音楽情報科学　39－19
(2001. 2. 23)

音声分析変換合成システムSTRAIGHTを用いたスキャットの生成について

河原英紀，片寄晴弘

和歌山大学、 CREST
〒640-8510 和歌山市栄谷930
kawahara@sys.wakayama-u.ac.jp

あらまし　音楽としての歌唱の魅力は，歌詞を伴うことに多くを負っている．しかし，歌詞の理解できない外国語の歌唱であっても，楽器としての人間の声の魅力を楽しむことができることも事実である．ここでは，楽器としての声そのものの魅力を楽しむスキャット，ヴォーカリーズ，口三味線，鼻歌などを対象として取り上げ，音声処理技術を用いて，その魅力の分析，再合成，加工を行うシステムの開発を狙う一連の研究プログラムについて紹介する．具体的には著者等が開発しているSTRAIGHTを分析合成エンジンとして利用し，基本的な反射弓を修飾する発声制御モジュール，韻律制御モジュール，音楽情報処理モジュール，インタラクション制御モジュール等を逐次更新して行く生態学的枠組に基づく開発戦略を採用する．本システムは，計算機音楽への直接の応用の他に，パラ言語情報処理技術に対してもユニークなベンチマークの手段を提供するものと考えられる．

キーワード　基本周波数，音声分析合成，発声，聴覚，音楽

Scat generation system based on STRAIGHT: a versatile speech analysis, modification and synthesis system

Hideki Kawahara and Haruhiro Katayose

Wakayama University,　CREST
930 Sakaedani, Wakayama, Wakayama, 640-8510 Japan
kawahara@sys.wakayama-u.ac.jp

Abstract　A research program to develop a versatile system for analysis, manipulation and generation of a specific vocal music *genre*; scat, vocalease, *kuchi-jamisen* and humming, is introduced. The one of the major aim of the program is to explore why vocal music is still attractive, even if their lyrics are not intelligible due to foreign languages. This may sound peripheral to the usual belief that lyrics is the central charm point of vocal music. However, we argue that this type of research is indespensable for understanding roles of para-linguistic components in speech and vocal music. The proposed system uses STRAIGHT as its central analysis, modification and synthesis engine, and will refine its constituent modules like voicing control, prosodic control, musical information processing, interaction control and so on organized as modifiers of the basic reflex arc, in an evolutional and developmental process. This program provides a unique test bed for para-linguistic processing algorithms as well as wide variety of opportunities in computer music applications.

key words　fundamental frequency, analysis and resynthesis, voicing, hearing, music

1 はじめに

既にリタイアした歌手の歌声を聞きながら,「この人にあの歌を歌ってもらえたら...」と想うことがある．華やかな技巧をちりばめた曲を聞きながら「この曲を自分が歌えたら...」そう想うこともある．それらの願いは，多くの場合，時期的な制約，能力的な制約，時間的な制約のために叶えられることは無い．しかし，この状況は変わりつつある．計算能力と記憶容量の爆発的な増加を背景とすれば，聴覚情報処理，音声情報処理（特にパラ言語情報），音楽情報処理，インタラクション技術，運動制御の計算理論等を総合することで，これらの願望を満たすシステムを実現することは工学技術の射程に入りつつある．そのような個人的な願望の充足のために高度の技術資源を集中することは，効率と速度の二十世紀的価値観からは，非常識なことであろう．しかし，人間の世紀となるべき二十一世紀の入口に立った現在，価値観の中心を個人の幸福に置く技術体系の構築を真剣に考えても良いのではないだろうか．

ここでは，筆者等によって開発が続けられている高品質音声分析変換合成システムSTRAIGHT[6, 9, 23] の応用として，上記の目標を実現する上での里程標となるシステムの構想を提案し，幾つかの要素技術に関する予備検討を行った結果について報告する．

2 目標の設定

ここで実現を目指すシステムでは，特定の言語に依存する歌詞を扱わないこととする．歌詞を扱わないことにより，声と表現，情緒，感動との関係を，より直接的に追求することを狙う．システムのアーキテクチャの選択にあたっては，計算機メタファーに基づく安易な工学的モジュールへの機能分割は行わない．低次の反射弓とそれを修飾する高次の反射弓や計画-制御モジュールが層状に積み重なった並列階層システム [26, pp.396-400] として構成することを狙う．いわば，進化と発達によって形成される脳のアーキテクチャに倣うのである．以下では，相互作用を通じて発達するシステムとしての発声・発話を簡単にまとめてみたい [13].

