2003－MUS－51　(12)
2003／8／5

会話音声と歌唱音声の基本周波数制御の動特性について

矢永龍一郎†　　河原　英紀†,††

E-mail: †{s035055,kawahara}@sys.wakayama-u.ac.jp

あらまし　同一の発声器官を共用していても、歌声と会話音声の物理特性には様々な違いが存在する。それらの中でも基本周波数の違いは、最も大きく、歌唱の本質的特徴でもある。ここでは著者の一人によって開発された実験手法（変換聴覚フィードバック）を用いて 6 名の被験者の会話音声と様々な演奏法による歌唱音声の動特性を測定した結果について報告する。これらの測定結果は、会話音声と歌唱音声の間には、歌唱音声における演奏法による動特性の違いを超える違いがあるものの、それらは制御情報の違いによるものであることを示唆する。

F0 control dynamic parameters in speech and singing

Test results using Transformed Auditory Feedback with a revised procedure

Ryuichiro YANAGA† and Hideki KAWAHARA†,††

E-mail: †{s035055,kawahara}@sys.wakayama-u.ac.jp

Abstract　While speech and singing are sharing the same speaking organ, their physical characteristics show numerous differences. Among those parameters, funcamental frequency (F0) is essential for singing and has the largest differences between two modes of voicing. This article reports dynamic parameters of F0 control in speech and singing using an experimental procedure TAF (Transformed Auditory Feedback), which was developed by one of the authors. Six (three male and three female) subjects were participated in this experiment and sang in several musical expressions. The results suggest that the difference between speech and singing is resulted mainly from the difference of their control commands and shares the same dynamic parameters, eventhough the difference is larger than differences between various musical expressions.

1. はじめに

会話と歌唱は、同じ発声器官を用いているにも関わらず、様々に異なる物理特性を有している。特に基本周波数は、旋律を成立させる本質的な要素であり、会話と歌唱において最も大きく異なる特性でもある。自然で豊かな表現能力を持つ歌唱システムの構築には、楽譜上の音符と表情記号や歌い手の特徴を表す動的なパラメタから基本周波数の軌跡を計算する方法を明らかにすることが必要となる。この動的なパラメタが、会話音声から求め得るものであるか、様々な演奏表現毎に求める必要があるものであるかは、システム構築の難易度を左右する基本的な問題である。

これまで、歌唱の基本周波数周波数制御モデルの構築に向けて [14]、音程の遷移の前後に認められるプリシュートやオーバーシュート [4], [5], [7]、微小変動、ヴィブラート等の特性や知覚的影響が調べられている。しかし、これらの研究は、様々な歌唱における特徴が発声器官の動特性の変化によるものか発声器官への制御指令の違いによるものなのかを明らかにするものではなかった。本研究では、音声の基本周波数制御における聴覚フィードバックの影響を利用してこれらの動的パラメタを直接求め、会話と歌唱における基本周波数軌跡の差異がどのようにして生じているかについて調べた結果について報告する。

2. 背　　景

著者の一人は、音声生成における聴覚系の影響を調べるための一般的方法として変換聴覚フィードバック (TAF: Transformed Auditory Feedback) を提案し、基本周波数の制御特性を調べて来た [11]。変換聴覚フィードバックでは、聴覚フィードバック経路に加える操作の量をできるだけ少なくすることで、正常な生成知覚相互作用を妨害せずに特性を測定することを狙っている。そのため、予測が困難でありかつ同一振幅で最大の S/N 比を確保することのできる M 系列信号を利用し、周波数の操作も 25cent 程度の微小なものとしている。TAF は、同期加算を組織的に行う方法であるとも解釈することができるため、高い S/N が必要となる伝達特性の測定も可能となっている。

TAF を用いた一連の研究により、周波数の操作に対して、逆方向の早い応答と遅い応答が存在すること、早い応答の潜時は、基本周波数に反比例する成分を含むこと、周波数の操作の効果

著者所属：† 和歌山大学, Wakayama University
††ATR 人間情報科学研究所, ATR Human Information Science labs.

