博 士 論 文

人間と計算機の協調的な音高判定技術と その応用に関する研究

Supervisor: Professor Dr. Kazushi Nishimoto

School of Knowledge Science
Japan Advanced Institute of Science and Technology

March 2013

要 旨

過去から現在に至るまで「声」を活用した情報処理技術へのニーズは高く，音楽情報処理においても古くから声の持つ有用性に着目した技術が開発されている．その一つが鼻歌入力（Voice-to-MIDI）である．Voice-to-MIDI は，歌唱されたフレーズを入力音として音高等を取得し，音符情報に自動変換する技術であり，ユーザはフレーズを歌唱するだけでよいため利便性が高い．そのため音楽制作や楽曲検索等に利用される他，様々な応用可能性がある．しかしながら，情報を取得する音を自動処理で選別するため，計算機が価値があるとした音とユーザが価値があるとした音とは必ずしも一致せず，結果，望んだ音情報の欠落や不要な音情報の混入等が起こる．これらは，計算機による完全自動処理では，音の取捨選択に関する人間の意思を反映させづらいことが原因と考えられる．そこで本研究では，音の時系列のうち人間が「価値がある」と判断した区間からのみ情報，特に音楽で重要な音高を取得できる，人間と計算機との協調的な音高判定技術の構築と，自動処理システムでは難しい課題への応用を試みる．本論文ではまず，音の取捨選択の仕組みとして，入力音に合わせて人間がリアルタイムに音区間の区切りをタップ入力し，計算機がその区間の音高を抽出する基盤システムを構築し，Voice-to-MIDI の従来からの課題である音高と音数の判定精度について評価する．次に，人間による音の取捨選択が可能な提案手法の特長をもって解決可能な課題として，Voice-to-MIDI システムの操作者自身の声以外を入力音とする事例を 2 例採り上げ，提案手法の適用を試みる．第一の事例は，自然音等の環境音が入力音の事例である．自動処理では音の区切りが難しい環境音に対して提案手法を適用し，人間による音の取捨選択で環境の音情景を音楽として再構成できる新しいリズム楽器を提案し，試用によって有用性を評価する．第二の事例は，他者が自由に発する非音楽的な音声が入力音の事例である．認知症患者が発する常同言語を入力とし，介護者が音を取捨選択することを想定した音楽療法支援システムを構築し，ケーススタディによって有用性を評価する．以上を通じ，まず提案手法の有用性，Voice-to-MIDI の応用可能性拡大への寄与を示す．次に人間と計算機との協調処理の観点および知識科学の観点から俯瞰を行うことによって，本論文の成果をまとめる．

目 次

図 目 次

第 1章

序論

1.1　本研究の目的

過去から現在に至るまで「声」を活用した情報処理技術へのニーズは高く，その代表例と言える音声認識技術や翻訳技術は，近年，スマートフォンの普及とともに実用性を獲得し，急速に浸透し始めている．

音楽情報処理においても声の持つ有用性に古くから着目されており，声を活用した技術が様々開発されている．その一つが鼻歌入力（Voice-to-MIDI）と呼ばれる，音響データから符号データへの変換技術である．Voice-to-MIDI は音楽制作や楽曲検索などに応用され，主にマイクを通じて歌唱されたメロディに対して，音高や音長などの情報を取得して音符情報に自動変換する技術である．ユーザは頭に浮かんでいるフレーズを歌唱するだけでよいため利便性が高く，音楽制作や楽曲検索にとどまらない応用可能性がある．

しかし，Voice-to-MIDI では通常，入力音（声とは限らない）が変換処理の対象になるか否かは処理アルゴリズムに依存するため，計算機が変換処理の価値があると見做した音と出力を受け取る人間にとって価値がある音が常に一致するとは限らない．その結果，望んだ音の情報の欠落や逆に不要な音の情報の混入などの問題が発生する．これらの問題は，計算機による完全自動処理では，音の取捨選択に関する人間の意思を反映させづらいことに原因があると考えられる．