発声は，母親等の養育者との相互作用から始まる．言葉を発するに到る前に，マザリーズと喃語での相互作用が長く続く．ここでは，養育者の声と自分の声の同一性の知覚が，相互作用と発達を導く．この期間を通じて，基本周波数の変化パターンと母音のような音によるバリエーションを中心とする喃語は，徐々に複雑さを増す調音運動をレパートリーに加え，言葉に到る．こうして最終的に獲得されるレパートリーは，それぞれの言語環境に固有のものとなる．しかし，言語依存部分の下には膨大な共通の基盤がある．この基盤は，発声と調音を制御する基礎的な機構から構成されており，言語に依存しない生物学的，生態学的拘束の下に形成される．スキャットの生成は，主にこの基盤を利用し，子音や母音等のレパートリーは，言語の語彙的，統語的，意味的内容からは切り離されて音の素材として利用される．

このようにして生成されるスキャットと同じものが生成できるようなシステムを工学的手段で実現することを目標とする．開発戦略としては，組織的ダウングレードとでも呼ぶべき方法と，発達過程の模倣という二つの戦略を採用する．

組織的ダウングレードでは，まず，十分に鑑賞に耐えるスキャットを，STRAIGHTを用いた分析変換合成により，作成するところから出発する．次いで，高い品質を保ったまま，STRAIGHTの合成器に渡すパラメタの時系列を，プロトタイプモデルの出力と置き換える．初期の段階では，プロトタイプモデルの出力が鑑賞に耐えるスキャットを生成するためには，プロトタイプの動作に多くの修飾を加える必要がある．モデルの詳細化と洗練のステップを踏む毎に修飾をサブシステムのプログラム化された動作で置換えて行くことにより，最終的には，楽譜と演奏意図を与えるだけで，鑑賞に耐えるスキャットを生成するシステムを実現する．

発達過程の模倣は，組織的ダウングレードによって作成されたシステムから出発する．ここでは，内部で利用する適切な目標軌跡を，強化学習の枠組みで，教師の演奏を模倣し練習することで形成することのできるシステムの実現を狙う．

3 技術課題

前節で示した目標を実現するためには，様々な技術的課題を解決することが必要となる．以下では，それらの課題について，この枠組みの下での従来の研究の位置付けに関する若干の議論とともに紹介する．また，それらの技術的課題の解決とSTRAIGHTの構成要素との関わりならびに解決策の実装について論ずる．

3.1 基本周波数の精密な制御

単純なパルス音源を用いるVocoder型の音声合成システムでは，高い基本周波数の音声のピッチ[1]を音楽に必要な精度で制御できないという問題があった．これは，標本化周期が基本周期に対して無視できない大きさとなることによる．正弦波合成型のシステム [15] では，問題とならない．パルス音源の場合であっても，標本化時刻と本来パルスの存在すべき時刻との差を補償するような直線位相特性の付与によって，この問題を回避することができる．STRAIGHTでは，群遅延操作を行う音源生成

[1]ここでは，ピッチは知覚される心理量，基本周波数は信号の物理的属性を表すものとして，使い分ける．

機構を用いているため，その機能を流用して離散化された時刻との差を群遅延で補償するような実装が行われている．

3.2 基本周波数制御の動特性

声の基本周波数は，楽器と比較するとはるかに複雑な軌跡を示す．しかし，少なくとも，伝統的な西洋音楽の歌唱では，離散的な表現である楽譜の旋律をそのまま楽器で演奏したものと同様な離散的な要素の組み合せと聞くことのできるような演奏も，それほど不自然では無く可能である．ここでは，まず，離散的な『目標』が与えられた場合の基本周波数の軌跡を，物理的には離散的なものとさせない要因とその影響を概観する．

3.2.1 基本周波数制御の生理的制約

発声には多くの器官が関わっている [17]．基本周波数は，最終的には，声帯の振動速度により決まる一つのパラメタで表される量である．しかし，その制御は，喉頭周辺だけでも15種類の筋肉に依存しており [17, pp.11-15]，呼気の供給に関与するものを加えると，20種類を超える過剰な自由度を持つシステムにより行われている．

このようなシステムにおいて，基本周波数を精密に制御することは困難な課題である [17, pp.279-306][4, 3]．しかも，基本周波数を一定にさせない多くの要因が存在する．基本周波数を一定に保つには，呼気の排出に伴う声門下圧の低下，心拍による声門下圧の変動や脈波による声帯質量の変動 [16]，筋肉への指令パルスの確率的変動等の影響を補償することが必要となる．

3.2.2 声道の狭窄の影響

子音の調音では，通常，声道の途中に狭窄あるいは閉鎖が形成され呼気流が妨げられる．呼気流への抵抗は，等価な声門下圧の低下につながる．その結果，多くの場合，基本周波数は子音の部分において低下する．これらの言語情報の干渉を受けて，基本周波数軌跡には，計画された軌道からの局所的な逸脱が重畳することになる．