に対する補償は、遅い成分が主に担っており、早い成分は補償の開始を早める効果があること、4Hz 付近の基本周波数の揺らぎは、聴覚フィードバックの効果が 4Hz 付近において正帰還となることによるものであること等が報告されている [11]〜[13]。しかし、これらの結果は、後で触れる測定手法に内在する問題もあり、学術論文としては刊行されていない。

発声の基本周波数の制御における聴覚フィードバックの関与は、Northwestern 大学の Larson らのグループによって、ステップ状の周波数シフトを用いても調べられている [8]〜[10]。この場合、ステップ状の周波数シフトは数 100cent に及び、被験者に容易に知覚される大きさのものが用いられている。Larson らの一連の研究により、ほとんどの応答は周波数シフトを補償する方向の応答であること、応答には、早い成分と遅い成分が含まれているらしいこと、早い応答成分は、音程の遷移する近くでも音程が一定の定常部でも同様に認められることが報告されている。

著者らによる結果と、Larson らによる結果は、ほぼ同じ傾向を示しているものの、幾つかの差異が残されている。まず、Larson らの実験では、周波数シフト量が多いときにはシフトと同方向に基本周波数が変化する（暴走する）事例が生ずることがあることが報告されている。これは、周波数シフトがおこなわれたことに気付いた被験者の制御が、混乱により破綻したことを示唆する。Larson らの報告にある、シフトの方向に対する応答の非対称性、シフトの大きさに対する応答の非線形性は、TAF では調べられない。いっぽう、Larson らの方法では、操作回数が TAF と比較して圧倒的に少ないため、伝達関数のような定量的な解析が困難であるという問題がある。ただし、TAF においても、遅い応答のパラメタを正しく推定するためには、これまでの実験で用いていた約 2 秒という M 系列の周期をより長くする必要がある。

もう一つ検討しておかなければならないことは、F0 の解析と音声合成での F0 制御に広く用いられている藤崎モデル [2] との関連である。このモデルは、喉頭の制御機構にまで踏み込んだ考察を背景として構築されており、F0 軌跡を簡単な低次のシステムの応答として即物的に実現する方法とは一線を画している。しかし、著者や Larson らの研究で見いだされた応答は、現在の藤崎モデルが用いている臨界制動よりも遥かに振動的である点において異なっている。人間における実際の F0 の制御がどのような動特性を有しているかを明らかにすることは、複数のモデルが必要か共通のモデルのパラメタ調整で済むのかという問いへの答を与えることにつながる。

歌唱システム [1] における基本周波数制御の適切なモデルを作成するためには、まず、両者の報告する「早いシステム」と「遅いシステム」の挙動とパラメタを明らかにすることが必要となる。そのためには、まず、Larson らの研究で指摘された応答の非線形性（非対称性も非線形性の一種として扱う）を考慮した解析法が必要となるとともに、測定に用いる M 系列の周期を、測定対象に含まれ得る応答よりも長いものとすることが必要となる。ここでは、前報で提案した新しい解析法を利用することにより、非線形性と低 S/N に起因する問題点を回避するとともに、後者の問題を回避するために約 3 秒の周期を持つ M 系列を用いることとした。

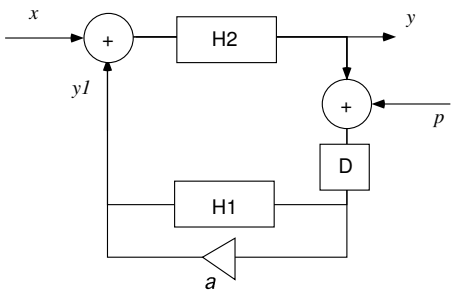

図 1　Hypothetical model of voice F0 control and probing points of TAF procedure

3.　変換聴覚フィードバック実験

これまでの知見に基づき、図 1 を基本周波数の制御における聴覚系の関与を表す作業仮説とした。ここで H1 は、ピッチ知覚のシステムを表し、H2 は、神経指令に基づいて動作する発声器官の機械的応答特性を表す。図中の a は、聴覚を介した基本周波数の自動制御機構のゲインを表す。x は、基本周波数の目標指令、$y1$ は、フィードバック情報、y は、発声された音声の基本周波数、p は、実験系による基本周波数の操作量を表す。これら全ての量を対数周波数で表すことにより、システムは第一近似として線形となる。随意運動の計算理論 [18] が示唆するように基本周波数の制御が習熟したスキルであるなら、このフィードバックゲインは、かなり小さな値となると考えられる。今回の実験の目的は、これらのサブシステムのパラメタに、歌唱と会話音声とでどのような共通性と相違点があるかを明らかにすることにある。

3.1　実験系の構成

実験系の構成を図 2 に示す。マイク (SONY C-536P) で収録された声は、ピッチ変換器 (EVENTIDE H3500S harmonizer) によって、1 秒間に 20 回の割合で長さ 63 の M 系列から作成された疑似白色信号に基づく変調が加えられた。変調の深さは、S/N と被験者への影響の妥協点から 25cent が用いられた。こうして変調された被験者の音声は、空気を伝わってくる自分の声をマスクするためのピンクノイズと混合されて、密閉型ヘッドフォン (SONY RH-40M) を介して発声者にフィードバックされた。

