

TR - H - 169

0024

音環境による発声および行動の 変化計測に関する基礎検討

山本 泰宜

河原 英紀

(奈良先端技術大学院大学)

1995. 10. 9

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

音環境による発声および行動の 変化計測に関する基礎検討

山本 泰宜 (奈良先端技術大学院大学)

河原英紀 (ATR 人間情報通信研究所)

平成7年 10月 7日

Abstract

これまでの変換聴覚フィードバックの研究によって発声が聴覚からの情報によって変化することを実証してきた。本実習テーマは、このような相互作用が、音量や音場のようにより広範な人間行動の指標に対しても測定可能であるかを確認するための予備実験としてのデータ収集、および測定手法の確立を目的とする。具体的には、まず音場における外部音原からの聴覚情報による影響測定の基礎データとして実験機器の音響特性と音場での特性データの収集を行なった。次いで、基準条件としてピッチを対象とした変換聴覚フィードバックの基礎データを収集した。最後に同一の被験者に対して音量を対象として変換聴覚フィードバック実験を遂行し、音声パラメタ間の相互作用の解析を行なった。本資料では、これらの基礎データを紹介し、具体的な測定条件や今後の課題について述べる。なお、本研究は山本の学外実習として1995年9月4日から9月29日の期間に行われたものである。

1 はじめに

音声生成時に対する聴覚の関与を示す現象として、遅延フィードバック時の発声障害、騒音下での発声に伴う Lombard 効果、失聴者の聴覚情報遮断に伴う発声の変化、音咽頭筋反射等が報告されていた。しかし、これらからは、通常の発声時に実際に聴覚情報がどのように音声生成過程の制御に利用されているかについての定量的な評価は困難であった。

河原らは、通常の発声時における音声知覚と生成の相互作用を定量的に明らかにすることを目的として、変換聴覚フィードバック (TAF: Transformed Auditory Feedback) と呼ばれる手法を開発した。この手法を用いることにより、音声の基本周波数の制御において、聴覚を介して約 150ms の遅れを伴う補償系と約 400ms あるいはそれ以上の遅延で特徴づけられる系が発声時に並列に動作していることが見いだされた。[5] [6] これらの系が、TAF という人工的な条件の下だけでなく自然な発声においても働いており、またそのパラメタが TAF で求められたものと定量的に一致することが、音声の基本周波数の揺らぎの時系列解析により確認されている。[7] ところで、音声生成機構における周波数の制御は、喉頭周辺の様々な筋肉の協調動作により実現されており、周波数の制御は、音量やスペクトル傾斜などの他のパラメタの変化を伴う。実際、多次元自己回帰モデルを用いた基本周波数と音量の多次元時系列の分析結果は、基本周波数と音量との間に強い相互作用があることを示唆している [4]。これらの相互作用には、発声機構のダイナミクスに依存する成分と聴覚系を介した制御を反映した成分とが含まれているものと考えられる。

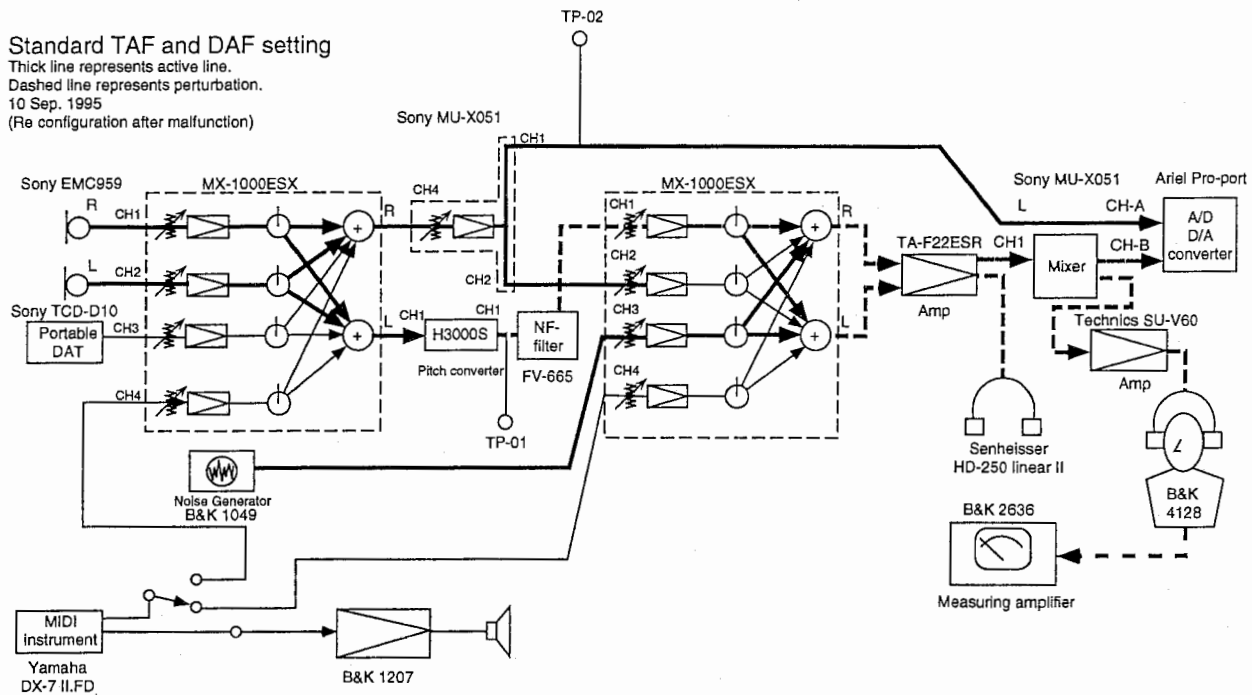


図 1: 実験装置の接続図

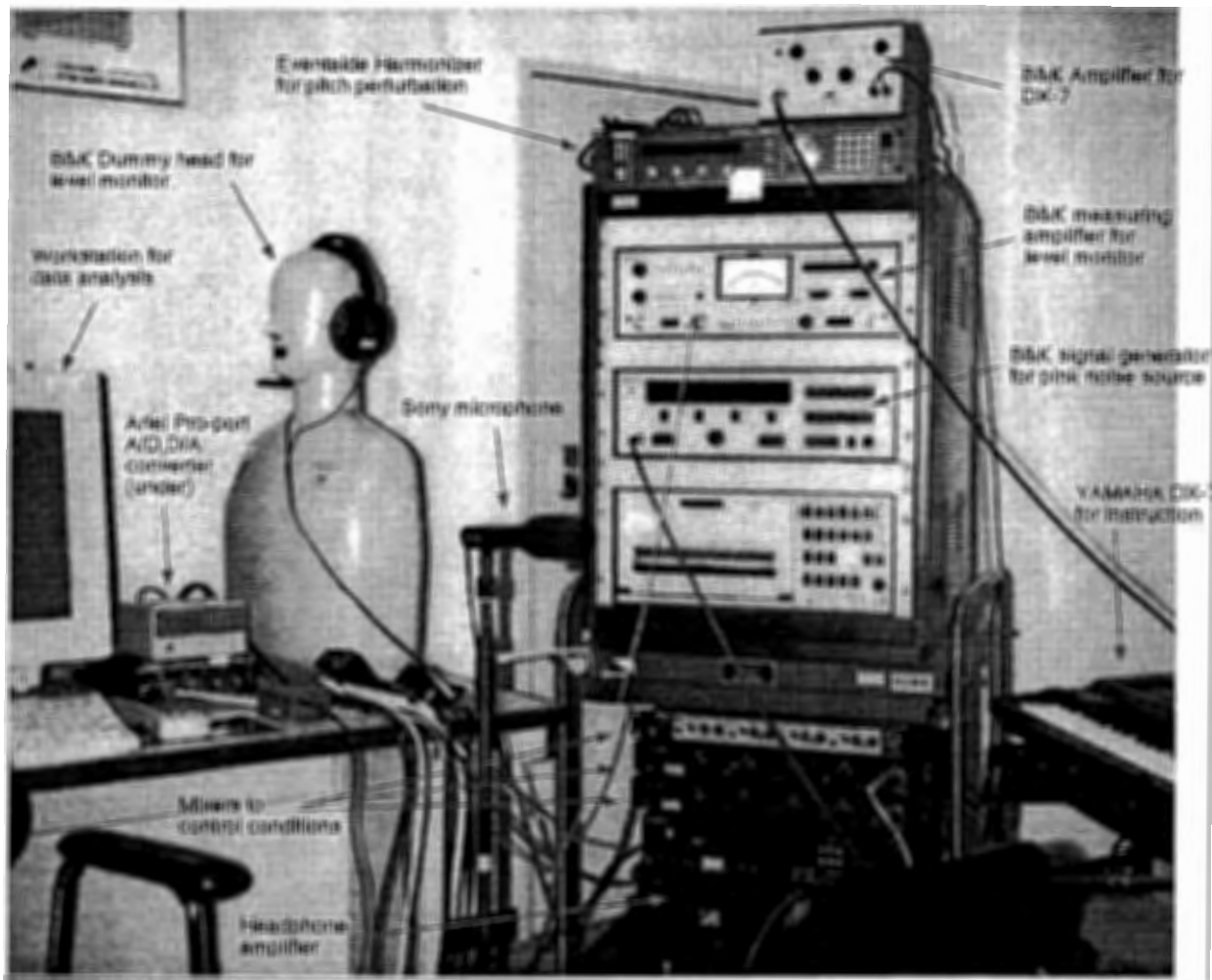
本報告では、この相互作用における聴覚の寄与を明らかにすることを狙い、従来のピッチ変調の他に音量を変調した摂動を加えた実験を行った結果について報告する。また、将来 TAF を音場において実施するための基礎データを得ることを狙い、音響系および電気音響変換系の特性の測定を行った結果についても報告する。以下では、まず、前半で実験系の構成と音響特性について説明し、後半で、その測定系を用いて行った音量の TAF に対する応答特性について説明する。

2 測定系の構成

TAF では、自然な側音の経路を、人工的なフィードバック経路で置き換え、音声のパラメタに変換を加えてその影響を解析する。この人工的なフィードバック経路は、基本的には、マイク、音声変換装置、アンプ、密閉型ヘッドフォンで構成される。実際には、様々な実験条件の設定と、安全措置としての音量のモニター、残留する骨導音等の影響をマスクするための雑音付加等のために、図 1 に示すような複雑な構成となっている。

図は、変換聴覚フィードバック (TAF) と遅延聴覚フィードバック (DAF: Delayed Auditory Feedback) の場合の設定を示す。各種の実験条件は、Mixer の入力ボリュームとパンポットにより設定する。図中の太線は、この実験条件において有効となっている信号経路を示す。破線は、変調された信号が流れていることを示す。

実験室での実際の実験状況における装置構成の写真を図 2 に示す。



Physical setup for TAF experiments

図 2: 実験装置の構成 (写真)

2.1 構成機器の概要

ここでは、TAF の実験系を構成する各々の機器の概要と実験系における役割について簡単に説明する。

1. ハーモナイザー (Eventid H3000S Ultra-Harmonizer) 音楽用音声加工装置。ピッチ変換、振幅変換用に使用。各種パラメタを MIDI 信号により制御。
2. 計測器。(B&K 2636) 計測アンプ。ダミーヘッドの人工耳を用いて、実験中の被験者の受聴する音圧をモニタするために使用。
3. 雑音源 (B&K 1049) 汎用信号発生装置。被験者が自己の発声した音声を聴くことで受ける影響を、本装置で発生させたピンク雑音を混合することにより軽減するために使用。
4. マイク (Sony ECM 959)。人工的フィードバック経路の入力部分。単一指向性ステレオマイク。(従来からの経緯で用いているが、モノラルマイクにした方が良い。) 特性

測定を確実にを行い、近接効果の影響を確認しておけば、使用可能。

5. ヘッドフォン (Senheisser HD-250 linear II)。人工的フィードバック経路の出力部分。平原らによる測定結果に基づき、耳覆い型のヘッドフォンの中から選定。低域特性が比較的平坦。
6. ダミーヘッド (B&K 4128) 内部にスピーカ内蔵。口 (人工口) の部分から音を出力。口の前の治具は、基準音圧測定用の 1/8" マイクの装着用。両耳にはコンデンサマイク内蔵。実耳のインピーダンスを近似的に実現しており、実耳レスポンスの代用として用いる。
7. ミキサー (Sony MX-1000 ESX and one Sony MU-X051) 写真の装置の最上段は MU-X051、上から 2 段目と 3 段目の装置は、MX-1000ESX。実験条件の設定は、これらのミキサーのボリュームおよびパンポットの操作により行なう。MU-X051 は、出力を分岐するために使用。9 月 5 日以降は、最上段に MU-X051 を更に一個増設し、故障した DAT を置き換えた。さらに外部にアンプ (Technics SU-V60) を加え、発声時の音量のモニタとした。
8. DAT (Sony DTC-1000ES) A/D 変換器への入力の前段のバッファアンプとして使用。9 月 5 日に故障したため、以降の実験では構成から外した。
9. A/D,D/A 変換器 (Ariel Pro-port) Sparc Server M660 (hakusan) の DSP 拡張ボードに接続。ボードとの間は、シリアル接続。
10. アンプ (Sony TA-F22ESR) 被験者用のヘッドフォンの駆動。
11. フィルター (NF Electronic FV-665) 46dB/oct の遮断特性を有するフィルタ 2 個を接続することにより、様々なフィルタ特性を実現。周波数選択的フィードバックの実験に使用。他の実験では、THRU としておく。
12. MIDI key board (Yamaha DX-7) 被験者への発声の音程の教示に使用。また、自己の発声を完全にマスクするための正弦波音を作成する場合にも使用。
13. Amplifier (Technics SU-V60) 音圧モニタ用のヘッドフォンを駆動するために使用。
14. 多チャンネルデジタルレコーダ (TASCAM DA-88) pitch TAF では未使用。音響特性等の測定で使用。8 チャンネルのアナログ/デジタル入出力を備える。同時録音/再生が可能。この機能により、高精度の時間軸整合が必要な M 系列による伝達特性測定を容易に行なうことができる。
15. Mac (Macintosh fxII) Harmonizer の制御用の MIDI 信号の送出に使用。制御用の MIDI 信号は、MATLAB により作成した Text ファイルを UNIX 上の t2mf により standard MIDI ファイルに変換した後、Mac に転送して File Typer を用いてクリエータとファイルタイプを変換して MIDI ファイルとし、Mac 上のシークエンサーソフト (Performer) により、無限ループを設定した後に Performer 固有フォーマットで格納することにより作成した。実験時の MIDI 信号の送出には、同じ Performer を使用した。