3.2.3 聴覚フィードバックの影響

発声時の基本周波数は，聴覚により常にモニターされて修正されている [10, 14, 5]．モニターされた結果は，速度が遅いが大きなループゲインを有する系と，速度は早いが小さなループゲインを有する系とにより，並列にフィードバックされ，ずれを補償しているようである．また，早い系であっても，基本周波数の変化に反応するまでには，全体として100 ms以上のむだ時間を含んでいる [10]．このような系を，フィードバックによって制御することは困難である．従って，制御が逐次的なのではなく，逐次的なフィードバック誤差学習によって動特性の方に学習が行われていると考えた方が良い．具体的には，大脳による調整への関与を得て，順モデルと逆モデルが形成されて，制御そのものは前向き制御で行われると見るのである．なお，最近の研究は，目標値そのものも，聴覚からのフィードバック情報に基づいてゆっくりと修正されて行くことを明らかにしている [5]．

3.3 基本周波数軌跡の計画

言語音声の基本周波数軌跡の計画に関しては，藤崎による先駆的な試みがある [4]．ただし，これは習熟した言語行動に関するモデルであり，素人の段階から玄人に到る様々な習熟段階の歌唱にまで適用して良いか否かは明らかでは無い．また，(may not be exactly critically-damped)[4, page 234] との保留をつけられながらも文章音声の良いモデルであるとされている臨界制動二次系は，少なくとも歌唱における基本周波数の動特性の記述のためには拘束が強過ぎる怖れがある[2]．ここは，既存のモデルを流用するのではなく，自分自身を含んだ環境と相互作用するシステムとして歌唱を捉え，基本的なレベルから議論を再構築すべきであると考える．ここで，急速に発展している運動制御の計算理論の枠組み [11, 26, 12] をこのレベルに適用することは，本質的な理解と適切なモデルの構築のための鍵となろう．

3.3.1 ピッチ感覚の時間特性

歌唱の制御に対して運動制御の計算理論の適用を困難にしているのは，目標と実現との比較が行われるレベルと表現が明らかではないことによる．物理量である基本周波数に対応する心理量であるピッチの知覚は，数100 msという大きな時定数を持つ遅いプロセスである（例えば [28, pp.53-57]）．しかも，基本周波数の情報が利用可能になるまでには，基本周期が5回繰返す程度の処理時間が必要であり [10]，逐次制御には間に合わない．基本周波数が早い速度で変化する場合のピッチ知覚（特に歌唱に関連するような）については，比較的単純な軌跡についての組織的な検討 [1, 2] が開始されたばかりである．なお，基本周波数の周波数変動の中の十数 Hz以上の速度で変化する成分は，ピッチ情報としてではなく，声に自然性を与える成分としての役割を担っていることが示唆されている [31]．

3.4 音源の群遅延パラメタの制御

音源の群遅延パラメタの制御は，STRAIGHTにより初めて音声合成に導入され [22]，自然性の向上に大きく貢献した．しかし，現在のシステムで用いられている既

[2]現状の理解において，先行研究とは [25, 18]，この点で異なった見方をしている．

定値は，少数の話声のサンプルを用いた検討によって決定されたものであり，一般的な話者グループに最適である補償は無い．また，予備的な検討によれば，歌声の場合には，話声と比較すると遥かに小さな群遅延の広がりが適していることが観測されている．STRAIGTHへの応用を目的として，合成のための群遅延パラメタを音声から抽出することを試みている．一つは，周波数領域での写像の不動点に基づく方法であり [8, 21]，もう一つは，時間領域での写像の不動点に基づく方法である [24, 7, 20]．このパラメタの利用は，だみ声から透明な声，気息性の声までの制御を可能にして歌唱音声の表現の幅を広げる．

3.5 基本周波数とスペクトル包絡の相関

基本周波数が変化すると，音源波形と声道形状の双方に依存するスペクトル包絡は，幾つかの要因により変化する．それらは，基本周波数の調節メカニズムに起因する構造的なものであったり [27]，局所的な声門の開閉比率の変化に伴う音源スペクトル形状の変化であったり，聴覚フィードバックに基づく響きの局所的な調整によるものであったりする．自然な歌唱音声の生成には，これらの規則を取り入れる必要があろう [29]．

3.5.1 包絡ピークと調波の相関

歌唱では，できるだけ音源から供給されるエネルギーが効率よく音となって放射されるよう，スペクトル包絡のピークは，基本周波数の調波の位置に調整される傾向がある [17, pp.231-232]．現在のSTRAIGHTに，基本周波数を変更した時のこの効果を導入することにより，歌声らしさをより強調することができる．

3.5.2 基本周波数とスペクトル傾斜の相関

同一人物であっても異なる周波数領域では，異なった発声法を用いる．それらの発声法は，異なった声帯振動様式に結びついており，音声のエネルギを供給する声門での駆動条件の変化として実現されている．これらの差違は，音源スペクトルの大局的傾斜に大きな影響を与える．