以上のような背景において，本論文は，音の時系列の中から人間が「価値がある」と判断した区間を選び出し，その区間の情報，特に音楽で重要な要素である

音高を取得する技術を核とした研究についてまとめたものである．この技術では，

- 人間による「価値がある」区間の選び出しの手段の提供
- 計算機による,「価値がある」区間からの音高の算出

を特徴としており，これを人間と計算機との協調的な音高判定技術と定義する．

研究の最初の目的としてまず，人間と計算機との協調的な音高判定技術の基盤となるシステムの構築および従来の Voice-to-MIDI の課題であった音高と音数の判定精度向上課題への適用を行う．基盤システムは，音響データから符号データへの変換技術である Voice-to-MIDI の手法を応用し，マイクからの入力音と同時に，人間がコンピュータや電子楽器のキーボードなどからリアルタイムに音区間の区切りをタップ入力し，計算機はその区間の音高を取得する仕組みを持つ（この基板システムをタップ併用型 Voice-to-MIDI システムとする）．音の取捨選択に関する人間の意思を反映させやすくなることで，音数抽出の正確さが増し，それが音高判定の精度向上にも寄与することが期待される．

次に，従来の自動処理による Voice-to-MIDI システムが適用困難な事例として，Voice-to-MIDI システムの操作者自身の声以外を入力音とする事例を 2 例採り上げ，提案手法の適用を試みた．

第一の事例は，提案技術を歌唱などの声だけでなく，その環境に応じて得られるあらゆる音の時系列に適応できるように拡張した新しいリズム楽器システムを構築することを目的とする．このリズム楽器は，環境に応じて得られるあらゆる音の時系列に対して，ユーザが受けた刺激から醸成された心象風景をタップによる音の取捨選択という形で織り込んで音情景を切り取り，最終的に音楽フレーズという形で環境の音情景を再構成することを目的としている．この楽器によって環境と人の表現を融合した新たな音楽の作成が可能となり，加えて，創作を通じて環境のありように対する新たな気づきや視点の変化が促進されることが期待される．

第二の事例は，他者が自由に発する非音楽的な音声に対して音を取捨選択する課題として，認知症患者に対する音楽療法を支援するシステムを構築することを目的とする．認知症患者の不安などの心的状態から生じる症状の一つとして，何度も同じ言葉を繰り返す「常同言語」というものがある．このシステムは，常同

言語を含めた患者の発声のうち，介護者がタップによって指定した区間の音高を取得し，その音高を基にして決めた患者を落ち着かせることを目的とした音楽を演奏する．将来的には，このシステムによって，患者が常同言語の繰り返しなどの状態から抜け出すこと，音楽的には初心者である介護者が音楽療法に携わることの支援が期待される．

1.2 本研究の背景

1.2.1 Voice-to-MIDI（鼻歌入力）の意義

計算機を用いた音楽制作はDTM（Desk Top Music）と呼ばれ，以前は，MIDI（Musical Instrument Digital Interface）[1][2][3] シーケンスデータと呼ばれる自動演奏データのみしか扱えなかったが，近年主流のAvid社Protools[4] やSteinberg社Cubase[5] などのDAWソフトでは，MIDIシーケンスデータだけではなく，楽器演奏を録音したWAVE波形データを用いて楽曲を作成することも容易になっている．WAVE波形データは，音色，音高やリズム，アンビエンスなどを含めた演奏そのままを記録・再生したい場合に有用であり，フレーズの演奏や歌唱を行うだけで楽曲を作成できる．しかし，演奏や歌唱の技術がある程度要求され，演奏ミスに対する音高やリズムの編集がわずかには可能であるが [6][7] ，音質劣化を伴うなどの問題により自在にはできない．

