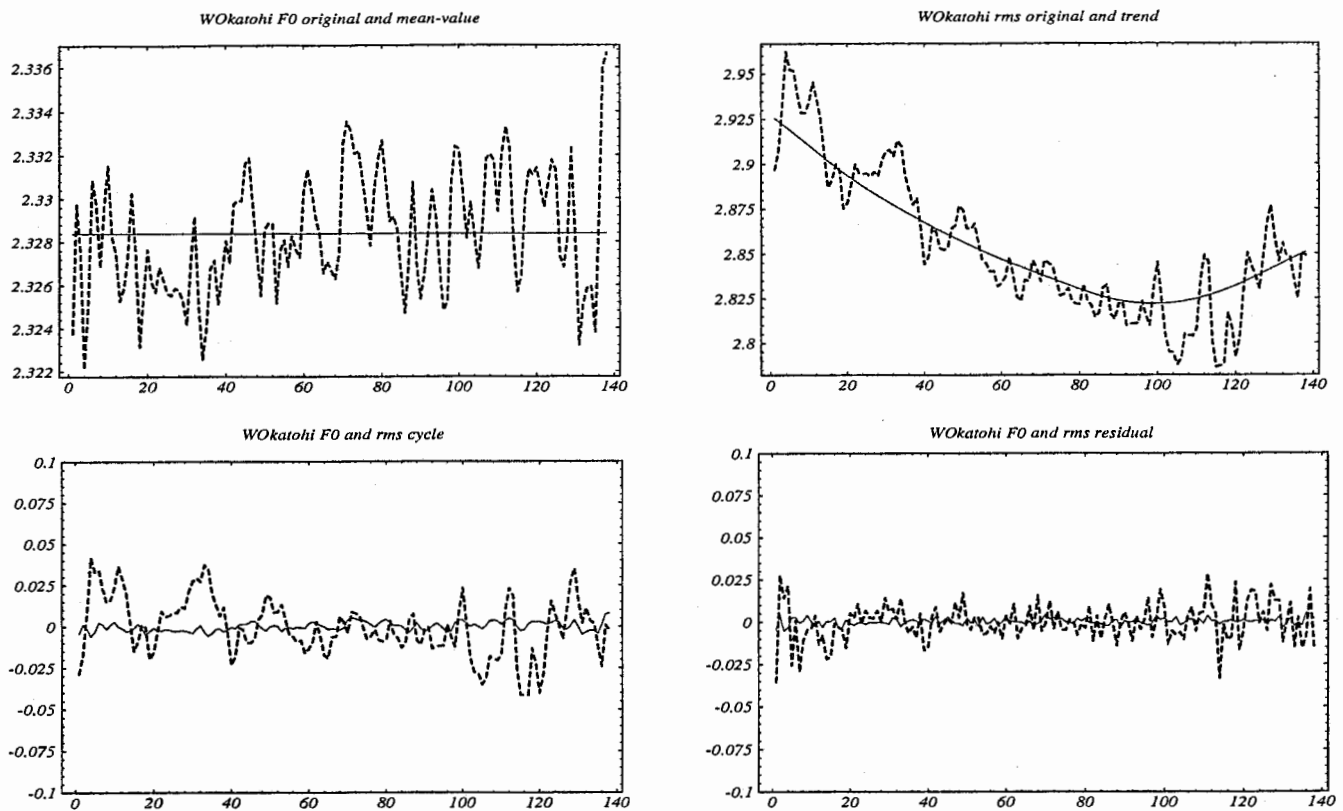


図 6: WO 状態 subject 1 の F0 と rms に平均非定常多変量フィードバックモデルをあてはめて分解されたそれぞれの成分と残差. 上段左: F0 の平均値 (実線) と原系列 (破線). 上段右: rms の平均非定常成分 (実線) と原系列 (破線). 下段左: F0 (実線) と rms (破線) の循環変動成分. 下段右: F0 (実線) と rms (破線) の残差.