3.5.3 歌唱フォルマント

オペラ歌手の歌唱の分析で発見された2000 Hz～3000 Hz付近でのエネルギーの強調は，歌唱フォルマント（singer's formant）と呼ばれている [17, pp.239-241]．クラシックの楽曲の歌唱においては，このようなスペクトル包絡の変型を模擬することも必要となる．

3.5.4 基本周波数の変換と個人性の保存

ここまでに示したような基本周波数とスペクトル包絡との様々な相関は，スペクトル包絡を変形し，同一人物であっても異なった基本周波数／歌唱法に属するスペクトル包絡の同一性／相似性は崩れる．しかし，そのような変形にもかかわらず，多くの場合に聴取者は同一人物の発声に一貫性を感ずることができる．これは，話声の基本周波数の10%の変化で，個人の同一性が失われるとする言語音声に対する従来の知見と鋭く対立する．

3.6 表情とスペクトル包絡の相関

もう一つ必要な操作軸は，基本周波数，個人性を保ったままの，表情の変化である．声門の閉止部分の長さと開放部分の長さの比と音色との関連についての検討は存在する．しかし，それ以上微妙な『明るさ』や『柔らかさ』『透明感』『だみ声』等の属性．更には感情との関連の検討が必要である．

3.7 調音の選択

速度の大きな旋律をスキャットで歌う場合，「ダバダバ～」「ドゥブドゥブ～」のように子音と組み合わせ，さらに二種類の異なった子音を交互に挟むことが良く行われる．ゆっくりとした旋律の場合には，同一母音の繰り返しや「ラ～ラ～」のように子音と組合わせる場合でも同一のものを繰返す傾向がある．この傾向の背景にある規則を抽出し，旋律の局所的速度に応じて，適切な子音の組を自動的に選択する仕組みを組み込む必要がある．

3.8 特異な声帯振動への対応

いわゆるクラシック音楽における歌唱では，周期性がはっきりとし，（ビブラートを除けば）安定した基本周波数を持つ声が用いられる．しかし，日本の演歌や様々な伝統歌唱，ポピュラー音楽の領域では，だみ声や叫び声，気息性の声等，多様な表現が用いられる．周波数領域の不動点による方法 [8] や，時間領域の不動点による方法 [7, 20] を用いると，それらの多様な表現に特徴的なパターン[3] を認めることができる．しかし，それらを定量的で操作可能なパラメタとして表現することと，パラメタの操作結果を再合成に反映させる機構の開発は，これからの課題である．

3.9 演奏意図のモデル化

意識の脳神経基盤は，未だに決着の着いていない困難な問題である．ここでは，意識は，自己の（観察可能な）内部状態を観測した結果として作り上げられる『説明』であるとする立場を取る．意識は，脳内で実際に進行している膨大な情報処理には直接関与することができず，

[3] 後で示す『だみ声』の例（図5）では，低いC/Nの領域を持つ不動点の軌跡が1オクターブの間隔で並行して走っている．通常の母音の場合には，安定して低いC/Nを示す領域を伴う不動点の軌跡は1本だけとなる．

結果として表出された（広義の）行動のみに基づいて出現するものであるとするのである．

演奏意図[4]は，この文脈の下では，演奏者の意図を観客が受取るときに，目的とした意図に受取られるように演奏するための修飾の方法に付与されたラベルである．聴衆と演奏者が共通の「意図から演奏（音として出現した）への順モデル」を持つ場合には，演奏者の意図があたかも直接的に観客に受取られることが可能となる．歌唱の場合には，器楽演奏と比較すると，生物学的に拘束される部分が多いため，演奏者の意図と観客が受取るものには共通部分が大きくなることが期待される．

しかし，意図のモデル化には，方法論的な問題がある．前節で説明した「ラベル」は直接的には観測できない仮説的な実体である．アンケートを用いることによって，間接的に，事後に，言語的解釈を経た後の結果として報告を得ることは可能である．演奏者や観客が，演奏中に意図し受取ったラベルをアンケートへの記入時まで保持でき，かつその（本来は連続的な）ラベルと（本来離散的である）言語的表現とが一対一に対応付けられるのであれば，この方法でも，意図のモデル化のための基礎データを得ることは可能である．しかし，一般的には，このような条件が成立していることを期待することはできない．むしろ，作成したシステムに，様々な意図で演奏した資料を模倣するように学習させた場合の制御パラメタの既定値からの修飾量のクラスタ分析と，それらクラスタと演奏者／観客の生理的指標やアンケート結果との相関の解析を通じて，明らかにすべき問題であろう．