変調用の疑似白色信号は Digital Performer(Mark of the Union) により、PC から MIDI 信号として Harmonizer に送られた。ヘッドホンから再生される音声の音圧レベルは、自然側音がマスクされるように高く設定された。実際に実験中に観測された音圧は、A 特性で 83〜85dB であった。このフィードバック音声による聴覚障害が生ずることを予防するため、音圧レベルを、HATS (B & K Type 4128) の人口耳により常にモニターするとともに、90dB を越えると [17]、リミッター (BEHRINGER　MDX2200) が作動して過大音圧が被験者に加わらないようにした。発声した音声と変調された音声は、ミキサー (TASCAM M-08) を通して 48kHz、16bit で DAT(SONY ZA5ES) に同時記録された。

3.2　実 験 方 法

被験者は、合唱及び声楽経験のある音楽専攻の大学生 (18

なお、以下の解析では操作量が 25cent という非常に小さなものであり、対数関数は一次関数で近似できるため周波数そのものを用いている。

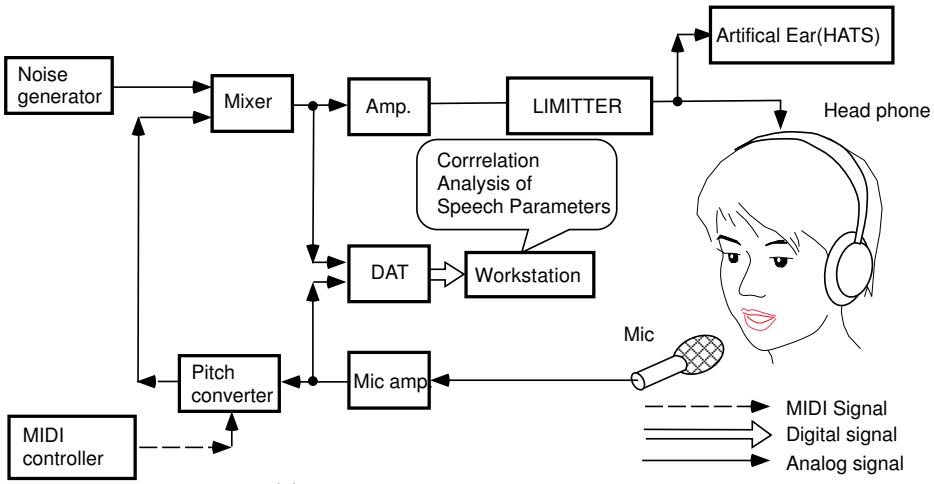

図 2　Schematic diagram of TAF experiment

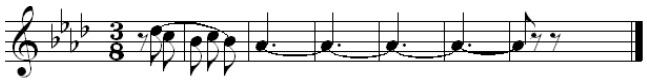

図 3　An example of a musical score used for an instruction

歳-22 歳) 計 6 人 (男性 3 人、女性 3 人) とした。被験者は、まず、母音「あ」を一定の音の高さで 9 秒間程度発声するように教示された。この条件では、通常の話し声と 3 種類の音の高さ (C3、G3、C4) の歌声（クラシック唱法）を収録した。次に、被験者は発想記号の付された楽譜を提示され、指示された発想で母音「あ」を用いて歌唱し、最後の音符を同じ発想の下で同様に 9 秒間延ばすように教示された。被験者に提示した楽譜の例を図 3 に示す。 1 つの課題に対して 2-3 分間繰り返し、意図的なヴィブラートを付けないよう指示した。実験は防音室内で行った。被験者とマイクとの距離は 40cm であった。

3.3　解 析 方 法

収録された数分間の長さの音声ファイルから、音声区間毎のファイルが切り出された。それぞれのファイルについて、周波数領域の不動点に基づく方法により、発声とフィードバックの両音声の基本周波数が 5ms 毎に抽出された。図 4 に女性被験者の発声した音声とフィードバックされた音声の基本周波数軌跡の例を示す。

変調信号を作成するために用いられた疑似白色系列の一周期分との相互相関を計算することにより、システムのインパルス応答が求められる。図 5 に、求められた相互相関の例を示す。この例から明らかなように、組織的な同期加算法と解釈できる疑似白色系列との相関の計算を行っても、S/N は低く、このままではシステムのパラメタを求めることは困難である。

まず、フィードバック音声についての相互相関を同期のための手がかりとして、実験全体の音声を用いることにより、50〜60 回の同期加算を行うことができる。こうして求められた応答を、前報で行ったシミュレーションのように対数的時間軸加工とフィルタリングを組み合わせて処理することにより、線形の応答成分以外を抑圧することができる。こうして処理された結果を図 6 に示す。