図 3: 多チャンネルデジタルレコーダ (写真)

2.2 機器の校正

それぞれの機器の校正について以下に簡単に説明する。

1. ハーモナイザー: 全体の入力レベルと出力レベルの設定を確認。既定値は、入力側-5dB、出力側-2dB。ここで報告する全ての実験を通じて、この既定値を使用した。クリップする場合には、調整が必要。実験の開始にあたっては、実験条件に適合したプログラムがロードされていることを確認。(プログラム 14 番がピッチ変換の TAF、プログラム 20 番が音量変換の TAF。) バイパス状態になっていないことを確認。(電源スイッチ右下の押しボタン付属の LED が赤く点灯していれば OK。)
2. 計測器: TAF の被験者の受聴音量のモニタ用には、以下のようにパネルを設定する。入力 30dB レンジ、出力 0dB、Detection log 30dB、fast、コンデンサマイク用プリアンプ入力、A 加重と帯域制限。
3. 雑音源: 20Hz~20kHz 帯域の pink noise が continuous 出力状態になっており、出力電圧が 100mV であることを確認。
4. マイク: 指向性が 90 度に設定されるようにスイッチを入れる。
5. ヘッドフォン: TAF および DAF の実験では、被験者の耳にするヘッドフォンから 78dB(A) のピンクノイズが出るように各部のゲインをまず設定する。現在の調整状態では、特別なことの無い限りアンプのゲインを調整するだけで良い。発声時に被験者が聞いている音量をモニターできるように、モニタ用アンプのゲインを調整して、ピ

ンクノイズのみの出力時に、モニター用アンプに接続されているヘッドフォンの左耳から出る雑音のレベルが 78dB(A) になるように設定しておく。

6. ダミーヘッド：校正は別途音響校正についての節で説明する。通常は、人工耳からの出力の読みが実音圧を表示するように計測アンプのゲインを設定しておく。
7. ミキサー：詳しくは、実験条件の設定の所で説明する。
8. A/D,D/A 変換器：過大入力とならないように、入力ボリュームを調整。
9. フィルター：部分変換聴覚フィードバックを行わない場合には、接続を THRU の位置にしておく。
10. DX-7: 被験者へのガイド音の提示用。ガイド音のレベルは、聞きやすい音量に設定。被験者の発声のマスク用の純音、疑似音声の発生用として使用する場合には、被験者の耳で所望の音圧が出るように、Mixer の第 4 チャンネルの入力レベルにより調整。
11. 多チャンネルデジタルレコーダ：録音インディケータの赤色 LDE が点灯しないように入力レベルを調整。
12. Mac: ピッチ変換用の制御信号の送出。信号の送出中に MIDI インタフェースの OUTPUT インディケータが点灯していることを確認。ピッチ変換 TAF は、P/H3000Smq5NoShift.pfm を使用。音量変換 TAF の実験では、P/intrpm5d15a.pfm と P/intrpm5d15b.pfm を使用。

3 実験系の基準特性の測定

TAF の実験系の音響特性については、1993 年 1 月から 2 月にかけての学外実習において豊橋技術科学大学の岩谷により測定された [1]。しかし、この測定においては、B&K の装置の既定のプログラムによる 1/3 オクターブ点での測定が行なわれたに過ぎない。音場での TAF を行なうためには、時間的応答を正確に測定することが必要である。ここでは、TSP 信号を用いた方法 [8] と M 系列信号を用いた方法とにより測定を試みた。その結果、金田の指摘するように [2] M 系列信号では非線形歪による影響が測定区間中にランダムに分散し実際の音響的反射と区別できないため、以下では、TSP を用いた結果についてのみ報告する。

3.1 音響系の校正

音響系の校正は、測定系と実験系の二段階に分かれる。測定系の校正は、毎回繰り返す必要は無い。毎週確認を行えば良い。この校正された測定系を用いて、実験系の基準特性を測定しておく。この基準測定に基づいて、実験系は、原則として測定毎に確認を繰り返す。

3.1.1 測定装置の校正

B&K の測定装置は多くの付属品を用いて測定を行うように設計されている。それらの付属品は、心理実験ブースの前にある鍵のかかるキャビネット内に収納されている。また、コンデンサマイクのヘッド等、湿気を嫌う部品については、その隣の乾燥キャビネットの中に格納されている。これらの付属品は、使用後に元の場所に戻すよう注意する。

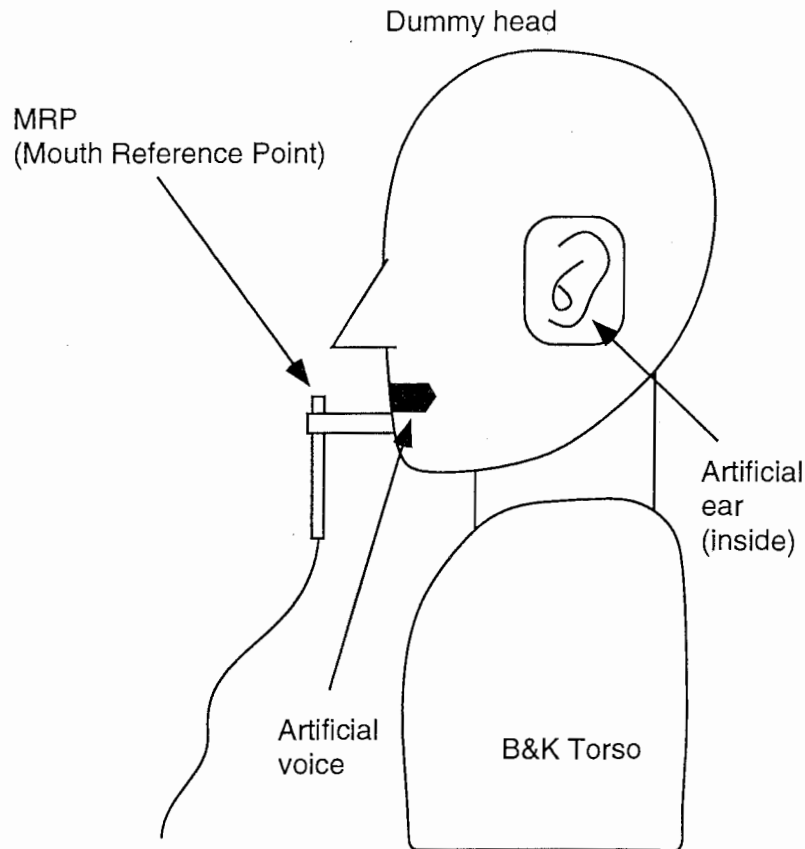


図 4: ダミーヘッドに関連する規定点

1. ダミーヘッド人工耳および計測アンプ：ピストンフォンを用いて、計測アンプの指示が実音圧を表示するように調節する。調整は、計測アンプの周波数荷重を flat に設定して行なう。調節後は、計測アンプの周波数荷重を A 特性に戻しておく。メータ表示用の Detection circuit は、メータの表示盤（交換可能）に適合する設定になっていることを確認する。表示盤左上の valid scale の表示が all level になっていれば正しい。1995 年 9 月 10 日の状態では、0.2dB 音圧を過大表示する。
2. ダミーヘッド人工声およびプリアンプ：ピストンフォンを用いて、MRP (mouth reference point) に置くコンデンサマイクを校正する。MRP は、ダミーヘッドの口の部分にある治具に 1/8" マイクを装着することで設定できる。計測アンプの指示と実音圧の関係が定められた値となるように調節する。1995 年 9 月 10 日の状態では、表示された音圧に 21.9dB を加えると実音圧が求められる。これらの規定点を図 4 に示す。

3.1.2 実験系の基準特性の測定

多チャンネルでの利得ならびに位相特性の測定が必要であるため、マルチトラックデジタルレコーダを用い、M 系列あるいは TSP 信号を用いて測定を行なう。特に、人工声を用いた音響計測では、音源の非線形性が問題となるため、MRP での音圧レベルが 80dB (flat 荷重) の場合について測定する。

1. 自然側音経路の特性測定：MRP (mouth reference point) から人工耳までの特性を測定する。人工声を適切な信号で駆動し、MRP のマイクと人工耳のマイクの出力の比

として伝達特性を測定する。計測器の flat 荷重での指示値を用いて絶対レベルを校正。

2. ヘッドフォン特性の測定：電気入力対音圧特性の測定。ヘッドフォンはダミーヘッドに装着して測定。計測器の flat 荷重での指示値を用いて絶対レベルを校正。
3. ヘッドフォン装着時の側音特性の測定：ヘッドフォンに対する電気信号が存在しない状態で、MRP のマイクと人工耳のマイクの出力の比として伝達特性を測定する。計測器の flat 荷重での指示値を用いて絶対レベルを校正する。
4. ヘッドフォン装着・人工的側音の存在する場合の特性測定：TAF や DAF の実験と同じ状況において、MRP のマイクと人工耳のマイクの出力の比として伝達特性を測定する。計測器の flat 荷重での指示値を用いて絶対レベルを校正する。この特性は、TAF 実験用のマイクロフォンの位置とアンプ系のゲイン設定の関数となるため、以下を基準系として定義する。マイク位置は、発話者の唇前方 20cm。被験者が受聴する雑音レベルは 78db(A)。500Hz におけるゲインは 0dB。
5. フィルタ特性の測定。挿入される HPF および LPF の特性特に位相および群遅延特性を測定する。測定には、TSP あるいは M 系列信号を用いる。

3.2 測定例

基準特性の測定例を示す。測定時の装置の配置を図 5 に示す。実験では、ダミーヘッドを椅子にすわらせて (position 1)、TAF の実験姿勢を模し、以下の条件で MRP (mouth referende point) から人工耳までの特性を測定した。ここでは、人工声として TSP 信号を用い [8]、MRP での音圧レベルが 80dB になるように設定した。TSP 信号の長さは、16384 サンプルであり、48kHz16bit で再生された。したがって TSP 信号の繰り返し周期は、341.3ms となる。A/D 変換は、この再生に同期して行なわれた。これらの条件の下で測定された周波数特性を図 6 ~ 図 11 に示す。また、周波数特性を計算するための応答波形は最初の 1000 サンプルを用いた。これは、20.83ms であり、直接音よりも 7m 長い伝播経路の反射波までを対象としたことを意味する。測定時間を 30 秒とし、同期加算により S/N の向上を図った。同期加算回数は、87 回である。