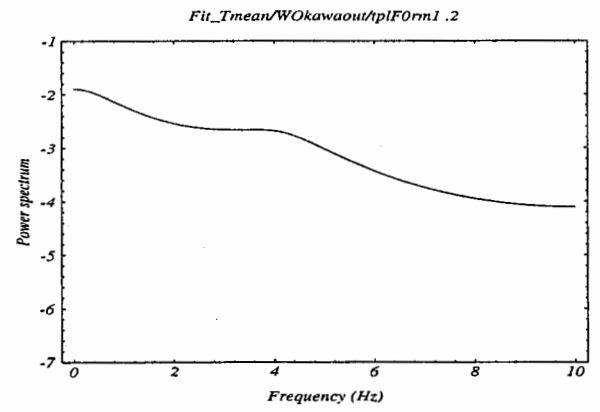
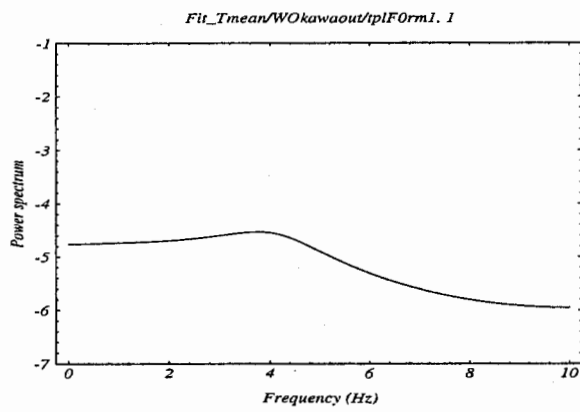
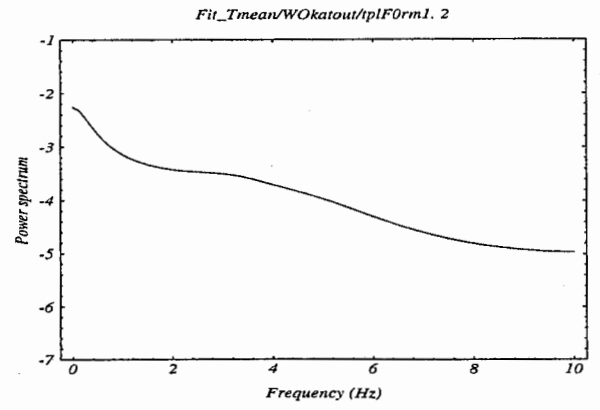
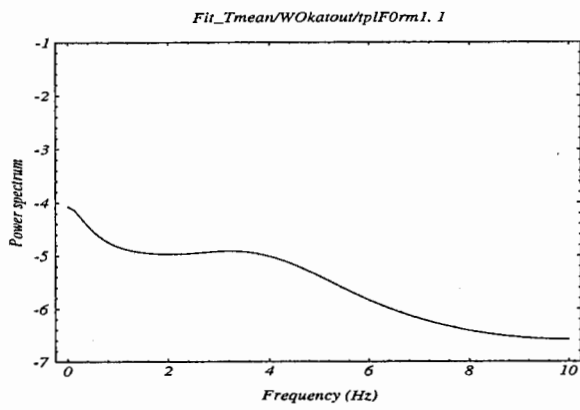


図 7: 状態 WO における推定された循環変動成分から得られた F0(左) と rms(右) のパワースペクトル. 上段: subject1. 下段: subject2.