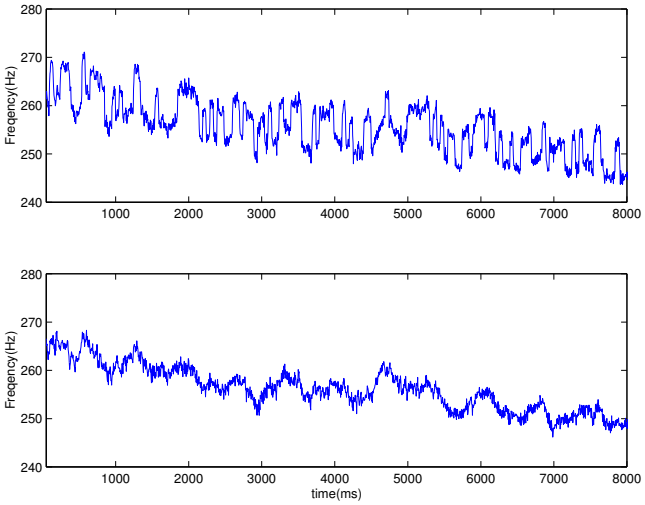

図 4　Extrated F0 trajectories. (upper: feedback speech F0, lower: produced speech F0)

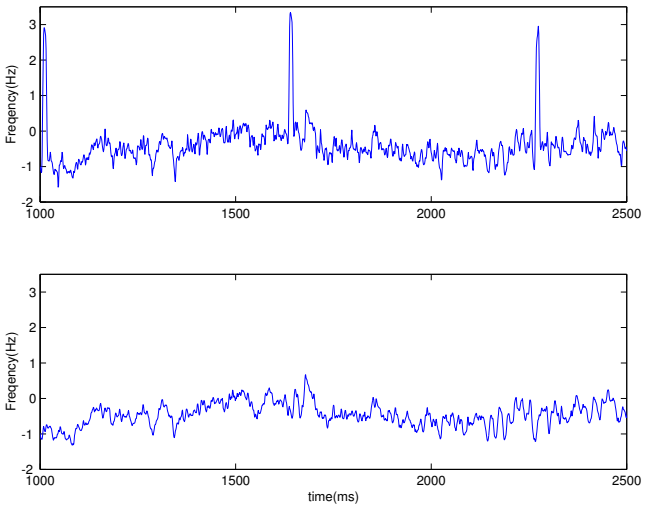

図 5　Cross correlation with the original PN sequence (upper: feedback speech, lower: produced speech)

この例には大きなトレンド成分がある。今回の解析では障害とならないため、トレンド成分を除去せずに解析を進めた。

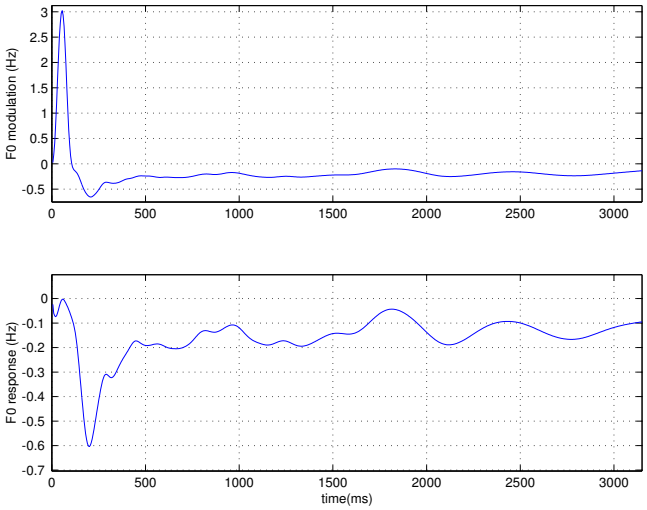

図 6　Processed correlations (upper: feedback speech, lower: produced speech)

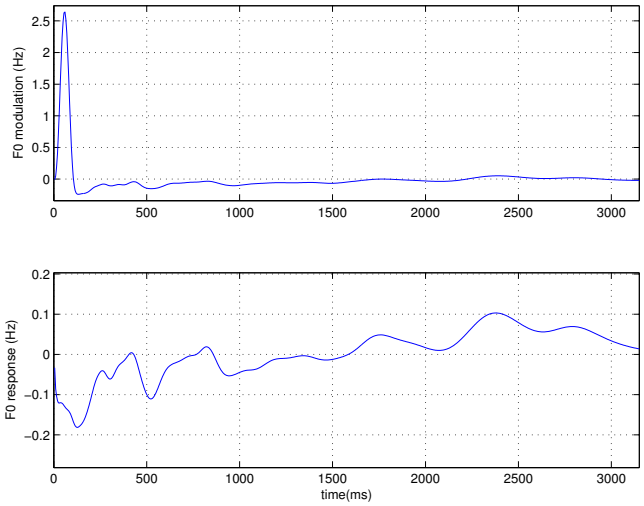

図 7　Abnormal response example by a male subject (upper: feedback speech, lower: produced speech)