条件 1 自然側音特性の測定 (1)。

ダミーヘッドはヘッドフォンを装着せずマイクを置かない。

条件 2 自然側音特性の測定 (2)。

ダミーヘッドはヘッドフォンを装着せず、MRP から 20 cm の距離にマイクを置く。人工的側音経路を構成する機器の電源は切。

条件 3 ヘッドフォン装着時の自然側音特性の測定。

ヘッドフォンを装着し、ダミーヘッドの MRP から 20 cm の距離にマイクを置く。人工的側音経路を構成する機器の電源は切。

条件 4 人工的側音経路の特性測定。

ヘッドフォンを装着し、ダミーヘッドの MRP から 20 cm の距離にマイクを置く。人工的側音経路を構成する機器の電源は入。

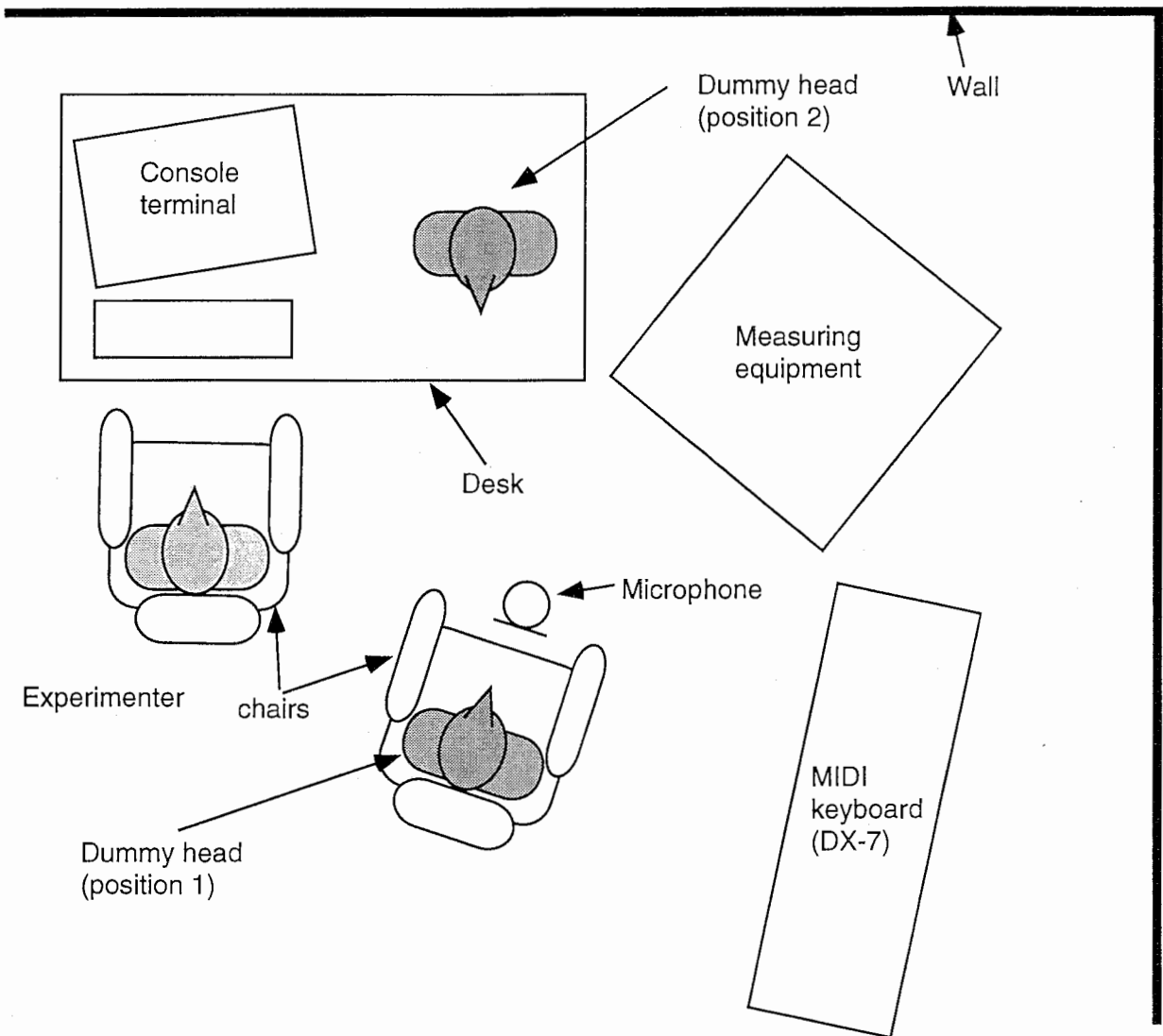


図 5: 測定時の機器配置

条件 5 マイク系の特性測定。

ダミーヘッドを position 2 に設置し、MRP から TP02 までの特性を測定する。

条件 6 ヘッドフォン系の特性測定。

ダミーヘッドを position 2 に設置し、2 段目の MX-100ESX の 4ch から人工耳までの特性を測定する。

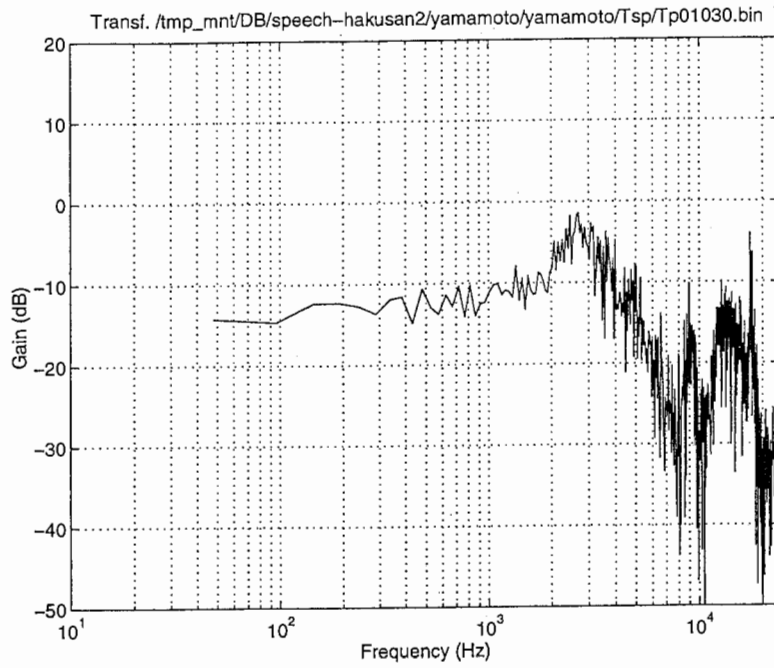


図 6: 自然側音 (1) (条件 1) の周波数特性

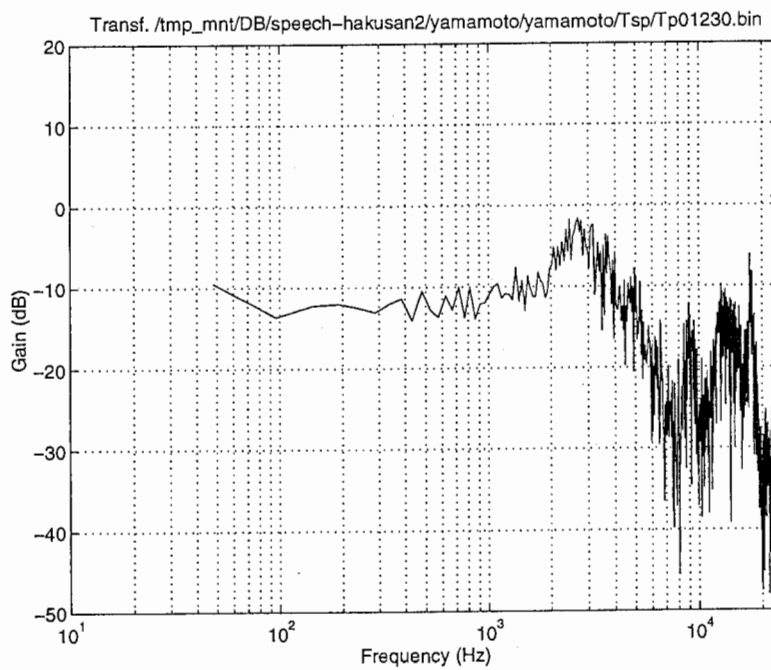


図 7: 自然側音 (2) (条件 2) の周波数特性

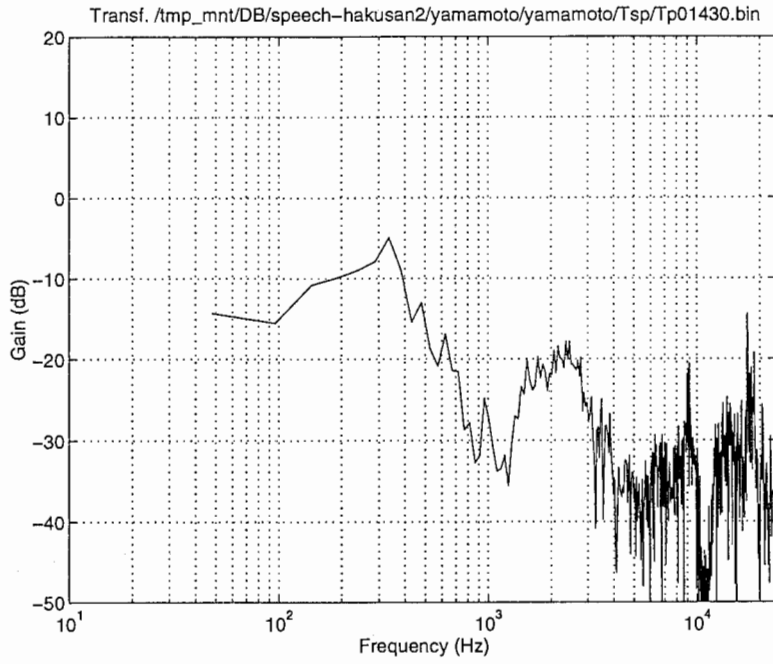


図 8: ヘッドフォン装着時の自然側音（条件 3）の周波数特性

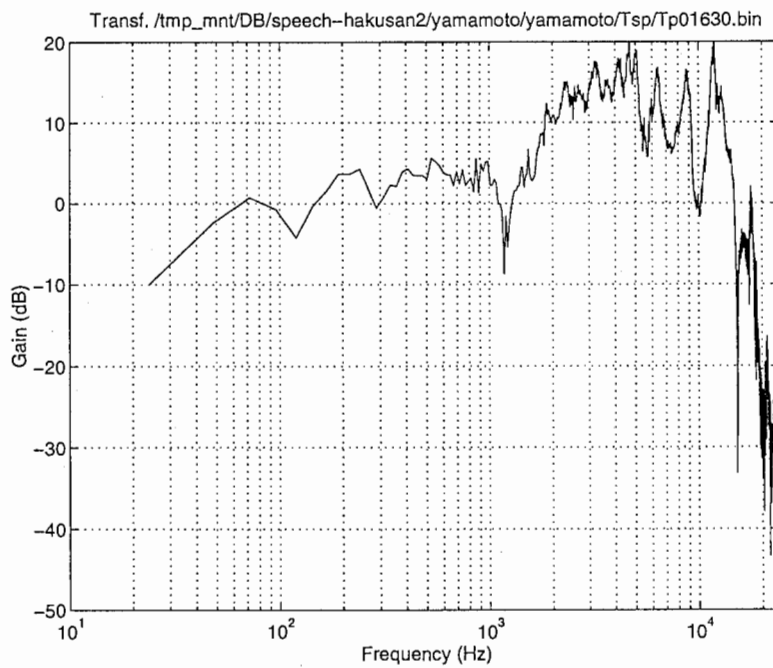


図 9: 人工的側音経路（条件 4）の周波数特性

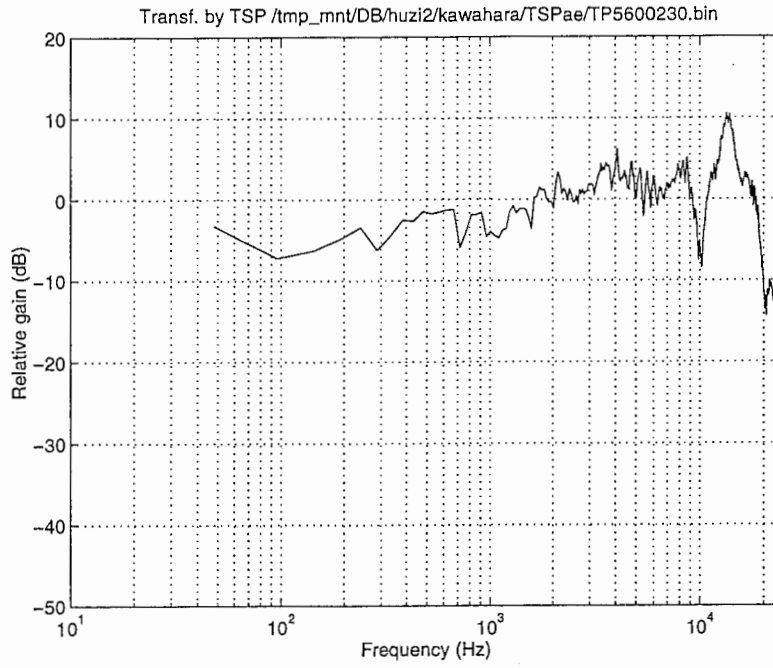


図 10: マイク系 (条件 5) の周波数特性

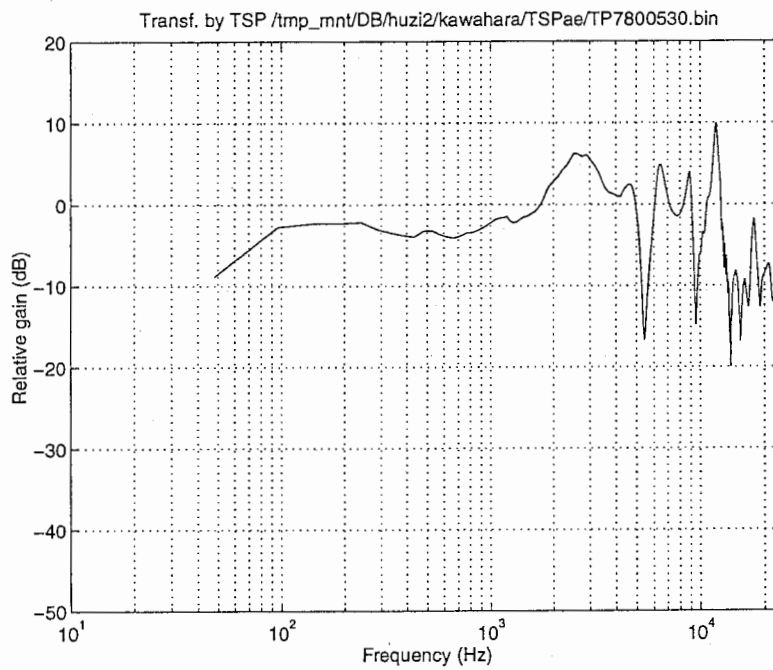


図 11: ヘッドフォン系 (条件 6) の周波数特性

3.3 測定結果

今回の測定結果は、1/3 オクターブの分解能の岩谷の測定に認められた特徴を共有しており、これらの測定間で再現性が保たれていることを示している。これらの共通の特徴に加え、本測定では、高い周波数分解能が確保されているため、前回の測定では明らかではなかった高い周波数領域での共振や反共振が明瞭に確認できる。

- 条件1 (図6) と条件2 (図7) の差は、マイクの有無である。特性にほとんど違いがないことからダミーヘッドのMRPから20cmの距離に置かれたマイクによる反射の影響はほとんどないことが分かる。3kHzの山は外耳道の共振によるものであり、5kHz以上でのゲインの低下は、頭部の遮蔽効果によるものである。周波数特性上の細かな山谷は、近くの物体からの反射波によるものである。図12に最初の20msの時間応答を示す。8ms, 12ms, 18msの周辺に明瞭に反射の影響が認められる。
- 条件1 (図6) と条件3 (図8) の差は、ヘッドフォンの装着の有無である。400Hz以上の成分が、条件1と比較すると大きく減衰しており、密閉による気導音の遮蔽効果を示している。200Hzから400Hzにかけては、自然側音の場合よりもゲインがやや高い。
- 条件3 (図8) と条件4 (図9) の差は、人工的側音の有無である。気導音の影響を避けるため、人工的側音経路のレベルを高く設定している。両者の差が最小の部分でも、人工的側音経路のゲインは10dBほど高い。したがって、気導音のみが自己の発声のフィードバックに寄与しているのであれば、この条件でフィードバックされる音声では人工的側音が支配的になる。
- 条件4 (図9) と条件1 (図6) の特性を比較すると、4kHzまではほぼ平行した形状であることが分かる。これは、音声の重要な帯域に含まれる成分に有害な偏りが生じないことを意味する。しかし、5kHz以上においては、人工的側音のゲインは遥かに高く、聴感上高域に偏った音色であることに対応している。
- 条件5 (図10) のマイク系の特性では、多くの山谷が認められる。これらは、マイク固有のものではなく、通常室内での収録であることによる近くの物体からの反射波の影響と考えられる。単一指向性マイクで通常問題となる近接効果の影響は、今回の収録条件の下では現われていない。
- 条件6 (図11) のヘッドフォン系の特性は、音声の知覚に重要な第一第二ホルムントの存在する200Hzから2kHzまでの帯域で平坦である。これは、音韻知覚の研究にとって有用な特性である。