3.10 インタラクションのモデル化

自分自身，伴奏，観衆等の状態を把握しインタラクションを制御するモジュールは，歌唱システムとして完結するためには，不可欠の要素である．しかし，この問題は，一般的な音楽演奏におけるインタラクションの枠組みで議論すべきであり [30, pp.206-217]，ここでは論じない．

4 実音声の分析と加工例

以下では，実際の歌唱音声の分析とSTRAIGHTを用いた加工の例を示す．図1,2は，異なった速度で歌われたスケールの基本周波数の軌跡である[5]．10年以上の西洋音楽の歌唱の経験のあるアマチュアの男女による歌唱である．上段は，ゆっくりとした試行の例を示す．下段の例の採取では，階段状のスケール感を保持できる範囲で，できるだけ速く歌うように教示した．スキャットに用いた言葉は，ゆっくりとした試行の場合には「ララララ～」，速い試行の場合には「ラバラバ～」である．

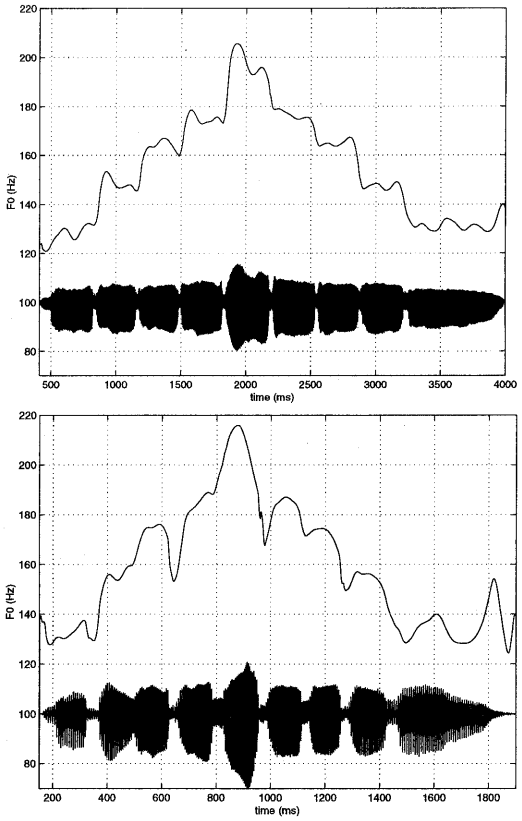

図 1: F0 trajectory examples by a male singer. (upper: slow performance, lower: fast performance)

データの収録には，Sony ECM-23F マイクロフォンを用い，44100 Hz, 16 bitの標本化を行った．基本周波数の抽出はSTRAIGHTに実装されている周波数領域での不動点に基づく方法を用いた．ゆっくりとした試行の場合には，5 ms，速い試行の場合には，2 ms毎に基本周波数を求めた．求められた基本周波数の誤差は，母音中央部では0.5 Hz以下である．ただし，/b/の子音部ではS/Nが低下し周期性も崩れるため，数Hz以上の誤差が発生している．

これらの図より，基本周波数を制御するシステムは，臨界制動二次系よりも遥かに制動不足の状態にあることが分かる．なお，これらの試行は，速い試行の場合でも，正しい音程を等間隔で歌っているように聞こえる[6]．速

[4]この言葉の示す概念は多数のレイヤーを有しており，この言葉は様々なレベルで使われる [30, pp.199-201]．ここでは，表現様式や手段ではなく，それらの操作を生み出す原因となる高いレベルを指す言葉として用いている．

[5]これらの図示，モデルの計算は，全て対数周波数を用いて行うべきである [4]．しかし，ここでは，便宜的に直線周波数を用いている．

[6]幼児期から馴染んでいる西洋音楽のスキーマに基づいて知覚レベ

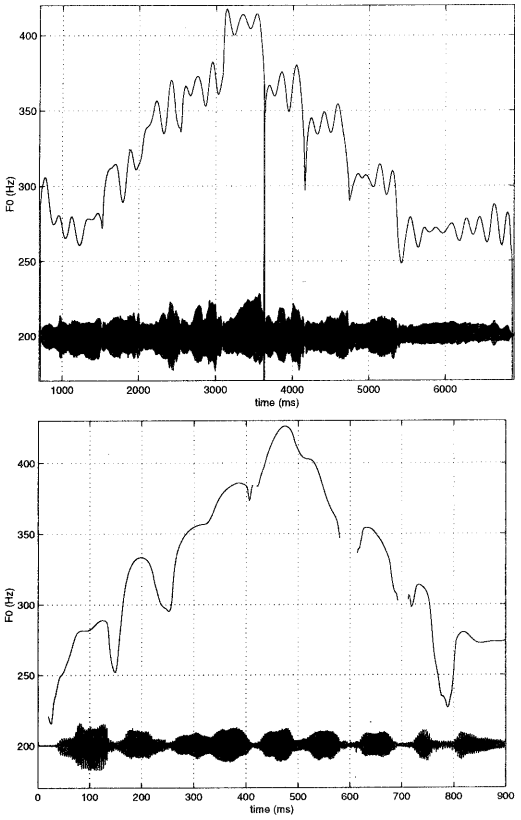

図 2: F0 trajectory examples by a female singer. (upper: slow performance, lower: fast performance)