4. 実験結果

前節で示した図 6 は、典型的な例である。この例に見るように、正方向への基本周波数の変動に対して、発声の基本周波数は百数十 ms 遅れで逆方向に動いて応答する。男性 3 人、女性 3 人の被験者のうち男性 1 人を除いては、同様な応答を示した。参考のため、他の被験者とは異なる応答を示した男性被験者の応答を図 7 に示す。

以下では、この特異な被験者のデータを除いたものを解析の対象とした。まず、図 8 に示すようにして、応答の潜時と強度を求めることとした。潜時を基本周期の関数として表したものを図 9 に示す。図から明らかなように、基本周期と潜時との間には高い相関がある。この相関は、人間が基本周波数の変化を検出するために用いているアルゴリズムの構造によるものであろう。この結果は、基本周波数変化の検出にはある個数の基本周期を観測することが必要であることを示唆しており、そこでは時間的手がかりが用いられていることを示唆する。

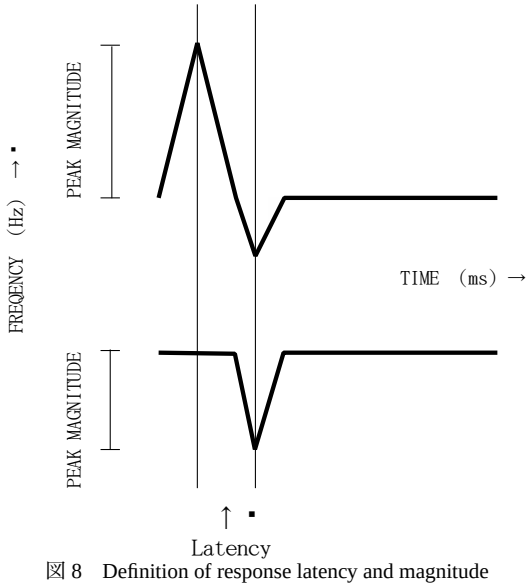

図 8　Definition of response latency and magnitude

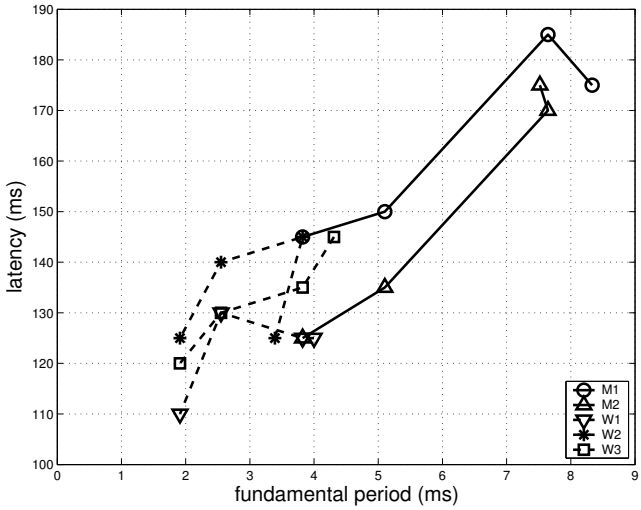

図 9　Latency as a function of fundamental period

その他の傾向を調べるため、探索的統計ツールである JMP（SAS Institute Inc.）を用いて、測定結果の分析を行った。なお応答の強度を示す応答のピークの高さの和は、セント値に変換した。分析結果を、箱ひげ図 (box plot) として図 10 に示す。左側が潜時、右側が応答の強度である。図の上段は、音高および発声の種類、中段が歌唱法、下段が被験者（男性：M、女性：W）を表している。図中には、データ点と分位置と平均値が示されている。箱ヒゲ図の上端と下端はそれぞれ 75%、25%の分位置、箱中の線は、中央値である。

これらの結果について、分散分析を行ったところ、潜時に関しては、被験者 (DF:4, F:187, Prob.<0.0001) と音高および発声の種類 (DF:3, F:13.2, Prob.<0.0001)、交互作用 (DF:12, F:8.1, Prob.<0.0001) に顕著な効果が認められた。応答の強度についても、被験者 (DF:4, F:21.7, Prob.<0.0001) と音高および発声の種類 (DF:3, F:120, Prob.<0.0001) に同様な傾向の効果が認められた。