Impulse response of natural side tone Tp01030.bin

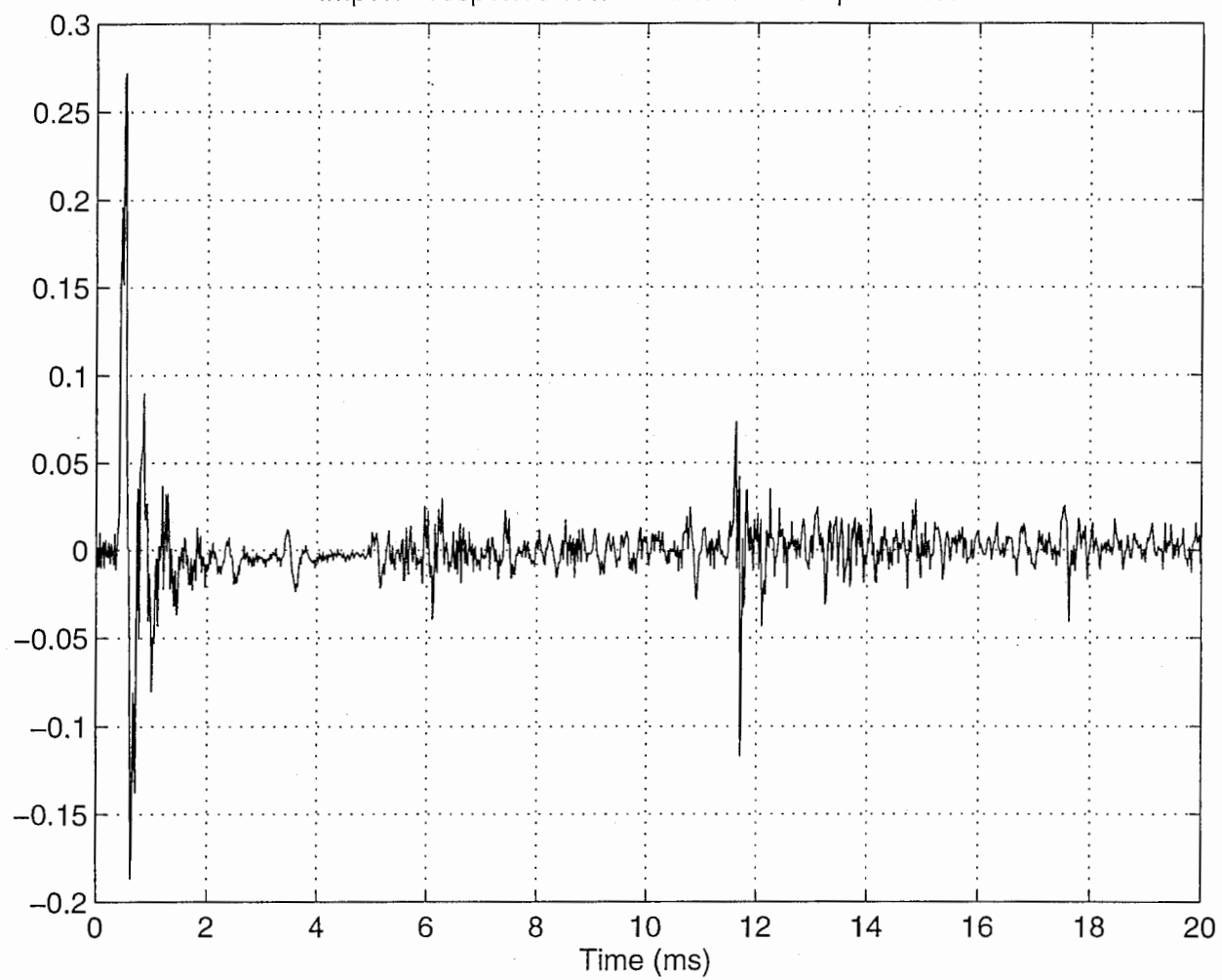


図 12: 自然側音経路のインパルス応答

4 変換聴覚フィードバック

以下では、このようにして特性が測定された実験系を用いて発声時の音声知覚と生成との相互作用を測定する。まず、変換聴覚フィードバックについて、簡単に整理する。

変換聴覚フィードバックは、通常の聴覚フィードバック経路を人工的な信号処理系を含む経路により置き換え、この経路を介してフィードバックされる音声のパラメタに系統的な摂動を加えることにより、その影響が発声された音声パラメタに与える影響を、摂動の特別な性質を利用して解析する方法である。原理的には、どのような音声パラメタに摂動を加えても良い。摂動は、正常な発声を妨害しないよう、検出可能な範囲でできるだけ小さな方が望ましい。また、意識的・無意識的な予測による誤差を避けるため、S/Nが許す範囲で長い周期を持つ、予測が困難な複雑な信号を用いる。

具体的な実験手続きは、以下のようなものである。被験者は唇をマイクから20 cmの距離において、1ないし2分間にわたって発声する。この時、被験者は、人工的なフィードバック経路を通じた変換された自己の発声を聞きながら発声する。発声された音声および発声時に被験者が受聴している音声は8 kHz, 16bitで量子化され計算機に取り込まれる。取り込まれた音声は50 ms長の \cos^4 窓を利用して分析開始位置を10 msずつずらしながらパラメタ抽出が行なわれる。摂動信号としてはM系列から作成した疑似白色信号を用いた。また気導音や骨導音によるフィードバックの影響を軽減するため、密閉型のヘッドフォンにするとともに、78dB(A)のピンクノイズを重畳させた。

これらのデータに基づいて、詳細は文献[5]に譲るが、以下のようなステップを踏んで分析を行なう。

- (1) 同期加算
- (2) 疑似白色信号との相互相関
- (3) 時間原点の推定と整列
- (4) 個別測定データの統合（測定を複数回に分けて行なった場合のみ）
- (5) 伝達特性と coherency の推定
- (6) 閉ループインパルス応答の推定
- (7) 開ループインパルス応答の成分への分解
- (8) 伝達特性の再構成と聴覚フィードバックによる影響の評価

4.1 ピッチ変換聴覚フィードバック

これまでの実験結果と対照させて検討するために、まず、ピッチを対象としたTAFの実験を行った。ピッチの変換量は、p-p値で50cent（半音の半分）とした。疑似白色信号の作成に用いたM系列は、長さが31の系列である。各ユニットは1/16秒の時間長を有する。これらの値は、これまでのTAFで用いてきた値であるが、後で討論の部分で説明する幾つかの理由から、詳細な検討に基づいた改訂が必要であると考えられる。

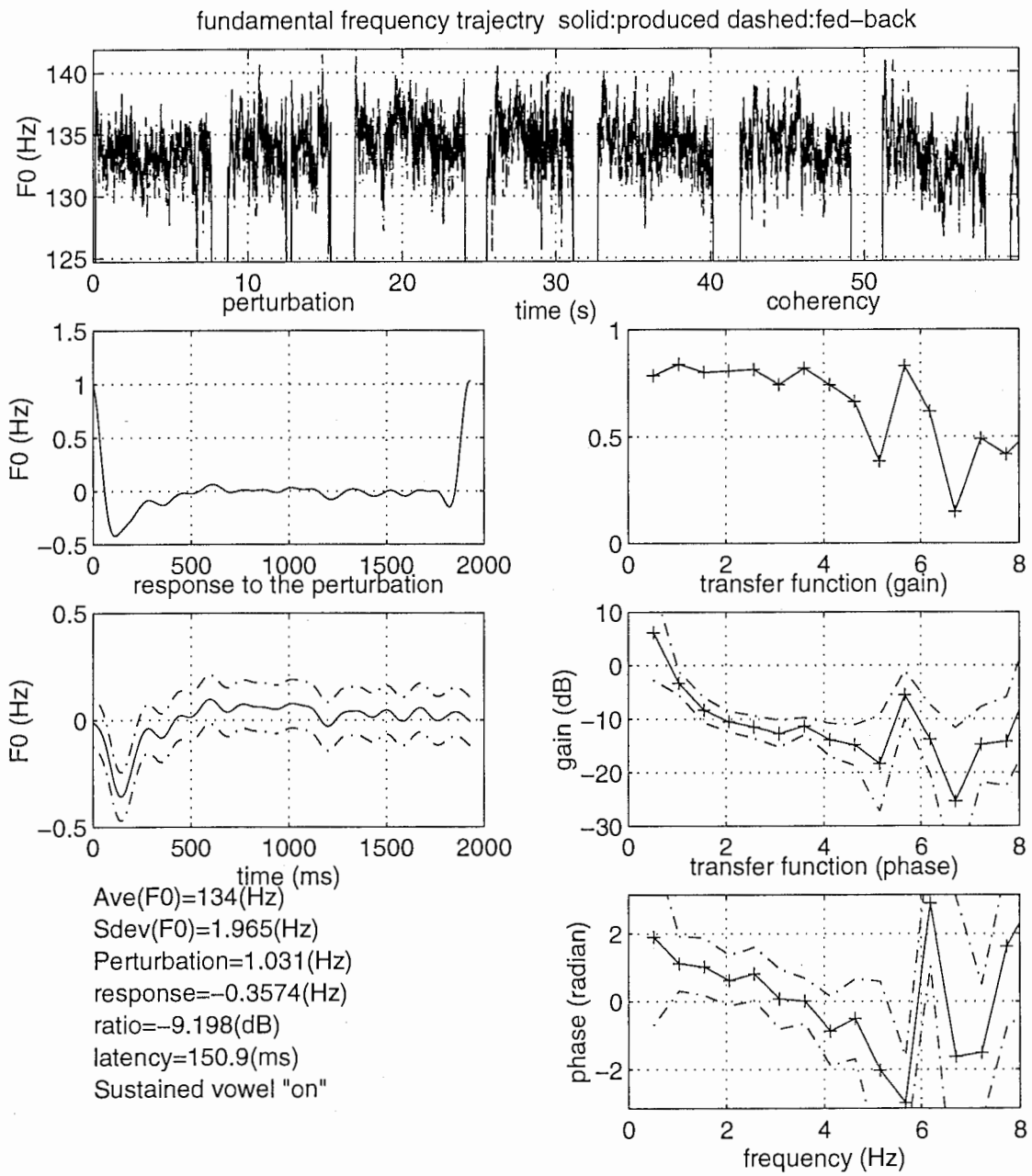
図13に分析例を示す。上段のグラフは話者の発声音とフィードバック音声の基本周波数を示す。左側2段目のグラフは被験者へのフィードバック音声の基本周波数と摂動に用いた疑似白色信号との相互相関を表す。これが知覚・生成システムへの入力となる摂動信号である。左側3段目のグラフは被験者が発声した音声の基本周波数と摂動に用いた疑似白色信号との相互相関を表す。これは線形システムと仮定した聴覚から音声生成への影響の閉ループ

でのインパルス応答の近似となっている。右側 2 段目以下は順に coherency、伝達特性の絶対値、伝達特性の位相を表す。図中の破線は、測定値の 90% 信頼区間を示す。伝達特性は先に述べた閉ループの近似インパルス応答と摂動信号のフーリエ変換の比として推定される。

図に示される応答は、従来の結果と同様に、明瞭な補償応答を示しており、摂動のピークと応答のピークの振幅の比として定義される応答の強さや、摂動のピークから応答のピークまでの時間間隔として定義される latency も、従来の測定結果の範囲に含まれている。

これまでの検討により、ピッチの摂動に対する応答は、2 個の 2 次系の応答の和に分解できることが知られている。ここで例として取り上げた応答も、図 14 のように比較的早い応答と遅い応答とに分解される。これらの応答の分解に当たっては、coherency の低下による誤差を避けるため、8Hz に帯域制限を行なった上で非線形最適化プログラムにより自乗誤差を最小にする評価基準の下で最適化を行なった。

図 14 の上の図の破線が実際に推定された開ループでのインパルス応答であり、実線が、2 つの 2 次系の応答から合成した応答である。下の図はそれぞれの 2 次系の応答を示したものである。これらの図の下には、それぞれの応答のパラメタを記した。早い応答には、入力から応答の開始までにむだ時間が認められる。図中では、この時間を fast delay として表わしている。fast delay は、音声波形からピッチ修正のための運動指令を作成するまでの情報処理に要する時間を表わすものと考えられる。



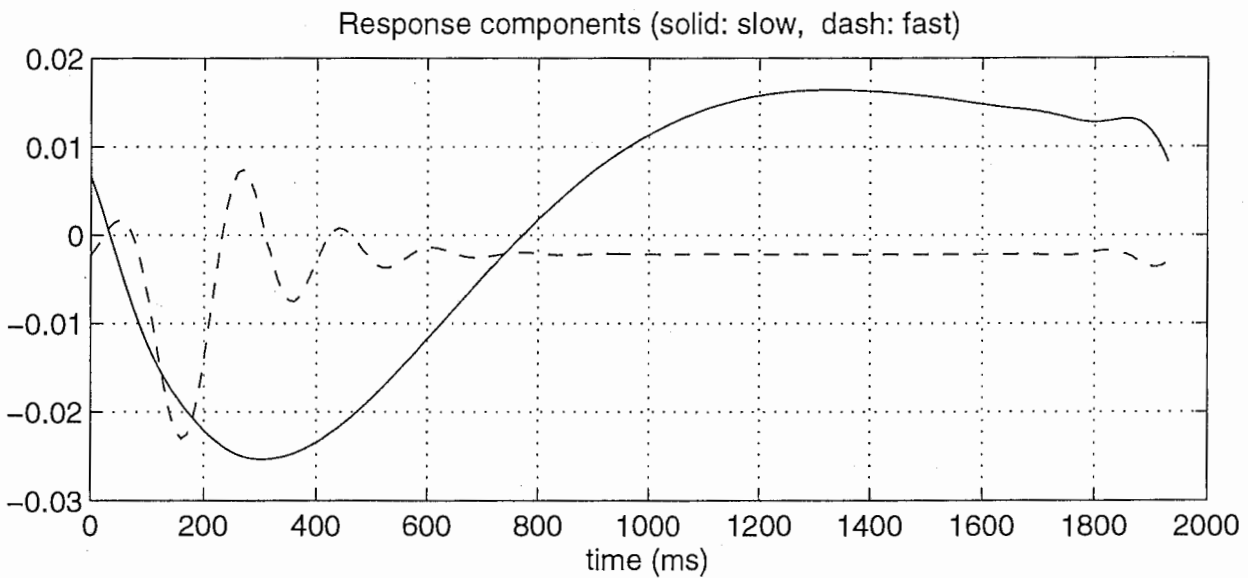
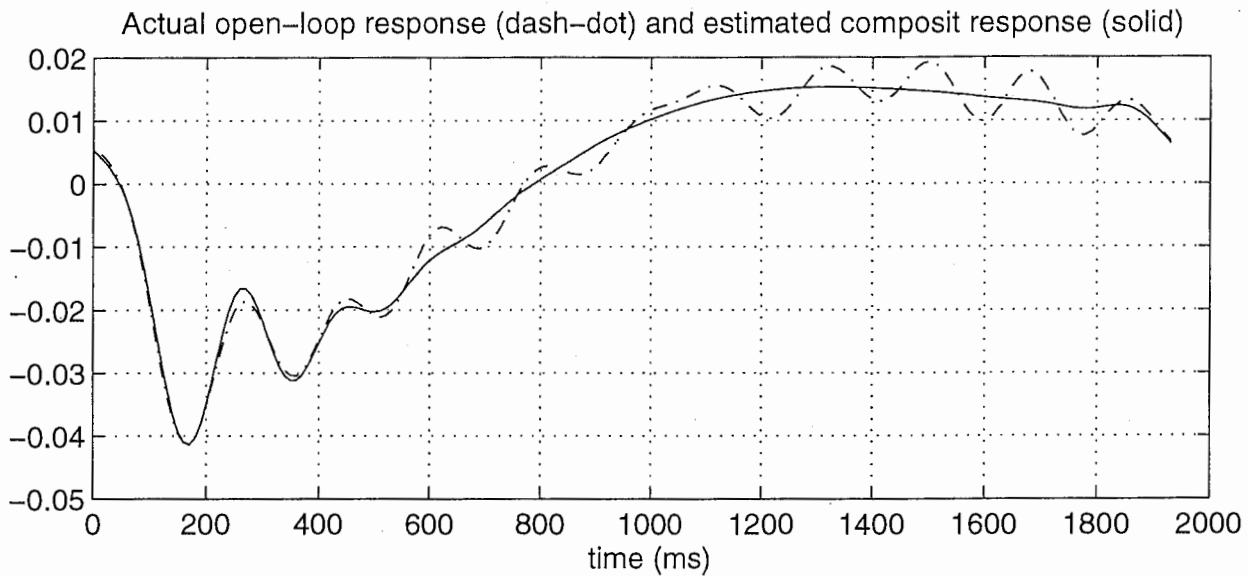
Path:/tmp_mnt/DB/speech-hakusan2/yamamoto/test2