一方，MIDIシーケンスデータは発音タイミングや音高などを指示するMIDIイベント情報が記録されているのみであり，実際にどのような音が出力されるかは，いわゆるMIDI音源のような音色ライブラリから選んだ音色によって変えることができる．よって音色，音高，リズムなどほとんどのパラメータについて自由度の高い編集が可能であり，再利用性も高い．そのためWAVE波形データよりも手軽に音楽表現の外在化に用いることができる．しかしながら，その表現自由度の高さゆえにMIDIシーケンスデータの作成は大変手間のかかる作業であり，例えば，生演奏を模したようなデータを作成するには，細かな演奏制御データ作成の手間に加えて，楽器の発音機構や奏法への知識や適切な音色の選択なども必要となる．

特に自作メロディやフレーズを入力したり，いわゆる“耳コピ”と呼ばれる頭で

記憶したメロディを入力したりする場合，つまり楽譜などの音高やリズムが記述された情報がない場合，まだ観念的存在 [8] であるメロディやフレーズから 1 音ずつ自ら音高やリズムを探って判別し，音符として入力する作業が必要となってくる．入力者がよく訓練を受け，絶対音感やリズム感のような比較的高度な音楽的素養や知識を獲得している場合，単旋律なら比較的容易に外在化可能と考えられるが，絶対音感保有者の存在は一般的とは言えない [9] ．ゆえにこの作業は，多くの音楽の知識や能力（特に音感）に乏しい者にとっては，初心者でなくとも労力を要する作業と言える．

この作業は，音感やリズム感を備えた他人にフレーズを歌唱することによって口承で伝えれば，彼らが“耳コピ”して音符情報に変換してくれることも可能である．しかし，常に誰かに代行を依頼するのは不便であるし，現実的ではない．

そこで，これを人間に代わって計算機に行わさせるのが Voice-to-MIDI（鼻歌入力）[10][11][12][13][14][15] [16][17][18][19] である．Voice-to-MIDI システムでは計算機が音符変換を担うため，ユーザは，創造したり記憶しているフレーズをマイクに向かって歌うだけでよい．よって，特に絶対音感や相対音感を持たないユーザや楽器演奏技術の無いユーザにとって非常に有用な入力方法である．Voice-to-MIDI は音楽制作用途の他，Query By Humming（QBH）と呼ばれる楽曲検索のインタフェイス [20][21] などに応用されている．

本研究では，マイクから入力される音響の音高取得のために Voice-to-MIDI 手法を用いる．しかし，そのままでは人間の意図した音の取捨選択が難しいため，リズムについては人間が注目した音区間を切り出せるような仕様とした．

1.2.2 Voice-to-MIDI における音の区切り

自動処理によるセグメンテーション（音の区切り）を行う Voice-to-MIDI において，音の区切り位置を検出する方法には

1. 音の大きさ（振幅）に対する閾値
2. F0（音高）の音高遷移
3. 音楽的特徴量の算出

4. 周波数構造の変化

などが挙げられる．

音の大きさに対する閾値を設ける方法では，閾値以上になったら音の開始点とし，再び閾値以下になったら終了点としていると推定されるものがある [12]．仕組みが簡単であり，導入しやすいのが利点である．一方，使用場所やマイク感度に合わせて事前に閾値を調整する必要があり，屋外のように閾値を頻繁に再調整しなければならない場所では非常に使いづらい．歌詞歌唱では 1 音ごとの区切りが不明確になるため，複数音が 1 音に認識されてしまうなどの問題も起こりやすい．

F0（音高）が音高遷移したこと検出する方法では，入力音の F0 に対して，50cent 以上離れたら別の音とするものがある [22]．また，入力音の F0 に対して 12 音階などの絶対的に固定されたスケールをあてはめ，F0 が現在の音高から別の音高に遷移したときを音の終了点および次の音の開始点にしていると推定されるものもある [7][13]．音の大きさに影響されにくいため，歌詞歌唱などにも対応できる．しかし，同じ F0 の音が複数続くと 1 音と認識されてしまう可能性が高くなる．また，歌唱の場合は F0 が安定しづらいため，音が細かく区切られてしまう可能性があり，ビブラート検出などアルゴリズムに工夫が求められる．

文献 [23] などの音楽的特徴量の算出や周波数構造の変化を用いる方法では，より確度の高い区切りの検出が可能と考えられるが，高度な処理になるほど計算コストの増大が起こりやすい．また，データセットによる学習が必要になることもある．