5. 動的パラメタの推定

前節では、データからボトムアップに求められた値の性質を

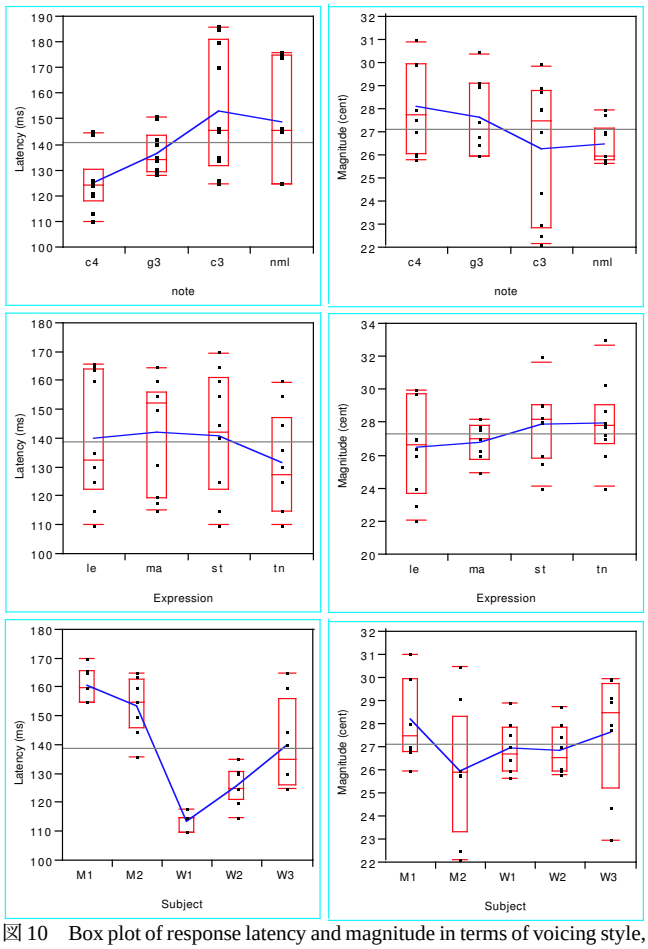

図 10　Box plot of response latency and magnitude in terms of voicing style, musical expression and subject.

調べた。歌唱システム構築のためには、作業仮説として用いている図 1 の各サブシステムのパラメタを求めることが必要である。ここでは、まず、特徴がはっきりとしている早い応答成分に対応する H2 として二次系を仮定し、非線形最適化によりパラメタを推定した。推定するパラメタは、極の周波数と帯域幅および図 1 中に a として表されている帰還ゲイン、そして、F0 抽出のための脳内での計算時間に相当する遅延時間である。最適化する評価関数としては、モデルの振幅周波数特性で正規化した測定された伝達関数とモデルの応答の自乗誤差を用いた。なお、この評価関数の値は、3Hz から 10Hz の範囲の値から計算した。遅いシステムである H1 については、まだ測定に伴うノイズが大きいため平均遅延時間と積分値として求めた応答の強さ（相対値）をパラメタとして求めることとした。

ここでもまず、前節と同様に、基本周期と遅延時間の関係について調べる。図 11 に、その結果を示す。図 9 と同様に基本周期と遅延時間が直線的に対応する傾向が認められた。ここで求められる遅延時間では、潜時に含まれていた発声器官の機械的応答による遅れが除かれていることになる。すなわち、ここでの遅延は基本周波数抽出と運動指令作成の情報処理の時間に相当する。

図 12 に前節と同様の箱ひげ図を示す。今回の目的変数は、遅延時間、フィードバックゲイン、共振周波数である。それぞれ、図中の左、右、下の図に対応している。それぞれの中の図に対応する要因は、前節と同様である。

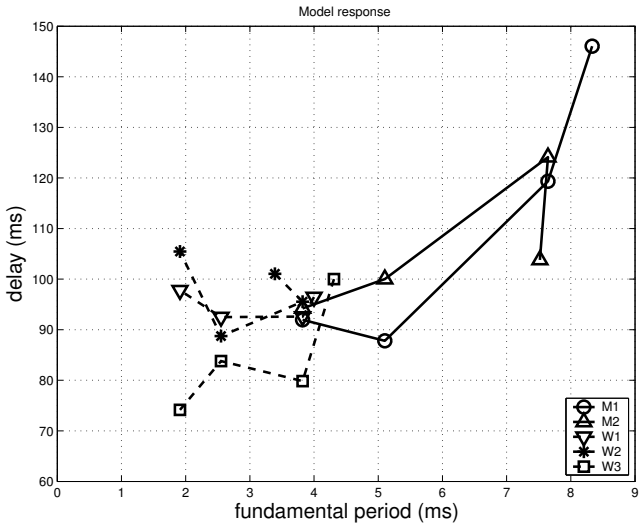

図 11　Delay as a function of fundamental period

これらの箱ひげ図の視察により、遅延時間は潜時、フィードバックゲインは応答の強度に類似した傾向があること、中心周波数は、被験者に強く依存するが、他の要因への依存度が少ない等の特徴が認められた。