File:testt Analysis date: 28-Sep-95 16:31:43

-rw-r--r-- 1 xyama dept1 1920441 Sep 4 14:27 testt.sd

analyzed by wholeprocedureSpl (1 of 1)

図 13: ピッチ変換聴覚フィードバックの分析例



/tmp_mnt/DB/speech-hakusan2/yamamoto/test2

Analyzed by decomposeCyc10 at 21:39:05-Oct-95 file name=testt

fastdelay=135ms df=0.22 fq=6.3Hz amp=0.1257 ppamp=0.03038

slowdelay=-22ms df=0.54 fq=0.58Hz amp=1.895 ppamp=0.04176

averageF0=134

resp highlimit=6Hz ffq highlimit=8Hz

図 14: ピッチ変換聴覚フィードバックの応答の要素への分解例

4.2 音量変換聴覚フィードバック

ピッチと並ぶ韻律の重要な要素である音量についても同様な知覚・生成間の相互作用が存在するか否かを検討するため、ピッチを対象としたTAFと同じ条件で、加える摂動として音量を変調した信号を用いた実験をおこなった。変調の大きさは、rms値の変動がp-p値で約4dBとなるものと、1.7dBとなるものの2種類を用いた。測定結果は、音量を音声波形の自乗平均値(rms)とし、dBで表わした。被験者は男女各2名。発声音の音名はG#、C、Eの3種類である。被験者は、母音「あ」を60秒間にわたって適当に息つぎをしながら発声するように指示された。

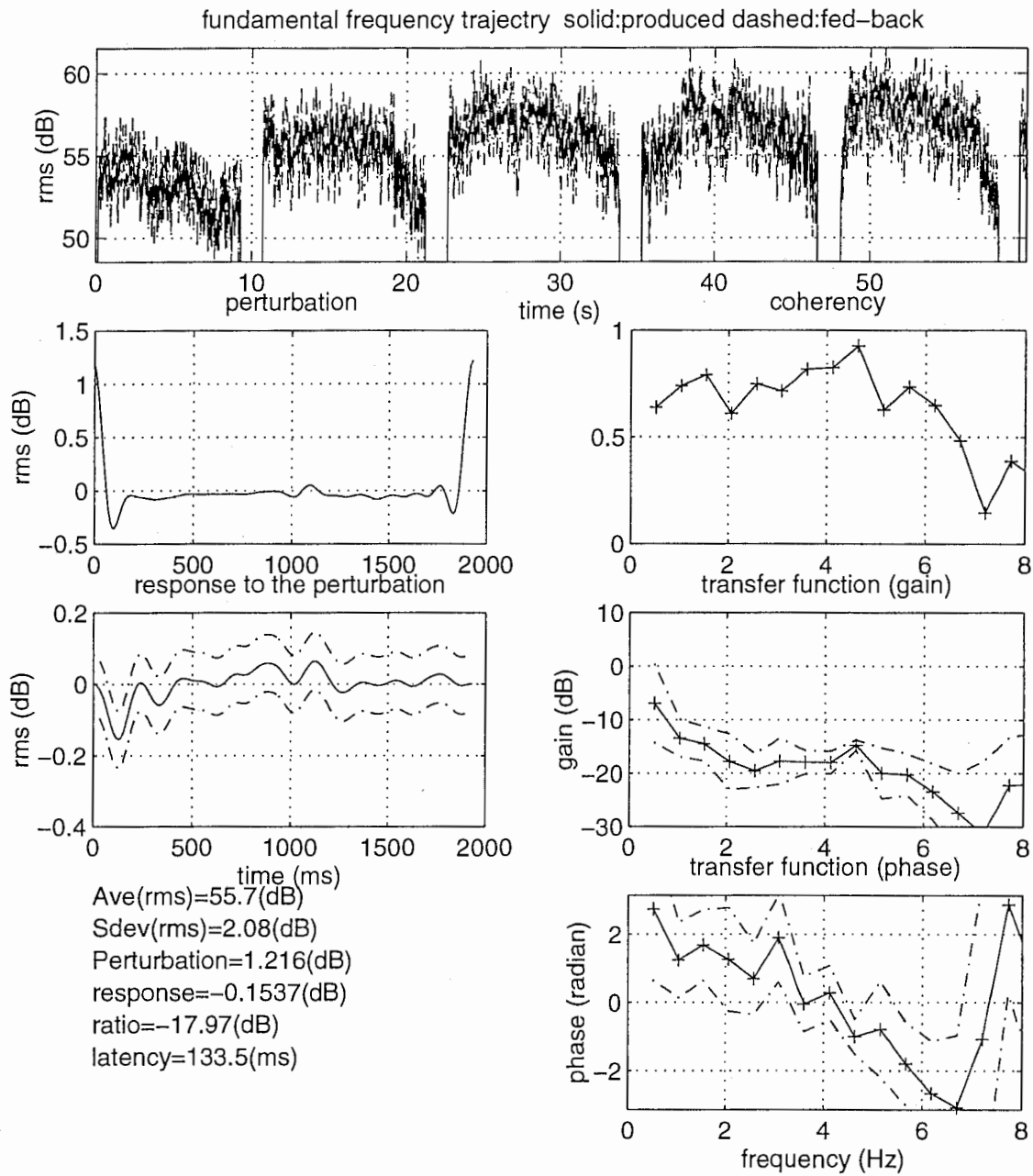
図15～図20に幾つかの代表的な分析例を示す。図15と図16はフィードバック音声の音量から発声された音声の音量への応答の例である。これらの応答は、ピッチからピッチへの応答の場合と同様に、摂動による変動を保証する方向の応答である。これらの応答波形と伝達特性は、coherencyが一般に低いことを除けば、ピッチからピッチへの応答に類似しており、同様に2次系への分解が可能である。図17に図15の応答を分解した例を示す。

音量から基本周波数への応答は、これまでに見てきたものと異なる様相を呈する。図19と図18に例を示す。なお、ここでは周波数をdBとして表わしている。音量の変動に対して、周波数の変動の大きさが1/10以下であることが、注目される。これらの例では、音量に対する周波数の応答は音量から音量への応答を微分したような波形を示している。これらの一例を図18に示す。21例の音量から周波数への応答の中で、微分波形と同じ符号であるものや、音量の応答と同様に負の応答（音量が上がると、周波数が下がる方向に反応する）を示すものは、14例であった。図19では、微分波形の符号が反転されたように見えるが、1件だけであり、測定時誤差と見たほうが良いであろう。以上の結果をまとめると、音量においてもピッチと同様な相互作用が存在するとともに、音量とピッチの間にも相互作用があることが確認されたと言えよう。

しかし、図20に示すように、フィードバック音声の音量から生成される音声の基本周波数、音量への応答がともにはっきりわからない例も6件存在しており、ピッチと比較すると、聴覚フィードバックの寄与は全体に小さいという印象がある。

4.2.1 パラメタ間相互作用

以上により、音量においてもピッチの場合と同様な知覚・生成相互作用が存在することについては確認されたといえよう。また、それらの応答が、おそらく副次的な作用としてピッチにも影響を及ぼしている場合のあることが示された。このような副次的な効果は、当然これまでのピッチを用いた変換聴覚フィードバックによっても観測されて良いはずである。実際、図21に示すように、ピッチから音量に関してもピッチからピッチへの応答と類似した応答が認められる。



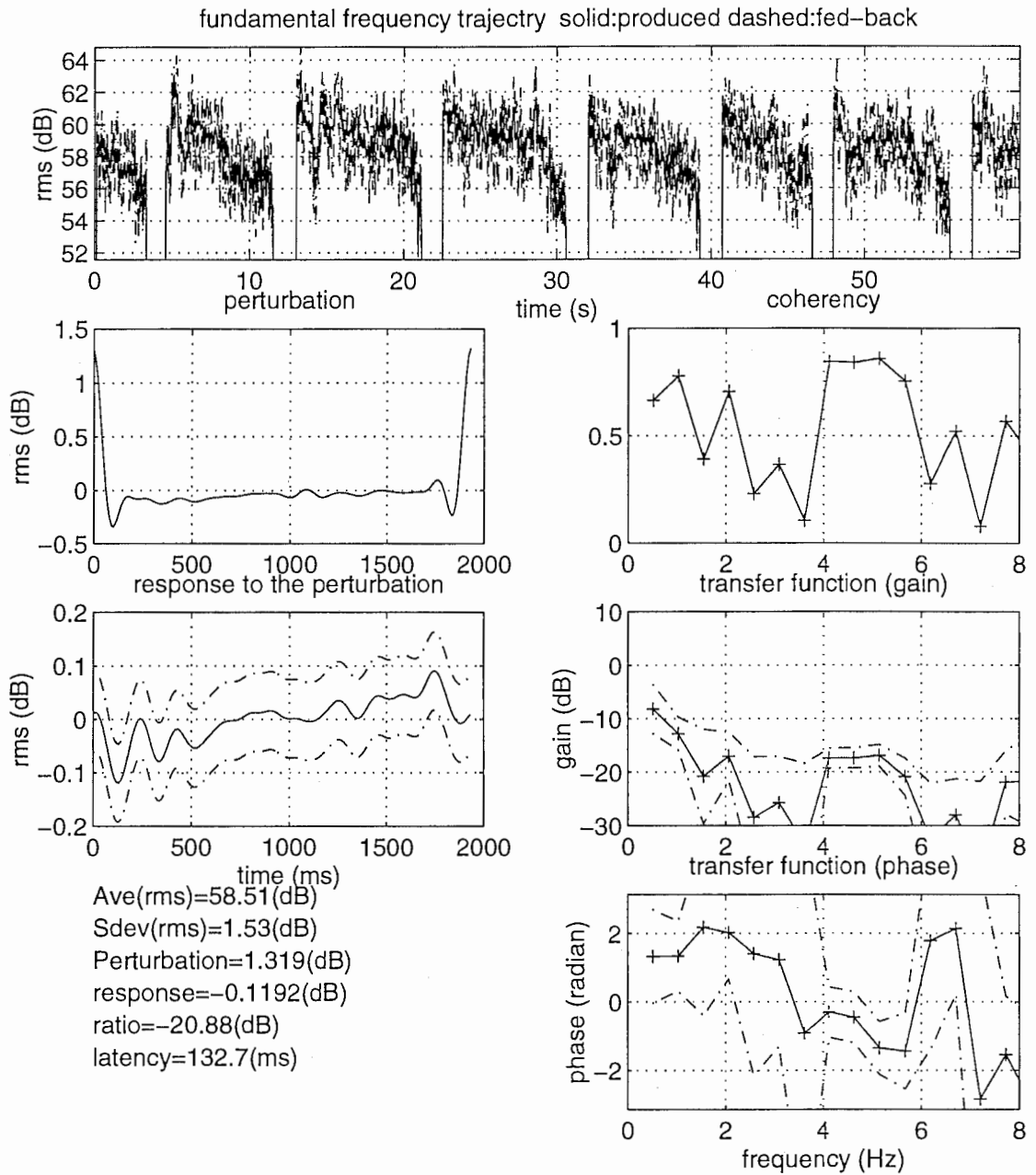
Path:/tmp_mnt/DB/speech-hakusan2/yamamoto/Taf_tanaka20aG#

File:testt Analysis date: 6-Oct-95 21:55:29

-rw-r--r-- 1 kawahara dept1 1920453 Sep 19 10:20 testt.sd

analyzed by wholeprocedureRms3L (1 of 1)