これらの音区切りの手法は，各々の処理アルゴリズムにおいて基準を満たした入力音響の全てが符号変換処理の対象となり，歌唱の符号化や楽曲検索などの Voice-to-MIDI 本来の用途には適用可能である．しかし，入力される全ての音響が符号化される必要がない用途，例えば歌唱であっても，一部分のみを取り出して符号化したい，というような用途への対応は難しい．

本論文で提案する音区切りの手法は，上記のような一般的な歌唱の符号データ変換用途への適用の他，入力される全ての音響が符号化される必要がない用途にも対応できる．

1.2.3　リズムのリアルタイム表現

阿部は，旋律（フレーズ）の歌唱や聴取における旋律の認識には

1. リズム構造の処理

2. 旋律線構造の処理

3. 調性構造の処理

の過程があり，それぞれの過程がある程度の独立性を保ちながらも相互関連し，統合的な処理がなされている，としている [24] ．フレーズ中の各音の音高が知覚できなかったり，調性が知覚できなかったりする人であってもリズムの知覚は比較的容易であり，同様にリズムを表現することも比較的容易であると考えられることを示す事例があることから，リズムは音楽フレーズを理解する上で最も基礎的な要素であると言える．

リズムが音高や調性よりも表現しやすいことを示す事例として，SongTapper.com[25] が挙げられる．これは，メロディリズムをタップ入力して検索クエリーとすることによって楽曲検索を行うサービスであるが，ここでは検索の際に音高が理解できている必要はない．もしクエリーがリズムではなく音高列の入力であれば，クエリー作成の難易度が格段に上がるであろう．また近年では，エアギター [26][27] というパフォーマンスが認知されている．これは既存楽曲に合わせて，あたかもそこにギターがあるかのように演奏のものまねを行うパフォーマンスである．実際のギター演奏と違って音は音源である楽曲から出ているので，演者は，ギターフレーズに同期してリズムを身体的に表現することに集中すればよく，場合によっては一切の練習をしなくても即パフォーマンスを行うことも可能である．

しかしながら，リズムが表現しやすいのは，リアルタイムに表現が可能な場合に限られるとも言える．大島らは Coloring-in Piano[28] および，それを MIDI シーケンスデータ入力に応用した 2 ステップ MIDI 打ち込み法を提案 [29] している．これらの研究では，MIDI シーケンスデータの入力において，最初に演奏ごとに変化することがない固定的な要素であるフレーズの音高をノンリアルタイムに入力しておくため，ユーザは MIDI キーボードのリアルタイム演奏によって演奏ごとに変化がある変動的な要素であるリズムの表現だけを行えばよい．ここで，リズム

にも音符や休符という離散的な表現手段が存在することを利用して，仮に 2 ステップ打ち込みの順番を逆にして音価をノンリアルタイムに入力し，音高をリアルタイムで表現することを考える．すると，特に楽譜などの情報が存在しないフレーズの場合，入力者が感じたリズムを音符や休符レベルに離散化する能力がなければ，各音の音価を入力する段階で入力した音価が適切かどうかを確認するのが難しいことが分かる．これが音高であれば，楽器を弾いて確認しながら入力していくことも可能である．また，リズムが適切に入力できたとしても，それは楽譜的な整った美しさはあるが，実際の演奏としては揺れがなくノリ [30][31] が機械的 [32] になるであろう．リズムは音楽が持つ構造的な性質上，音符や休符という離散的表現が用いられることは多いが，本来時間的な揺れやノリも含めて表現される方が音楽としては好ましいと思われる．これらのことから，リズムはリアルタイムに表現することに対してより親和性が高いと考えられる．

本論文で提案する音区切りの手法では，人間がリズムをより扱いやすい形態にするため，リズムをリアルタイム入力させて音を区切ることとした．これにより，音の発音開始を予測し，音の開始に合わせて区切るだけではなく，聞こえてきた音に触発を受けた結果を即座に音の区切りに反映させることも可能となる．