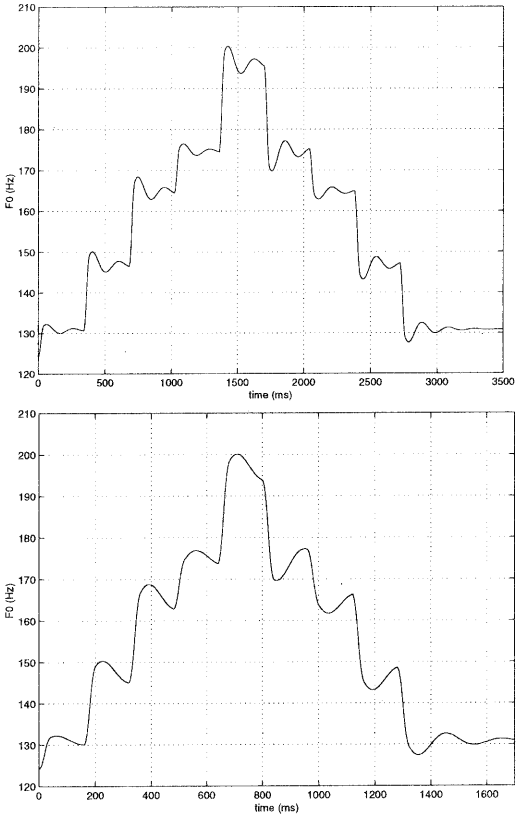

図 3: F0 trajectories gerated by a preliminary model. (with a feed forward control based on an internal model.)

い試行の基本周波数の軌跡からは，これらがどのような機構で，等間隔に正しい音程で演奏されたスケールであるように知覚されるのかを説明することが困難な課題であることも分かる．この問題は，ピッチ知覚機構の研究と並行して進める必要のある課題である．

4.1 軌跡生成の予備検討

図3に，変換聴覚フィードバック実験[19, 10]により求められた音声の生成知覚における基本周波数制御モデルに，そのモデルを順モデルとして内部に持つ前向き制御システムの入力として階段状のスケールを入力して生成した基本周波数軌跡の例を示す．上段は，一つの音符の長さが340 msの比較的ゆっくりとしたスケールの場合，下段は，一つの音符の長さが160 msの速いスケールの場合である．ここでは，子音の狭窄による局所的な基本周波数の低下はモデル化していない．この予備的モデルでは，ピッチ知覚の周波数特性，運動制御における制御周期，音声の生成知覚システムにおける時間遅れを考慮している．しかし，内部基準の順応，筋紡錘からの内部フィードバックによる10 Hz付近の極，随意系を介したピッチの補償システムはモデル化していない．参考のため，前向き制御系を外した場合の基本周波数の軌跡を図4に示す．これは，いわば歌唱の経験が全く無い場合をシミュレートしていることに相当する．

前向き制御を取り入れたモデルによる基本周波数の軌跡は，実際の音声で観測された軌跡の特徴を捉えている．しかし，これらの軌跡をそのまま利用してスケールの歌唱を生成しても，人工的な印象を完全には拭えない．

ルでのあいまいなピッチがスケールに乗るように解釈し直されているという可能性は排除できない．これらの試料は後述のウェブページに載せておくので，各自で判断頂きたい．

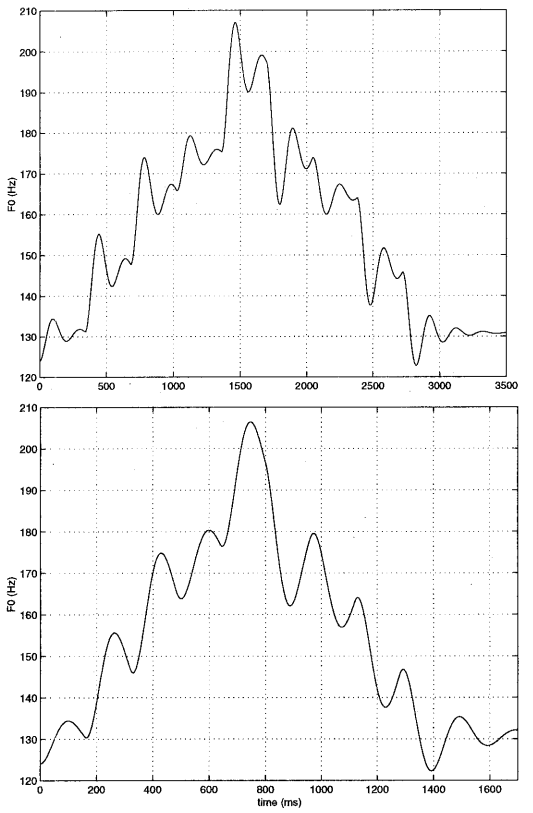

図 4: F0 trajectories generated by a preliminary model. (without a feed forward process.)