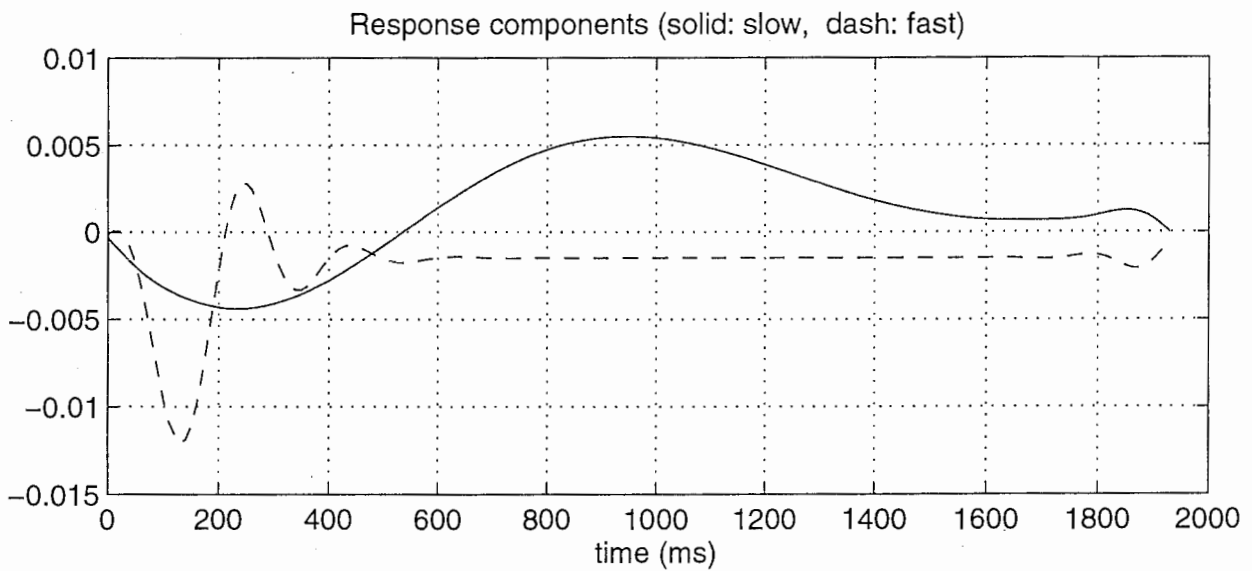
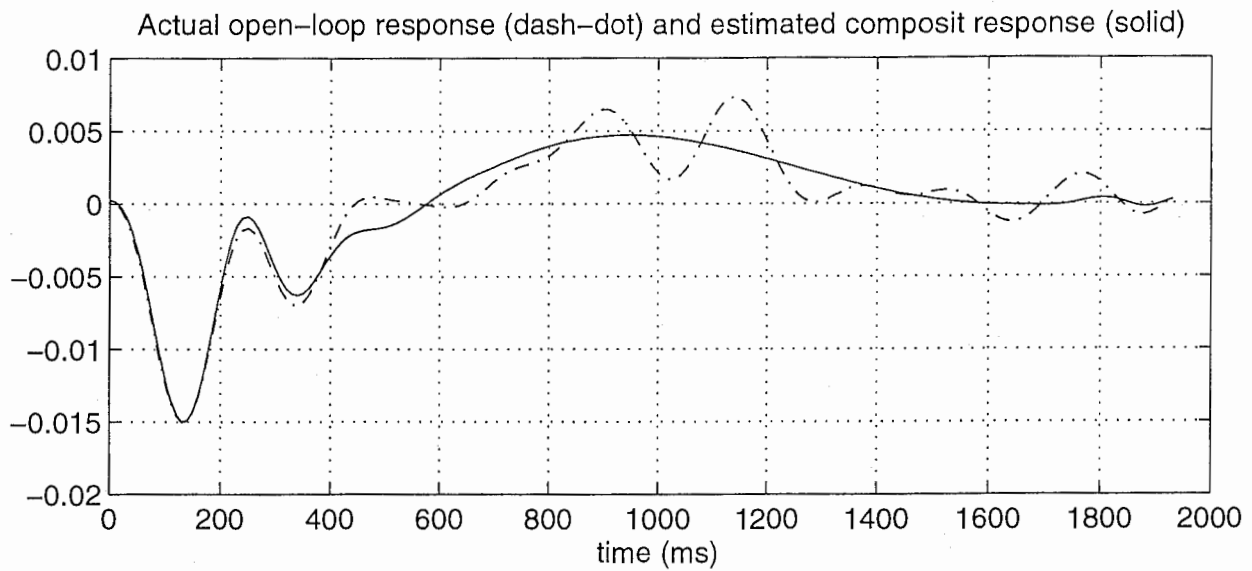
図 15: 生成される音声の音量への応答の例 (1)



Path:/tmp_mnt/DB/speech-hakusan2/yamamoto/Taf_tanaka20aC
 File:testt Analysis date: 6-Oct-95 22:2:25
 -rw-r--r-- 1 kawahara dept1 1920437 Sep 18 15:45 testt.sd

analyzed by wholeprocedureRms3L (1 of 1)

図 16: 生成される音声の音量への応答の例 (2)



/tmp_mnt/DB/speech-hakusan2/yamamoto/Taf_tanaka20aG#

Analyzed by decomposeCyc10 at 22:29:30 5-Oct-95 file name=testt

fastdelay=91.3ms df=0.35 fq=5.4Hz amp=0.07322 ppamp=0.01485

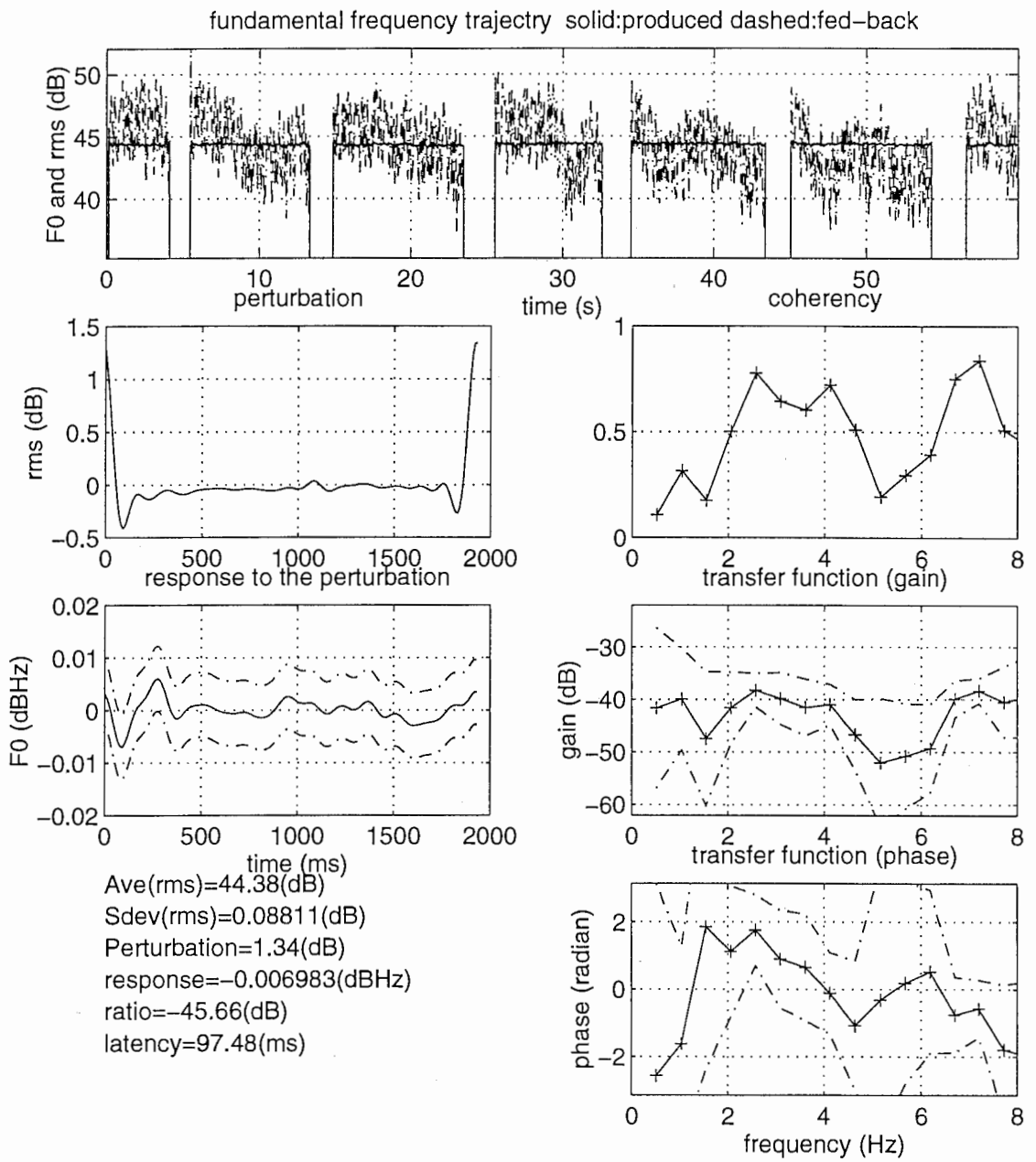
slowdelay=-50.1ms df=0.22 fq=0.72Hz amp=0.2149 ppamp=0.009895

averageF0=112

resp highlimit=6Hz ffq highlimit=8Hz

(1 of 1)

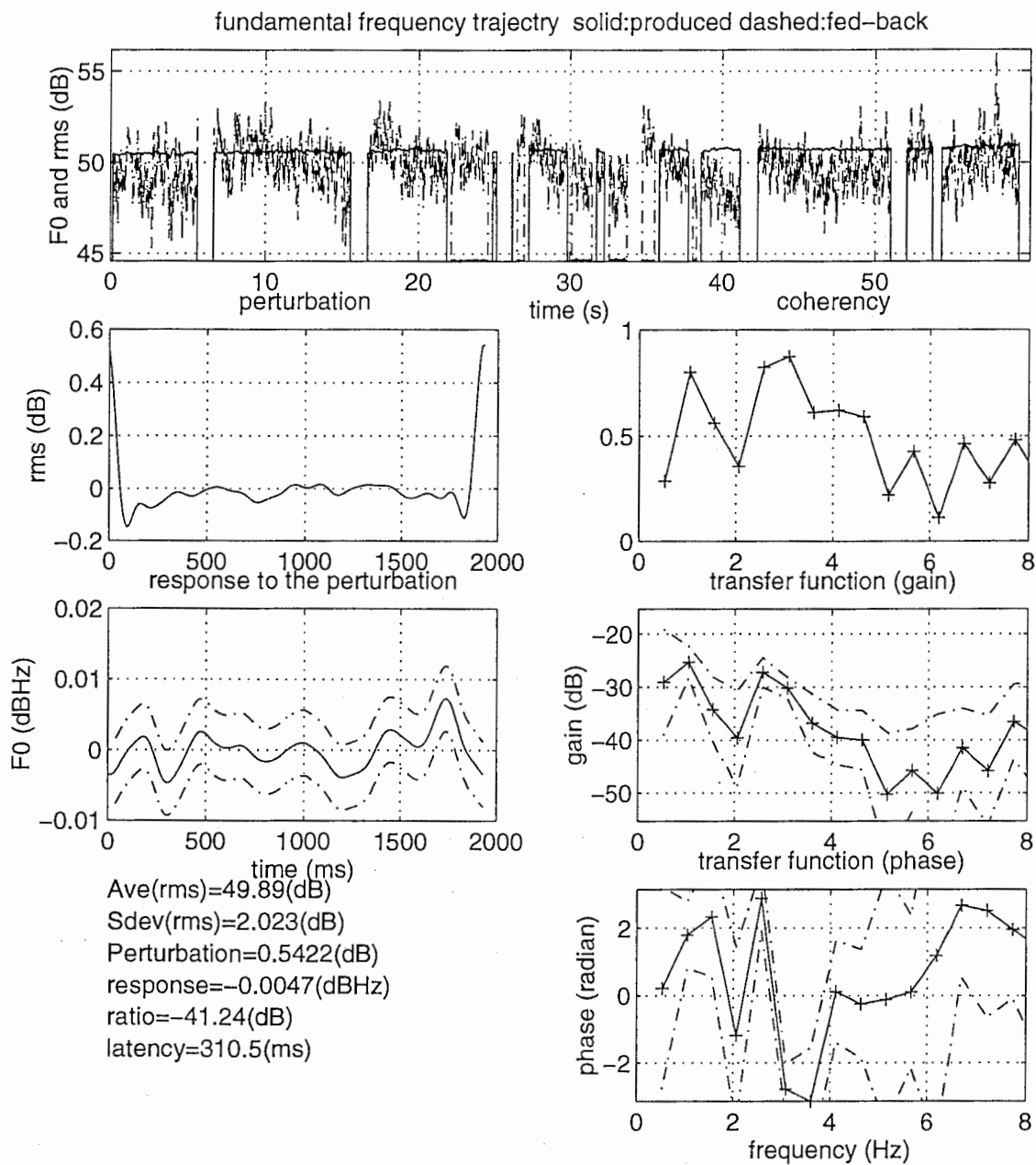
図 17: 音量から音量への応答の分解例 (1)



Path:/tmp_mnt/DB/speech-hakusan3/yamamoto/Taf_yamamoto20aE
 File:testt Analysis date: 6-Oct-95 22:21:10
 -rw-r--r-- 1 kawahara dept1 1920453 Sep 19 09:52 testt.sd

analyzed by wholeprocedureRms2L (1 of 1)

図 18: 音量から基本周波数への応答の例 (1)