1.2.4 人間と計算機との協調処理

何らかの作業行為において，計算機を用いて支援したり，高精度化や効率化を図ることは広義に協調処理と言える．その点において，計算機上における音楽制作である DTM もそれ自体が人間と計算機との協調処理である．しかし，その協調のレベルは高いとは言えない．

音楽制作をよりミクロなレベルで協調するシステムや手法として，半田らのシステム [34] が挙げられる．人間と計算機とが協調して採譜するシステムであり，発音時刻の候補を音楽情景分析器で求めて表示し，人間が音の有無や音高の上行・下行の情報を入力するシステムである．また，1.2.3 節で述べた 2 ステップ MIDI 打ち込み法 [29] は，システムが音高を管理し，人間がリズムを制御するという役割の分担を行うことによって協調を行い，表情豊かな MIDI シーケンスデータの入力を実現している．文献 [35] は，HCI（human-computer interaction）による協調

を行いながら音楽制作を支援するためのフレームワークである．

計算機が人間を支援するシステムでは，人間が計算機に与える入力情報は最低限であるか，人間のより自然な行為から人間が無意識のうちに追加的な情報を得ることが望ましいと考えられる．しかし，人間が陽に多少の追加的な入力情報を与えることによってよりよい結果を得られることがある．つまり，入力として人間が計算機に伝える意味のある情報を増やすことにより，計算機が人間の意図を汲み取りやすくなる．その結果，両者の緊密性や協調性を一層高めることができる．第 2 章で述べる本論文の基盤となるシステムは，人間のタップによる音区切り情報を追加することによって，より協調的な処理を実現している．

1.3 本論文の構成

本論文は，音の時系列の中から人間が「価値がある」と判断した区間についてのみ情報，特に音楽で重要な要素である音高を取得する，人間と計算機との協調的な音高判定技術の構築を行い，従来の自動処理による Voice-to-MIDI では難しい課題への適用を試みることを目的とする．

第 2 章では，第 3 章と第 4 章に先立って，それらの基盤となる，人間のタップによる音区切り情報の入力と計算機の音高抽出を用いた協調的な音高判定技術を提案する．また，提案手法をタップ併用型 Voice-to-MIDI システムとして実装し，評価を行う．このシステムは，Voice-to-MIDI を応用し，マイクからの入力音と同時に，人間がコンピュータや電子楽器のキーボードなどからリアルタイムに音区間の区切りをタップ入力し，計算機はその区間の音高を取得する仕組みを持つ．このシステムは，従来の計算機の自動処理による Voice-to-MIDI システムと比較すると，人間の介入を増やすことによってさらに多くの情報を与えることができるため，より人間と計算機が協調して音符への変換を行うことができる Voice-to-MIDI システムと言える．評価実験では，Voice-to-MIDI 技術の課題であった音高と音数の判定精度向上の課題について評価した．また，楽器経験の有無やタップの有無の変換精度への影響の評価を行った．具体的には，9 名の被験者に歌唱しながらそのフレーズのリズム区切りを入力させ，歌詞歌唱などの任意の発音の歌唱を許容する既存 Voice-to-MIDI システムとの変換精度の比較によって評価を行った．

第 3 章では，歌唱などの声だけでなく，環境に応じて得られるあらゆる音の時系列に自らの心象風景を織り込んで環境の音情景を再構成するための新しいリズム楽器システム“ EnvJamm ”の提案と評価を行う．EnvJamm は，第 2 章で実装したタップ併用型の Voice-to-MIDI システムに対して，検出できる音域を拡大したシステムであり，評価実験では，被験者 1 名による屋外における試用とシステムをよく理解している著者による試用を行い，アンケートによる評価を行った．