遅いシステムについては、パラメタの抽出におけるノイズが大きいため、被験者以外の要因についての効果は明らかではなかった。幾つか値の不確実なものが含まれているが、単純な平均は、遅い応答の重心が 300ms 付近にあることを示した。これは、意識的な応答に関連する P300 の潜時と同程度であり、H1 に対応する遅いシステムがピッチの知覚に関わっているとする仮説に整合する。

6. 考　　察

推定された動的パラメタの会話音声と歌唱での違いは、歌唱における発想による違いよりも大きい。しかし、それらの違いの大部分は、基本周波数の違いによって説明される。ただし、同じ基本周波数であっても、会話音声では動的パラメタは、広い範囲に散らばる傾向がある。これらの結果は、会話音声と歌唱では、発声器官の機械的特性や基本周波数を抽出する脳内処理などの基本的な部分では共通のシステムと処理を用うことを示唆する。言葉を代えるなら、会話と歌唱の違いは、制御指令の違いに帰着される可能性を示唆する。これは、歌唱システムを構築する上では好都合な知見である。動的パラメタをある限られた条件で求めておけば、わずかな変更で他の条件でも利用が可能になるからである。

7. ま　と　め

基本周波数軌跡に対する聴覚フィードバックの影響をインパルス応答として計測し、基本周波数制御モデルの動的パラメタが会話と歌唱とでどのように変化するかを調べた。全ての実験条件において、聴覚的にフィードバックされた基本周波数の変化に対し、生成される音声の基本周波数は逆方向の応答をしめした。応答の潜時は、会話と歌唱とで顕著に異なり、歌唱内でも基本周波数に大きく影響された。また、被験者による大きな違いが認められた。強度については、被験者の要因が顕著で

あった。機械的応答を表すと想定されるシステムの動的パラメタとして、共振周波数を求め、フィードバックゲインと、挿入される遅延時間を求めた。その結果、遅延時間とフィードバックゲインは、それぞれ応答の潜時と強さと同様の依存性を示した。また、共振周波数では、被験者の要因が顕著な効果を示した。これらの結果は、会話と歌唱の基本周波数軌跡の違いは、フィードバック情報への依存度の違いと制御指令の違いによって生み出されていることを示唆している。また、潜時と遅延時間の基本周波数に対する依存性は、基本周波数を抽出するための聴覚系での処理遅延を反映したものであることを示唆している。今後は、こうして求められたパラメタを用いて様々な演奏スタイルのための制御指令を求める逆問題を解き、歌唱システムの構築に向けて検討を進める予定である。

文　　献

[1] 河原英紀, 片寄晴弘, “高品質音声分析変換合成システム STRAIGHT を用いたスキャット生成研究の提案”, 情報処理学会論文誌, Vol.43 No.2, pp.208–218, Feb.2002.

[2] Fujisaki, H. and Hirose, K., “Analysis of voice fundermental freqency contours for declarative sentences of Japanese” J.Acoust.Soc.Jpn.(E), Vol.5 No.4, pp. 233–242, 1984.

[3] 酒井弘, “新版発声の技巧とその活用法”, 音楽之友社, pp. 25–26, 1990.

[4] 小田切わか奈, 森大毅, 粕谷英樹, “歌唱のピッチ遷移に関する検討”, 日本音響学会秋季講論集, pp.537–53, 2000.

[5] 矢田部学, 粕谷英樹, “歌声の基本周波数の動特性”, 日本音響学会秋季講論集, 3-8-6, 1998.

[6] 小田切わか奈, 粕谷英樹, “歌声のヴィブラートの分析、合成、知覚に関する検討”, 日本音響学会秋季講論集, 1-7-5, 1999.

[7] 北風裕教, 赤木正人, “歌声に含まれる基本周波数の微細変動成分の知覚に関する研究”, 日本音響学会秋季講論集, 2001.

[8] Burnett, T. A., Freedland,M. B.,Larson, C. R., and Hain, T. C., “Voice f0 responses to manipulations in pitch feedback”, J.Acoust.Soc.Am., Vol.103, pp. 3153–3161, 1998.

[9] Hain, T. C, Burnett, T. A., Larson, C. R., and Swathi Kiran, “Effects of delayed auditory feedback(DAF) on the pitch-shift reflex”, J.Acoust.Soc.Am., Vol.109, pp. 2146–2152, 2001.