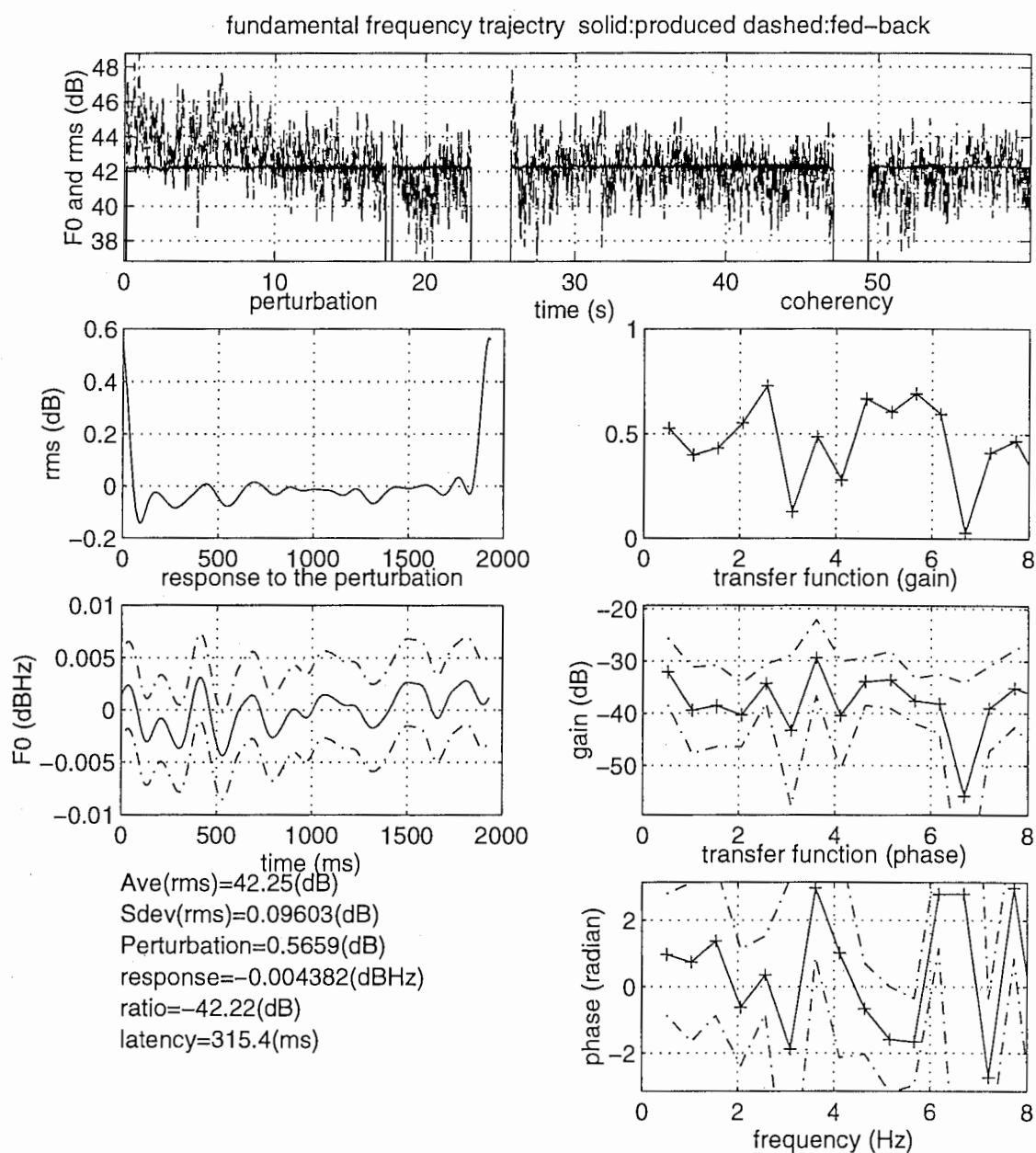


Path:/tmp_mnt/DB/speech-hakusan3/yamamoto/Taf_kubo20bE

File:testt Analysis date: 6-Oct-95 22:17:20

-rw-r--r-- 1 xyama dept1 1920449 Sep 19 14:58 testt.sd

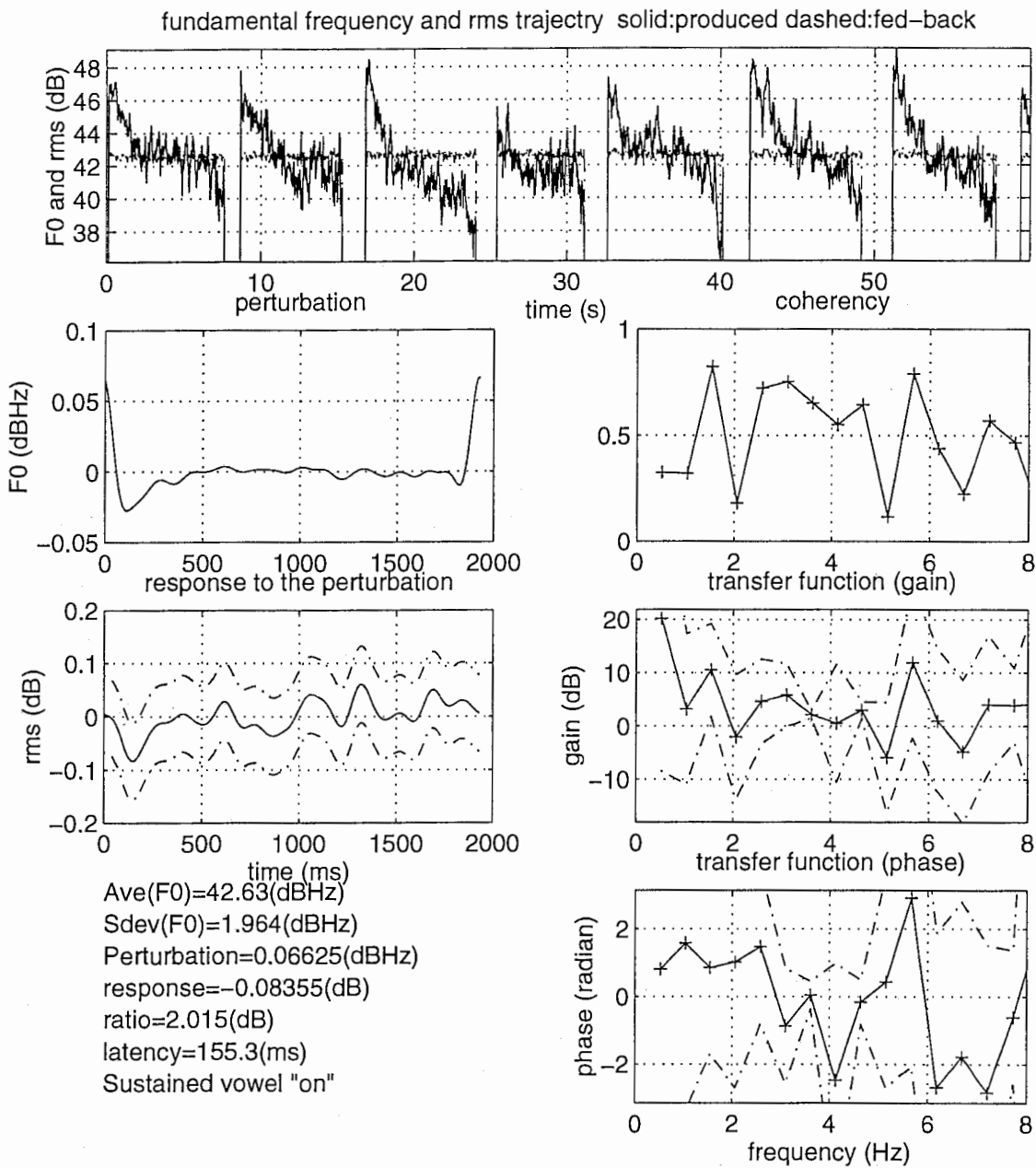
図 19: 音量から基本周波数への応答の例 (2)



Path:/tmp_mnt/DB/speech-hakusan2/yamamoto/Taf_kato20bC
 File:testt Analysis date: 6-Oct-95 22:28:27
 -rw-r--r-- 1 kawahara dept1 1920449 Sep 19 14:30 testt.sd

analyzed by wholeprocedureRms2L (1 of 1)

図 20: 生成される音声の基本周波数への応答がはっきりしない例



Path:/tmp_mnt/DB/speech-hakusan2/yamamoto/test2

File:testt Analysis date: 6-Oct-95 22:42:57

-rw-r--r-- 1 kawahara dept1 1920441 Sep 4 14:27 testt.sd

analyzed by wholeprocedureSpRmL (1 of 1)

図 21: ピッチ変換聴覚フィードバックにおけるピッチから音量への影響

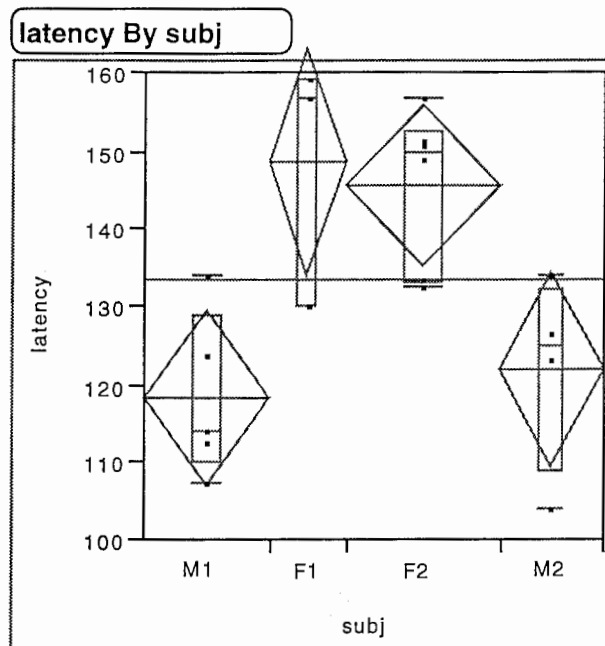


図 22: 被験者と最大応答までの時間の関係

5 実験結果の解析

以下の解析では、応答のパラメタとして、応答の強さ (ratio) と、最大応答の位置 (latency) を用いた。音量から音量へのフィードバックでは、2 個の 2 次系への分解が可能なので、それらに関するパラメタである、速い系の遅延 (fdly) と自然周波数 (ffq) についても解析を行なった。これらの実験において主要な要因は、被験者 (subj)、発声の音高 (note)、変調の大きさ (modlvl) であるが、予備的な検討の結果、音高による影響は男、女それぞれの中では有意にはならなかったため、以下では被験者と変調の大きさについての解析結果を示す。

5.1 音量-音量間相互作用

以下に示すように、分散分析の結果、latency と fdly という時間に関連するパラメタについてのみ、被験者および変調の強さの要因の影響が認められた。

5.1.1 応答の強さ (ratio)

被験者と変調の大きさを要因として、2 元配置の分散分析を行なった。その結果、被験者の要因 ($F(3,3) = 0.0040$, ($p = 0.9996$))、変調の強さの要因 ($F(1,1) = 0.9263$, ($p = 0.3585$))、交互作用 ($F(3,3) = 0.4434$, ($p = 0.7272$)) のいずれもが、有意とはならなかった。

5.1.2 最大応答までの時間 (latency)

被験者と変調の大きさを要因として、2 元配置の分散分析を行なった。その結果、被験者の要因 ($F(3,3) = 8.33$, ($p = 0.0045$)) と変調の強さの要因 ($F(1,1) = 7.61$, ($p = 0.0202$)) とが 5% 水準で有意となり、交互作用 ($F(3,3) = 0.065$, ($p = 0.977$)) は、有意とはならなかった。

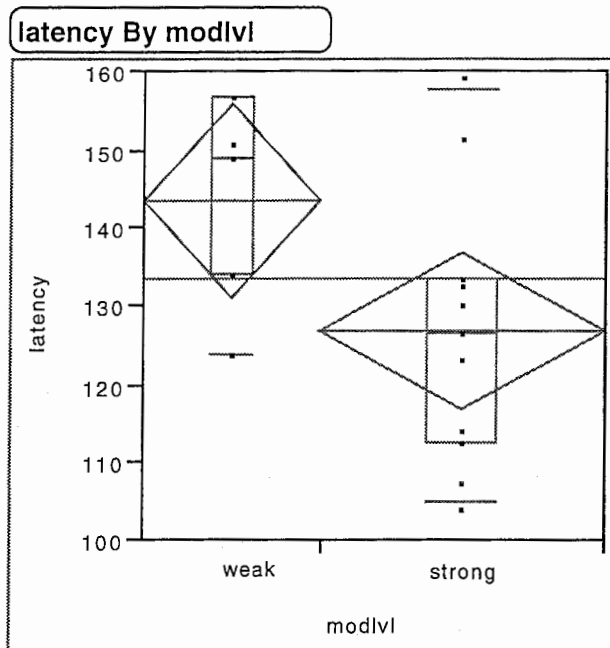


図 23: 変調の大きさと最大応答までの時間の関係

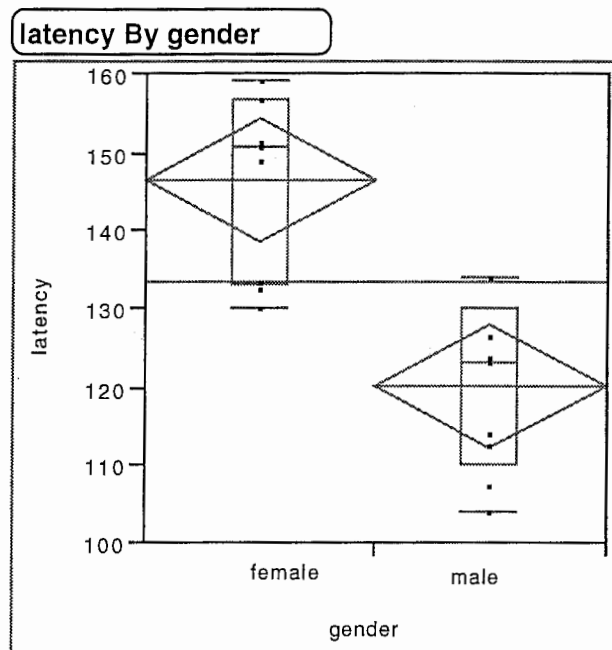


図 24: 性別と最大応答までの時間の関係

図 22に被験者と latency との関係を示す。図 22は、ダイヤモンドによって平均値と 95%信頼幅を、矩形によって、10%, 25%, 50%, 75%, 90% の分位値を表わす。図中の M1, M2 は、男性被験者を、F1, F2 は、女性被験者をそれぞれ表わしている。

図 23に変調の大きさと latency との関係を示す。図 23は、ダイヤモンドによって平均値と 95%信頼幅を、矩形によって、10%, 25%, 50%, 75%, 90% の分位値を表わす。

ここで、被験者による差を性別を要因とし変調の強さを他の要因として再度分散分析を行なうと、被験者の要因 ($F(1,1) = 30.20$, ($p = 0.0001$)) と変調の強さの要因 ($F(1,1) = 9.74$, ($p = 0.0075$)) とが 5%水準で有意となり、交互作用 ($F(1,1) = 0.205$, ($p = 0.657$))

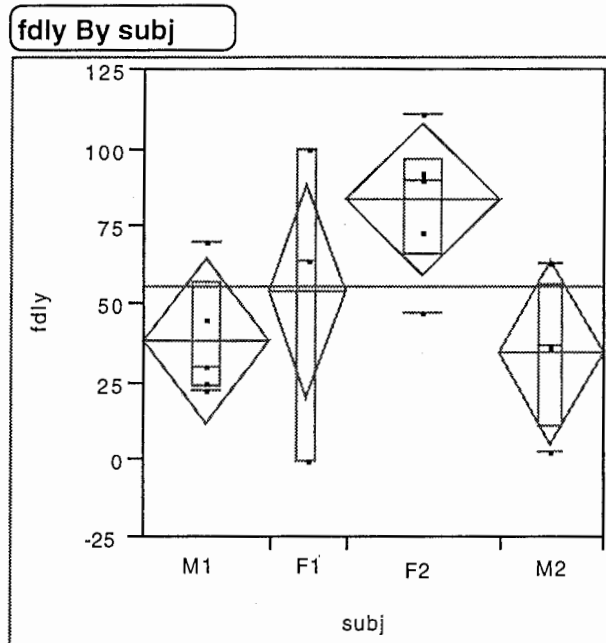


図 25: 被験者と速い成分のむだ時間の関係

は、有意とはならなかった。図 24に性別と latency との関係を示す。図 24は、ダイヤモンドによって平均値と 95%信頼幅を、矩形によって、10%, 25%, 50%, 75%, 90% の分位値を表わす。