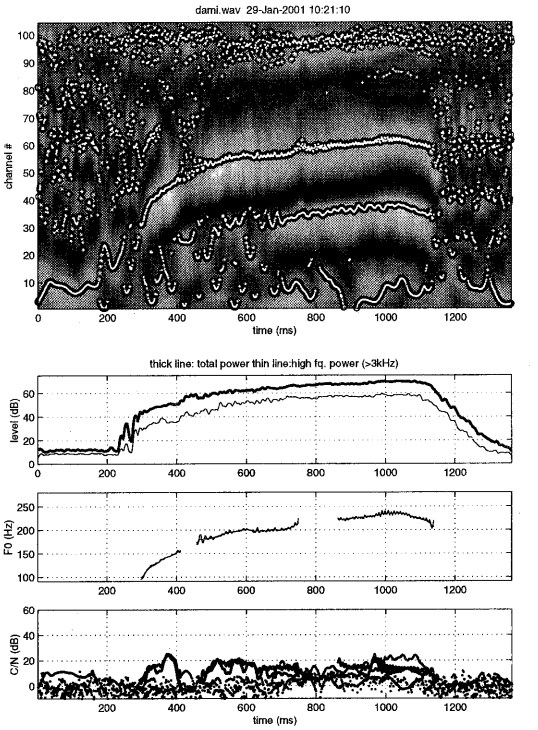

図 5: Source information for a "dami-goe" produced by a male performer.

4.2 特異な発声

図5に,『だみ声』の音源情報をSTRAIGHTにより分析した例を示す. 話声を対象としたシステムであるため, 出力としての基本周波数抽出結果には, 所々に欠落がある. しかし, 不動点とC/Nを表す最上段のマップには, 明瞭に, 声帯振動が二重周期を持つモードに移行していることが示されている. また, 最下段のそれぞれの不動点のC/Nも非常に低く, 声帯の振動が異常な状態にあることが分かる.

STRAIGHTによる分析結果を用いて, この『だみ声』の再現と, 基本周波数や, 速度, スペクトル包絡の変換による個人性やスタイルの変換を試みた. 予備的な実験の印象では, 基本周波数軌跡の欠落する部分でやや人工的な響きが生ずるものの, 自然な『だみ声』が得られた.

なお, これらのサンプルと, 加工音声の例は, 以下のページで参照可能となっている. 是非, 本資料での説明がどの程度成立しているのかを, 自分の耳で確認して欲しい.

`http://www.sys.wakayama-u.ac.jp/~kawahara/scat/`

5 まとめ

聴覚情報処理, 音声情報処理, (特にパラ言語情報処理) 音楽情報処理の総合的ベンチマークとして, スキャット生成システムの構想を提案し, 基本となる要素技術について紹介した. ここで提案したような研究プログラムは, 望む人の声で, 望む感情と表情で, 任意の曲の歌唱を合成するという, より挑戦的な目標の実現のための適切な里程標であろう. このような具体的な中間目標を設定することによって, 声のレタッチソフトや数多くの表情/表現のプラグインが開発されると共に, 聴覚, 音声, 音楽への理解が大きく深まることを期待している.

謝辞 本研究は，科学技術振興事業団CREST「脳を創る」領域の『聴覚脳プロジェクト』の支援を受けている．また，一部に科学研究費（基盤C：11650425）の支援を受けた．

参考文献

[1] Christophe d'Alessandro and Michaele Castellengo. The pitch of short-duration vibrato tones. *The Journal of the Acoustical Society of America*, Vol. 95, No. 3, pp. 1617–1630, 1994.

[2] Christophe d'Alessandro, Sophie Rosset, and Jean-Pierre Rossi. The pitch of short-duration fundamental frequency glissandos. *The Journal of the Acoustical Society of America*, Vol. 104, No. 4, pp. 2339–2348, 1998.

[3] Glenn R. Farley. A biomechanical laryngeal model of voice f0 and glottal width control. *The Journal of the Acoustical Society of America*, Vol. 100, No. 1, pp. 3794–3812, 1996.

[4] Hiroya Fujisaki and Keikichi Hirose. Analysis of voice fundamental frequency contours for declarative sentences of japanese. *J. Acoust. Soc. Jpn. (E)*, Vol. 5, No. 4, pp. 233–242, 1984.