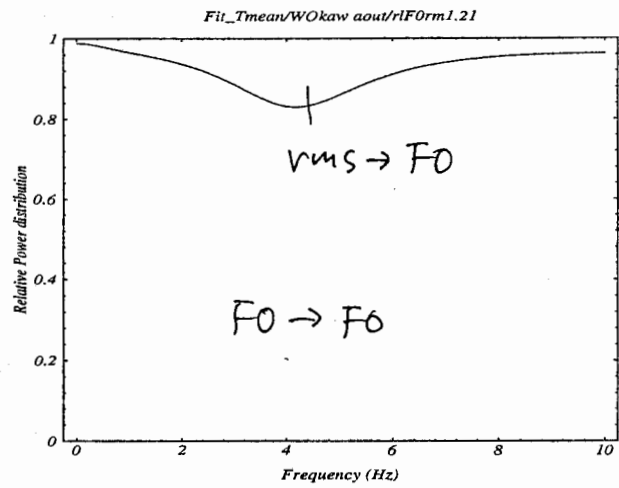
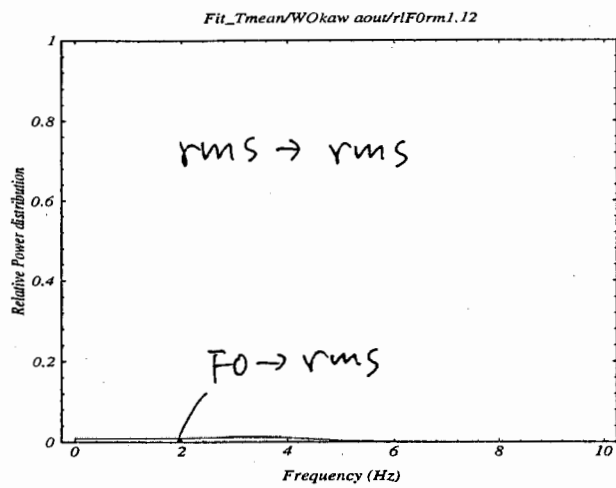
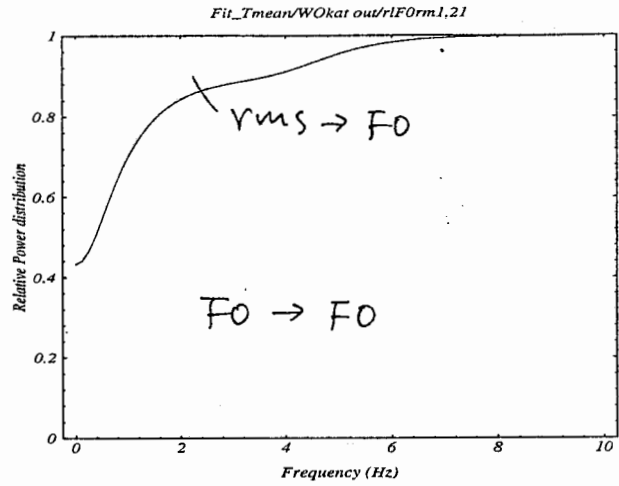
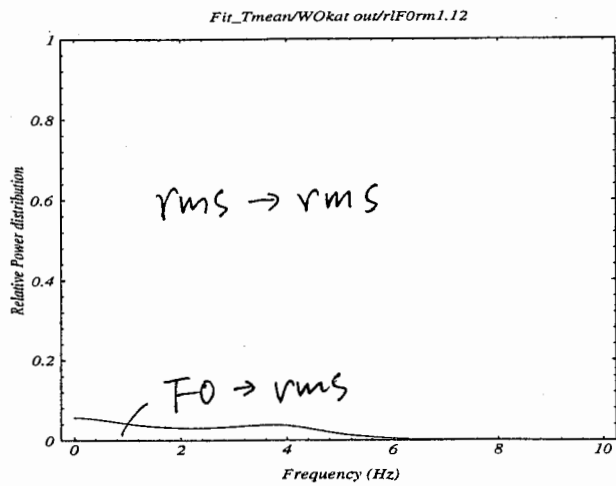


図 8: 状態 WO における推定された循環変動成分から得られたノイズ寄与率. 上段: subject1. 下段: subject2.