第 4 章では，他者が自由に発する非音楽的な音声に対して音を取捨選択する必要がある課題として，認知症患者に対する音楽療法を支援するシステム“MusiCuddle”を提案し，評価を行う．患者が，不安などの心的状態から生じる症状の一つである，何度も同じ言葉を繰り返す行為（「常同言語」と呼ぶ）から抜け出させたり，落ち着かせるために，MusiCuddle は患者の発声から得た音高に応じた音楽フレーズを操作者の指示によって自動演奏し，患者に聴かせるという仕組みを持つ．操作者は，タップ入力によって音高を取得する区間を指定する．音楽フレーズには，Iso-principle に基づいた，患者の精神状態に寄り添えると思われるものなどを使用する．評価は，認知症患者 1 名に対して試用を行うケーススタディの形で行った．

第 5 章では，これらのシステムから得られた結果を基にして，提案手法の有用性，Voice-to-MIDI の応用可能性拡大への寄与を示し，また，人間と計算機との協調処理の観点および知識科学の観点から議論を行う．

第 2章

メロディリズムタップを用いた音数・音高の判定手法

本章では，3 章および 4 章に先立って，それらの基盤となる，人間のタップによる音区切り情報の入力と計算機の音高抽出を用いた協調的な音高判定技術を構築する．これを計算機を用いた音楽制作における MIDI シーケンスデータ入力法の一つである Voice-to-MIDI（鼻歌入力）において，音高ならびに音数について判定精度を向上させる課題に適用するため，歌唱と同時にメロディリズムをタップ入力するタップ併用型 Voice-to-MIDI システムとして実装し，評価を行う．

2.1　はじめに

計算機上での音楽制作は DTM（Desk Top Music）と呼ばれ，MIDI シーケンスデータと呼ばれる自動演奏データや楽器演奏などを録音された WAVE 波形データを組み合わせて楽曲を制作する．

このうち，MIDI シーケンスデータは発音タイミングや音高などを指示する MIDI イベント情報が記録されているのみであり，実際にどのような音が出力されるかは，いわゆる MIDI 音源のような音色ライブラリから選んだ音色によって変えることができる．よって音色，音高，リズムなどほとんどのパラメータについて自由度の高い編集が可能であり，再利用性も高い．そのため WAVE 波形データよりも手軽に扱える．しかしながら，その表現自由度の高さゆえに MIDI シーケンス

データの作成は大変手間のかかる作業であり，特に生演奏を模したようなデータを作成するには，細かな演奏制御データ作成の手間に加えて，楽器の発音機構や奏法への知識や適切な音色の選択なども必要となる．

そのことを示すように，以前より奏法の再現などに関する各種の解説書 [36][37][38][39] が発売されている．また，ピアノ自動演奏データ作成システムの性能を競うコンテスト RENCON[40] が開催されている．RENCON では，将来的なショパンコンクール優勝を目標に，研究・製品を問わず多彩なシステムが，自動作成や入力支援による手間の軽減と演奏の自然さや表現力の両立を目指して競われる．その他，文化的な面から MIDI シーケンスデータ作成の大変さが示された興味深い事例もある．YAMAHA : Vocaloid2[41] シリーズとして「初音ミク」[42] が発売されてまもなく,「弱音ハク」というキャラクターがインターネット上のコミュニティで登場した．これは,「初音ミク」を上手く歌わせることの難しさを「初音ミク」になぞらえて表現したもので，キャラクタライズによる親しみの中にも，期待を込めて購入したものの MIDI シーケンスデータ作成の大変さに直面した様子が伺える．

以上のような状況に対して，MIDI シーケンスデータの入力法や作成法はさまざま提案されている．ここで，入力支援法まで含めて入力法・作成法を概観する．

まず，フレーズやアイディアなどを制作者が直接入力する方法として，リアルタイム入力である，鍵盤楽器などの MIDI 規格対応楽器の打鍵・離鍵などの演奏情報をそのまま記録する方法が挙げられる．そして，ノンリアルタイム入力である，画面上の楽譜やピアノロールなどのグラフィカルエディタにマウスで音符を入力する方法，また，MIDI 楽器で音高や音量（ベロシティ）を入力し，マウスや PC キーボードでリズムを指定するステップ入力と呼ばれる方法がある．これらは多くの MIDI シーケンサや DAW で採用されており，基本的かつ一般的な入力法と言える．しかし，入力の手間を軽減できるなどの工夫が十分になされているとは言えない．そこで，大島らは Coloring-in Piano[28] を MIDI データ入力に応用した 2step MIDI 打ち込み法を提案 [29] している．これは，最初に演奏ごとに変化することがない要素である音高列を入力しておき，次に MIDI キーボードのリアルタイム演奏によって，演奏ごとに変化がある要素であるリズムを入力する方法である．中野らが提案した Vocaloid のための演奏データ入力ツールである Vocalistener[43]