[5] Jeffery A. Jones and K. G. Munhall. Perceptual calibration of f0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, Vol. 108, No. 3, pp. 1246–1251, 2000.

[6] Hideki Kawahara. Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited. In *Proceedings of IEEE int. Conf. Acoust., Speech and Signal Processing*, Vol. 2, pp. 1303–1306, Muenich, 1997.

[7] Hideki Kawahara, Yoshinori Atake, and Parham Zolfaghari. Accurate vocal event detection method based on a fixed-point analysis of mapping from time to weighted average group delay. In *Proc. ICSLP'2000*, Beijin China, 2000.

[8] Hideki Kawahara, Haruhiro Katayose, Alain de Cheveigné, and Roy D. Patterson. Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity. In *Proc. Eurospeech'99*, Vol. 6, pp. 2781–2784, 1999.

[9] Hideki Kawahara, Ikuyo Masuda-Katsuse, and Alain de Cheveigné. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. *Speech Communication*, Vol. 27, No. 3-4, pp. 187–207, 1999.

[10] Hideki Kawahara and J. C. Williams. Effects of auditory feedback on voice pitch. In Pamela J. Davis and Neville H. Fletcher, editors, *Vocal Fold Physiology*, chapter 18, pp. 263–278. Singular Publishing Group, 1996.

[11] Mitsuo Kawato. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, Vol. 9, No. 6, pp. 718–727, 1999.

[12] Mitsuo Kawato, K. Furukawa, and R. Suzuki. hierarchical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, Vol. 57, pp. 169–185, 1987.

[13] Patricia K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom. Linguistic experience alters phonetic percpetion in infants by 6 months of age. *Science*, No. 255, pp. 606–608, 1992.

[14] Charles R. Larson, Theresa A. Burnett, Swathi Kiran, and Timothy C. Hain. Effects of pitch-shift velocity on voice f0 responses. *The Journal of the Acoustical Society of America*, Vol. 107, No. 1, pp. 559–564, 2000.

[15] Robert J. McAulay and Thomas F. Quatieri. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Trans. ASSP*, Vol. 34, pp. 744–754, 1986.

[16] Robert F. Orlikoff and R. J. Baken. Fundamental frequency modulation of the human voice by the heartbeat: Preliminary results and possible mechanisms. *The Journal of the Acoustical Society of America*, Vol. 85, No. 2, pp. 888–893, 1989.

[17] Ingo R. Titze. *Principles of voice production*. Prentice Hall, 1994.

[18] 小田切わか菜, 森大毅, 粕谷英樹. 歌声のピッチ遷移に関する検討. 音講論, pp. 537–538, 9 2000.

[19] 河原英紀. 声を使って聴覚を探る. 日本音響学会誌, Vol. 53, No. 9, pp. 731–737, 1997.

[20] 河原英紀, Parham Zolfaghari. 群遅延情報を利用した音声の駆動情報の多重解像度分析について. 信学技報, EA2000-35, pp. 63–70, 8 2000.

[21] 河原英紀, Parham Zolfaghari, Alain de Cheveigné, Roy D. Patterson. 周波数から瞬時周波数への写像の不動点を用いた音源情報の抽出について. 信学技報, SP99-40, 7 1999.

[22] 河原英紀, 津崎実, Roy D. Patterson. オールパスフィルタの位相操作による時間構造制御とその知覚への影響について. 聴覚研究会資料, H-96-79, pp. 1–8, 1996.

[23] 河原英紀. 聴覚の情景分析が生んだ高品質Vocoder: STRAIGHT. 日本音響学会誌, Vol. 54, No. 7, pp. 521–526, 1998.

[24] 河原英紀, 阿竹義徳. 音声の群遅延特性に基づく声門閉止等のイベント抽出について. 信学技報, SP99-171, 2000.

[25] 矢田部学, 遠藤康男, 粕谷英樹. 歌声の基本周波数の動特性. 音講論, pp. 383–384, 9 1998.

[26] 川人光男. 脳の計算理論. 産業図書, 1996.

[27] 本多清志. Biological mechanisms for tuning voice fundamental frequency. 喉頭, Vol. 8, No. 2, pp. 109–115, 12 1996.

[28] 難波精一郎. 聴覚ハンドブック. ナカニシヤ出版, 1984.

[29] 田中公人, 阿部匡伸. 基本周波数の変更量に応じてスペクトル包絡を変形する音声合成方式. 信学論, Vol. J83-DII, No. 8, pp. 1722–1732, 2000.

[30] 片寄晴弘. 第4章：音楽情報処理. 文字と音の情報処理, 岩波講座：マルチメディア情報学, pp. 163–224. 岩波書店, 2000.

[31] 北風裕教, 赤木正人. 基本周波数の微細変動成分に対する知覚. 信学技報, SP99-168, 2000.