[10] Burnett, T. A., Larson, C. R., “Early pitch-shift response is active in both steady and dynamic voice pitch control”, J.Acoust.Soc.Am., Vol.112, pp. 1058–1063, 2002.

[11] 河原英紀, “声を使って聴覚を探る”, 日本音響学会論文誌, Vol.53, No.9, pp.731–737, 1997.

[12] H. Kawahara and J. C. Williams, “Effects of Auditory Feedback on Voice Pitch ”, in Vocal Fold Physiology eds. Davis and Fletcher, Singular Publishing Group , pp.263-278, 1996.

[13] H. Kawahara, “Transformed Auditory Feedback: The collection of data from 1993.1 to 1994.12 by a new set of analysis procedures”, ATR Technical Report, 2002.2.

[14] 斎藤毅, 鵜木祐史, 赤木正人 “歌声における F0 動的変動成分の抽出と F0 制御モデル”, 日本音響学会聴覚研究会資料, H-2001-92, 2001.

[15] Hideki Kawahara, Haruhiro Katayose, Alain de Cheveigne, Roy D. Patterson, “Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity”, Proc. EUROSPEECH’99, Vol.6, pp.2781–2184, 1999.

[16] Hideki Kawahara and Ryuichiro Yanaga, ”Filtering on Non-Linear Time Axis and its Application for Measuring Perception to Production Transfer Functions in F0 Control”, Speech Dynamics by Ear, Eye, Mouth and Machine, Kyoto, Japan, June 2003.

[17] 日本産業衛生学会, ” 許容濃度等の勧告”, 産業医学, 32(5) pp381–401, 1990

[18] 川人光男, ” 脳の計算理論”, 産業図書, 1996

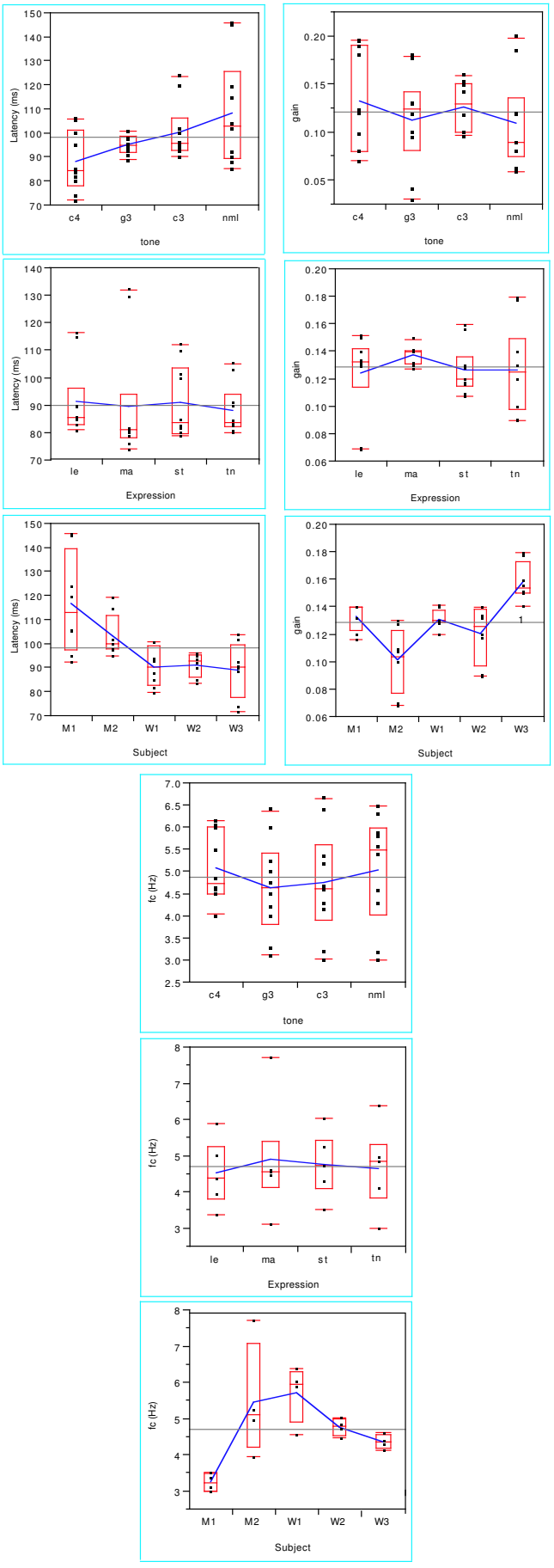

図 12　Box plot of delay, feedback gain and center frequency in terms of voicing style, musical expression and subject.