では，作成者が歌唱を与えることによって声質の調整や表情付け，歌詞のアライメントが行われ，簡便に表情豊かなVocaloid演奏データの作成が可能である．

入力を支援する方法として，Internet : Singer Song Writerシリーズ[10]やFinaleシリーズ[11]ではユーザが選択したテンプレートや表情記号を当てはめることによって演奏表情付け支援を行う機能がある．また，MIDIシーケンスデータに対して表情付けを行うソフト[44]などがある．近年では，Steinberg社が制定したVSTiなどの規格に対応したソフトウェア音源が数多く発売されており，中には作成の手間を減らしながら出力音の質も高めるために特定楽器の入力に特化したエディタと音源をセットにした製品[46][47]がある．エディタと音源がセットであるため，特定の音源を鳴らすためだけに簡素化された専用演奏データを作成するだけでよい．その他，物理モデル音源も発音機構に依存した演奏表現が簡易なデータで再現可能となることから[48][49]一種の入力支援と言える．

一方作成法，つまり人間が入力したいフレーズを入力するようなレベルの音楽表現の外在化ではなく，より粗いレベルの音楽表現の外在化に対応した方法として自動作曲[50][51][52][53]や自動編曲・アレンジ支援[12][13]などの方法がある．これらでは通常，ジャンルなどおおよその方向性は人間が決めるが，実際の演奏データを作成するのは計算機である．

近年では，スキャナを用いた楽譜認識[45][54]も実用的になってきている．

MIDIシーケンスデータの編集は自由度が高く便利であるが，作成に大変手間がかかる作業である点について，上記のような入力法や作成法の提案による解決は図られている．しかしながら，フレーズを入力したいという欲求に対して上記の入力法では，入力の際にメロディやフレーズ中の各音の音高やリズムを把握していなければならないということに根本的な問題がある．

特に自作メロディやフレーズを入力したり，いわゆる“耳コピ”と呼ばれる記憶にあるメロディを入力したりする場合，つまり楽譜などの音高やリズムが記述された情報がない場合，まだ観念的存在であるメロディやフレーズから1音ずつ自ら音高やリズムを探って判別し，音符として入力する作業が必要となってくる．入力者がよく訓練を受け，絶対音感やリズム感のような比較的高度な音楽的素養や知識を獲得している場合，単旋律なら比較的容易に外在化可能と考えられるが，絶対音感保有者の存在は一般的とは言えない[9]．ゆえにこの作業は，多くの音楽

の知識や能力（特に音感）に乏しい者にとっては，初心者でなくとも労力を要する作業と言える．

そこで，これを人間に代わって計算機に行わさせるのが Voice-to-MIDI（鼻歌入力）[10][11][12][13][14][15] [16][17][18][19] である．Voice-to-MIDI システムでは計算機が音符変換を担うため，ユーザは，創造したり記憶しているフレーズをマイクに向かって歌うだけでよい．よって，特に絶対音感や相対音感を持たないユーザや楽器演奏技術の無いユーザにとって非常に有用な入力方法である．Voice-to-MIDI は音楽制作用途の他，Query By Humming（QBH）と呼ばれる楽曲検索のインタフェイス [20][21] などに応用されている．

しかしながら，従来の Voice-to-MIDI システムには問題があった．

Voice-to-MIDI システムの処理は，一般に

1. 歌唱区間の検出
2. 1 音毎の区間検出
3. その区間内で短時間 F0 推定を繰り返し，当該区間全体にわたる短時間 F0 の集合を取得
4. その F0 推定情報からの区間音高判定
5. 得られた音高・音長から音符列を作成