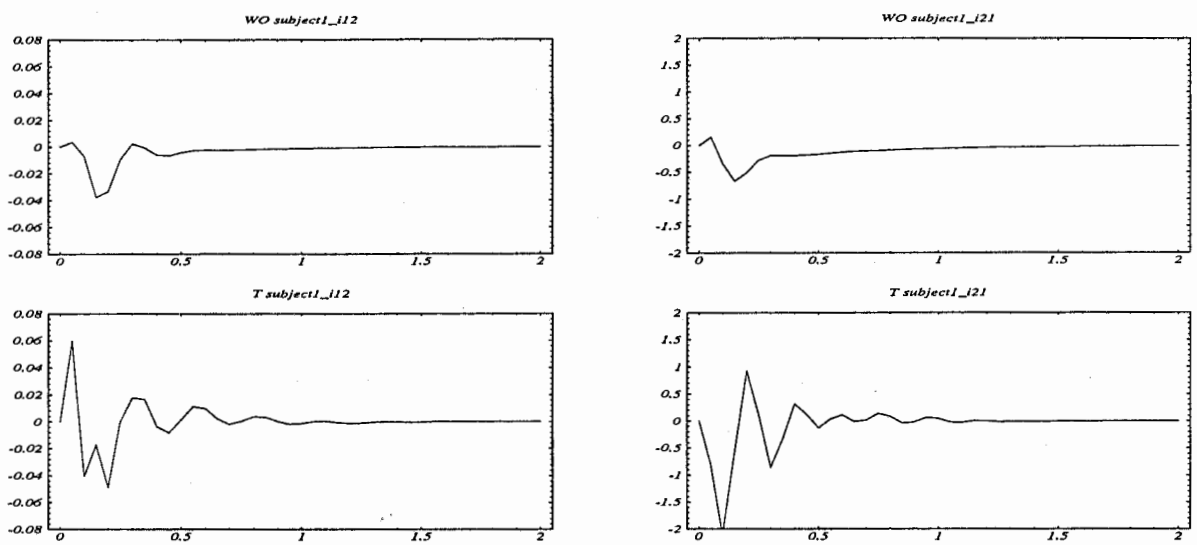


図 9: 状態 WO(上段) と TAF 条件下(下段)における subject1 のインパルス応答. rms にインパルスを入れたときの F0 の応答(左). F0 にインパルスを入れたときの rms の応答(右).

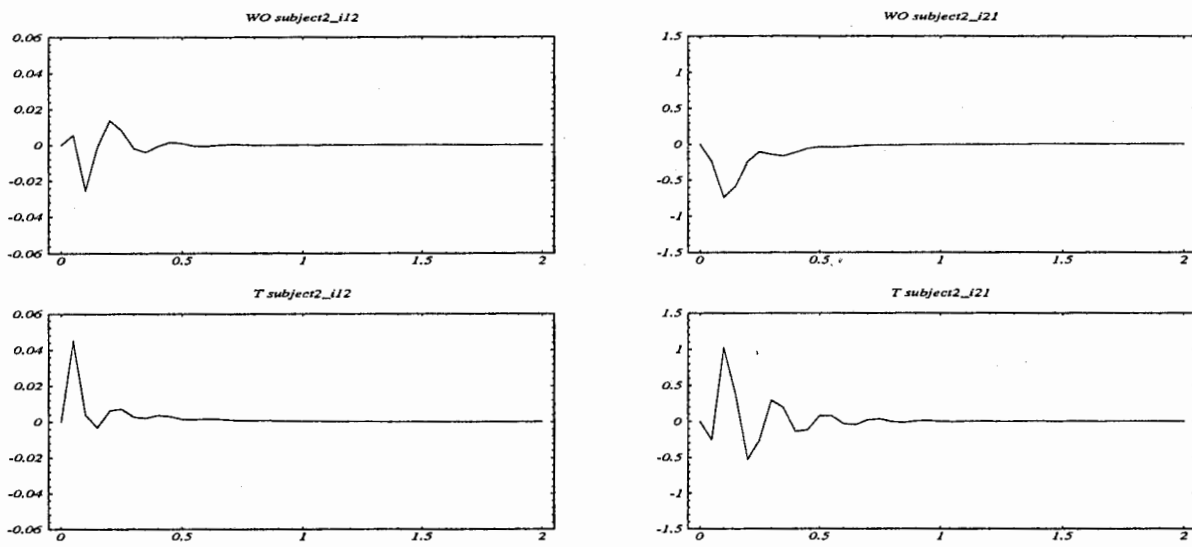


図 10: 状態 WO(上段)と TAF 条件下(下段)における subject2 のインパルス応答. rms にインパルスを入れたときの F0 の応答 (左). F0 にインパルスを入れたときの rms の応答 (右).