5.1.3 速い成分のむだ時間 (fdly)

被験者と変調の大きさを要因として、2元配置の分散分析を行なった。その結果、被験者の要因 ($F(3,3) = 4.54, (p = 0.0295)$) が 5%水準で有意となり、変調の強さの要因 ($F(1,1) = 1.98, (p = 0.1889)$) と交互作用 ($F(3,3) = 2.51, (p = 0.117)$) とは、有意とはならなかった。

図 25に被験者と fdly との関係を示す。図 25は、ダイヤモンドによって平均値と 95%信頼幅を、矩形によって、10%, 25%, 50%, 75%, 90% の分位値を表わす。ここでも、同様に女性の被験者の方が、むだ時間が大きい傾向が認められる。

5.1.4 速い成分の自然周波数 (fdly)

被験者と変調の大きさを要因として、2元配置の分散分析を行なった。その結果、被験者の要因 ($F(3,3) = 1.148, (p = 0.3767)$)、変調の強さの要因 ($F(1,1) = 0.941, (p = 0.3547)$)、交互作用 ($F(3,3) = 0.0411, (p = 0.9882)$) のいずれもが、有意とはならなかった。

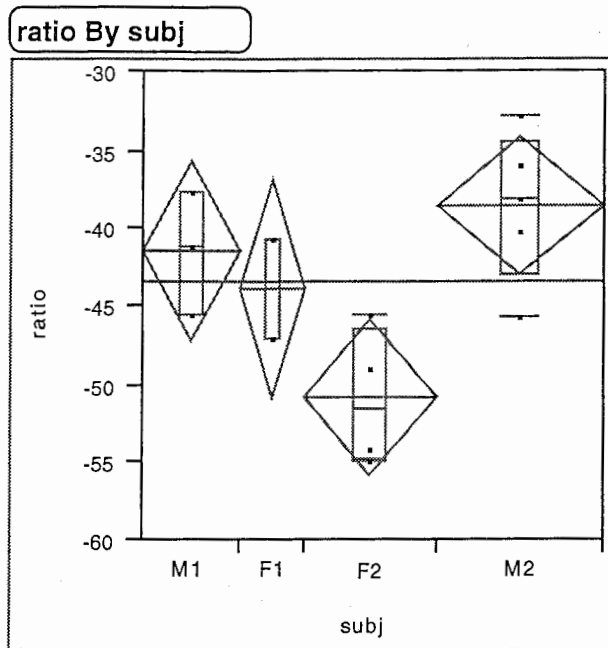


図 26: 被験者と応答の強さの関係

5.2 音量-周波数間相互作用

5.2.1 応答の強さ (ratio)

被験者と変調の大きさを要因として、2元配置の分散分析を行なった。その結果、被験者の要因 ($F(3,3) = 4.264$, ($p = 0.0621$)) と変調の強さの要因 ($F(1,1) = 5.0281$, ($p = 0.0661$)) は 10% の水準で有意となったが、交互作用 ($F(3,3) = 0.141$, ($p = 0.9318$)) は有意とはならなかった。

図 26 に被験者と応答の強さとの関係を示す。図 26 は、ダイヤモンドによって平均値と 95% 信頼幅を、矩形によって、10%, 25%, 50%, 75%, 90% の分位値を表わす。

図 27 に変調の強さと応答の強さとの関係を示す。図 27 は、ダイヤモンドによって平均値と 95% 信頼幅を、矩形によって、10%, 25%, 50%, 75%, 90% の分位値を表わす。変調が強くなると相対的な応答は弱くなっている。これは、音量への摂動に対する周波数の応答が、絶対値で見ると、両方とも同程度の大きさであることを意味している。

5.2.2 最大応答までの時間 (latency)

被験者と変調の大きさを要因として、2元配置の分散分析を行なった。その結果、被験者の要因 ($F(3,3) = 10.4484$, ($p = 0.0085$)) は 5% の水準で有意となったが、変調の強さの要因 ($F(1,1) = 0.2025$, ($p = 0.6685$)) と交互作用 ($F(3,3) = 1.062$, ($p = 0.432$)) とは有意とはならなかった。

図 28 に被験者と fdly との関係を示す。図 28 は、ダイヤモンドによって平均値と 95% 信頼幅を、矩形によって、10%, 25%, 50%, 75%, 90% の分位値を表わす。

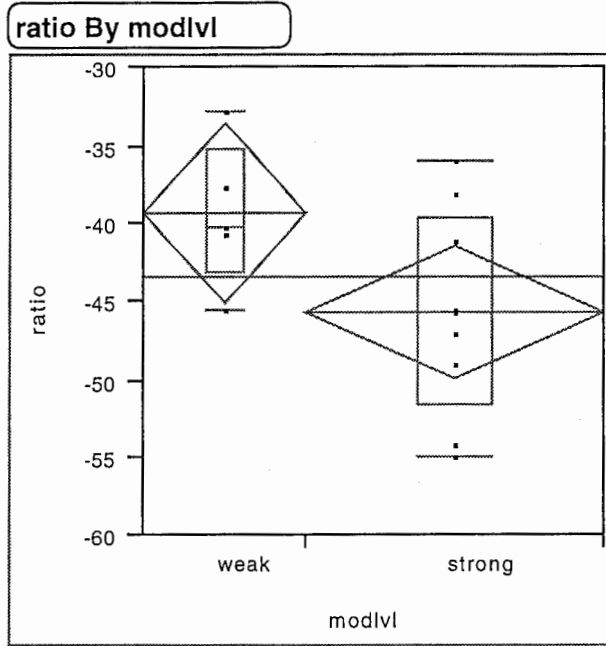


図 27: 変調の強さと応答の強さの関係

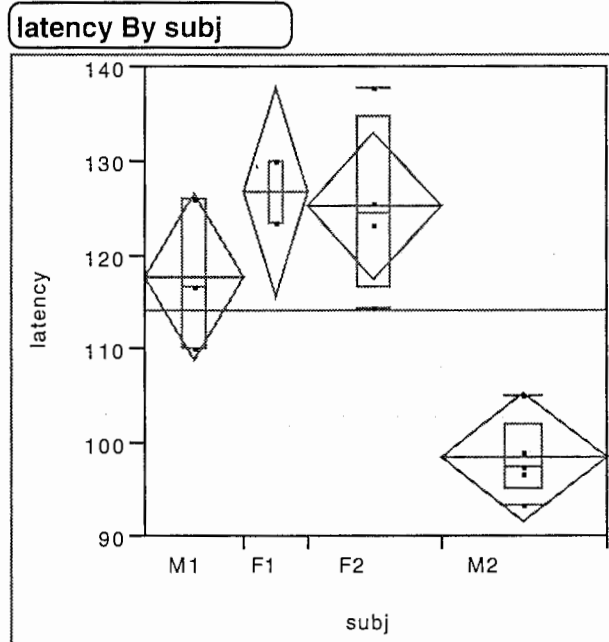


図 28: 被験者と最大応答までの時間の関係

6 課題

当初の目的の一つであった音場の測定と、音場の空間的変動に対する人間の応答特性の実験系の構築や予備実験には至らなかった。今後の課題としたい。

M 系列を用いた分析では、非線形性による応答が全区間にわたって雑音状に分布する問題がある。変換聴覚フィードバックで測定しようとする人間の知覚生成系は、見方によっては非線形性のかたまりである。何らかの工夫により [3]、これらの非線形特性を評価することが必要である。

また、周波数については、変調は対数周波数軸上で線形に掛けられていた。この場合、知覚的にも線形の変調がかけられていたことになる。それに対し、音量の変調は、線形の軸上で線形に掛けられており、知覚的には、非線形の変調が掛かっていたことになる。特に p-p 値で 4dB の変調では、この影響が M 系列の問題と併わさって、誤差を大きくしていた可能性がある。これらの影響を評価するとともに、知覚的に線形な音量の変調を用いた検討が必要である。

これまで用いてきた M 系列は、標本化周波数が 16Hz であった。ナイキスト周波数は 8Hz となるが、これは、発声系におけるピークの周波数と一致する場合があります。疑似白色信号から MIDI 制御信号を作成する場合に信号の補間と平滑化を行うのであれば、時間長と周期を独立に決定することが出来るので、より高い次数の M 系列を用いて、標本化周波数を測定対象の周波数範囲から追い出すことも可能である。これらについては、今後の課題として検討を進める必要がある。

7 まとめ

本実習では、これまで音声の基本周波数に対する摂動を用いた変換聴覚フィードバックによって発見された相互作用が、音量や音場のようなより広範な人間行動の指標に対しても測定可能であるかを確認するための予備実験としてのデータ収集、および測定手法の確立のための検討を行なった。具体的には、まず最初に音場における外部音原からの聴覚情報による影響測定の基礎データとして実験機器の音響特性と音場での特性データの収集を行なった。この測定結果は、1993 年の測定結果を大局的には追認するものであるが、高分解能の測定を実施したため、時間的応答、周波数特性の細部の特徴等を明らかにすることができた。

次いで、基準条件としてピッチを対象とした変換聴覚フィードバックの基礎データを収集した。最後に同一の被験者群に対して音量を対象として変換聴覚フィードバック実験を遂行し、音量に関しても、基本周波数の場合と同様な相互作用が存在することを発見した。また、同時に、これらの音量への応答が音声の基本周波数に対しても影響を与えることを見出した。これらの発見は、発声過程の多次元的動作の総合的解明につながるものである。

最後に、今後検討すべき様々な課題について論じた。

8 謝辞

本実習は ATR 人間情報通信研究所第一研究室で行われたものである。研究をすすめるにあたり、また種々の実験を行うにあたり被験者として協力して頂いた方々、及び常日頃、適切な御助言を頂いた第一研究室の方々に感謝いたします。

参考文献

- [1] 岩谷、河原：“変換聴覚フィードバックの基礎検討”、ATR Technical Report、TR-H-009 (1993.2).
- [2] 金田：“インパルス応答の測定における誤差の検討”、音響学会講演論文集、2-3-17、pp.567-568 (1995.3).
- [3] 柏木、孫：“非線形システムのボルテラ核の同定”、計測自動制御学会論文集、31、8、pp.1054-1060 (1995).
- [4] 加藤、河原：“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”、ATR Technical Report、TR-H-166 (1995.10).
- [5] 河原：“音声知覚・生成相互作用の伝達特性について”、音響学会聴覚研究会資料、H-95-35 (1995).
- [6] 河原：“フィードバック音声の基本周波数への摂動に対する周波数成分の分解について”、信学技報、SP-95-105 (1995)
- [7] 河原、加藤、ウィリアムス：“音声の基本周波数の揺らぎに対する聴覚フィードバックの寄与について”、音響学会講演論文集、2-3-18、pp.451-452 (1995.10).
- [8] 鈴木、浅野、曾根：“音響系の伝達関数の模擬をめぐって（その2）”、日本音響学会誌、45、pp. (1989)

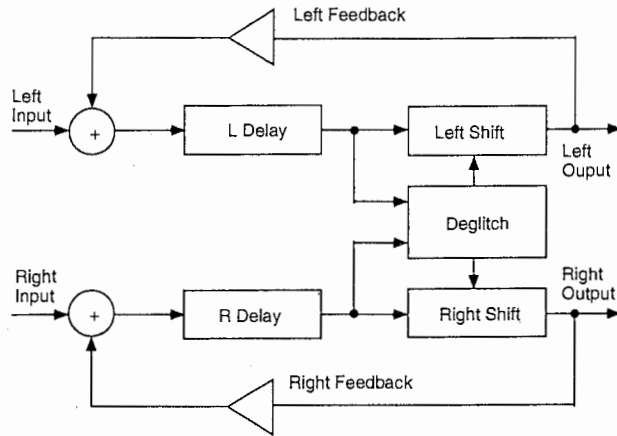


図 29: 周波数の変調に用いた Algorithm 102

A Harmonizer のプログラム

Harmonizer のプログラムは、内蔵されている既存のプログラムを変更して作成した。それぞれのプログラムは、独自のブロック構成を持っている。これらのプログラムは、Algorithm と呼ばれる。

A.1 周波数の変調

周波数の変調は、図 29 に示すように Algorithm 102 (Dual Shift) から作成された。L Delay を 0ms とし、Left Feedback を 0 として、Left Shift が外部の MIDI 信号により変調されるように設定した。

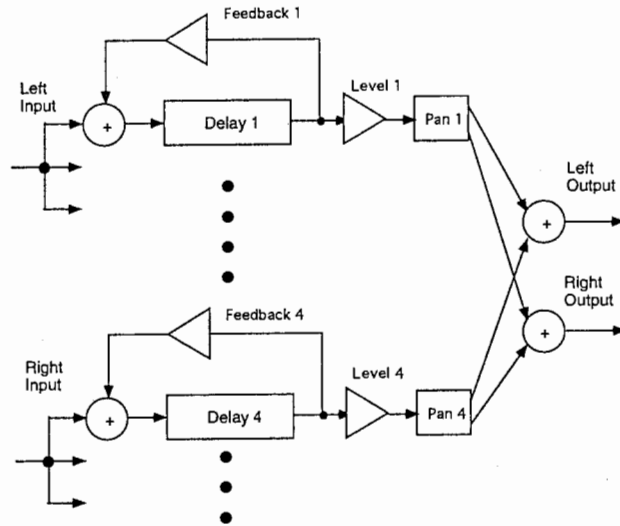


図 30: 音量の変調に用いた Algorithm 105

A.2 音量の変調

周波数の変調は、図 30に示すように Algorithm 105 (Swept Combs) から作成された。Delay, Feedback, Level はすべて 0 として R channel に Pan しておき、加工された信号と未加工の入力信号の混合比率を表わすパラメタである Mix が外部の MIDI 信号により変調されるように設定した。加工信号のレベルが 0 であるので、Mix を増加させると、未加工の入力信号の比率が減少し、結果として信号のレベルが減衰する。このようなプログラムは直観的ではなくバグの混在の原因となるが、Harmonizer を使用して MIDI で音量を制御するには、この方法以外は無いようである。