という処理段階に分類できる（(1) が明確に存在しなかったり，(2) の区間検出と (3) の短時間 F0 推定と短時間 F0 集合取得の処理順序が前後したりするなど，全てのシステムがこの通りとは限らない）．

この各段階で得られた結果は，いずれも連鎖的に次の処理の結果に影響を与える．例えば，(2) の処理で誤った区間が検出されると，音数が変化するのみならず，(3) の処理で区間内での短時間 F0 の分布も変化し，結果として (4) の処理で誤った区間音高判定が行われてしまう．したがって初期の段階での誤りは，それ以降の段階の誤りにもつながり，最終的に得られる音数や音高の変換結果をきわめて精度の悪いものとしてしまう．これを防ぐためには各段階においてできるだけ高い精度の処理結果を出すことが必要となる．とりわけ，歌唱区間の検知および 1 音

毎の区間検知の精度を上げることは，それ以降の処理段階への波及効果が大きいので，極めて重要である．

ところが，歌唱区間や1音毎の区間を計算機処理によって検知することは容易とは言えない．人間が次の音に遷移したと思っていても，計算機がその遷移を捉えきれず，前の音とつながった1音として認識されたり，逆に人間がまだ次の音に遷移させてないと思っていても，計算機が過剰に反応し，1音が複数音に分割されてしまったりする．各音の区間誤検知が発生すると，リズムの誤変換だけではなく，音高の誤変換にもつながる．また，意図しないノイズに反応してしまい誤変換されてしまうこともある．

このため，いくつかの Voice-to-MIDI システム [10][12] では,「タタタ〜タタ」のように全ての歌詞を「タ」に置き換えて明確に区切って歌う「タタタ歌唱」のような，特殊な歌唱方法が求められる．これにより区間誤検知が減り，一定の水準の処理結果が得られるようになる．しかし，たとえば初めに歌詞を作ってからメロディを作曲する「歌詞先作曲」[55][56] の場合，歌詞の持つイントネーションなどがメロディに大きく影響するため，歌詞をそのまま歌唱することが不可欠であり,「タタタ歌唱」は使用できない．よって，歌唱スタイルを制限せず，任意のスタイルの歌唱によって MIDI シーケンスデータを入力することができる Voice-to-MIDI システムの実現が求められる．

また，区切りを分かりやすく歌って区切るのではなく，音高の変化によって音を区切るシステム [13][22] もある．この方式であれば，歌詞歌唱にも対応可能である．しかし，同一音高の連続箇所が適切に区切れない可能性がある．また通常，歌唱の音高はピアノなどの楽器音と比較して不安定であり，音高遷移に伴ってオーバーシュート [57] などの意図的ではない音高の揺れが発生してしまう．よって1音歌唱する間に複数の音高を遷移すると区間誤検知が発生してしまう．歌唱中のビブラートなどの意図的な表情付けも含めて不安定さが歌唱における「人間らしさ」にもつながっている [58] とも言えるが，歌唱音高の不安定さは，Voice-to-MIDI では音高判定精度の低下を起こす要因となりうる．そのために区間誤検知に対処できる Voice-to-MIDI システムの実現が求められる．

Voice-to-MIDI は，MIDI シーケンスデータの入力に非常に有用な手法であるにも関わらず，まだ上記のような問題が十分に解決されているとは言えない．

そこで，この問題に対応できるVoice-to-MIDI変換手法の実現に向けて，人間のタップによる音区切り情報の入力と計算機の音高抽出を用いた計算機との協調的な音数・音高判定手法を提案する．次に，提案手法の実装システム（タップ併用型Voice-to-MIDIシステム：以下TVM）と，歌詞歌唱などの任意発音の歌唱を許容する既存のVoice-to-MIDIシステムとで従来からの課題であった音数・音高の判定精度について比較する．

2.2 先行研究

文献[59][60]では音声認識のために，本研究と同様に発声に併せたタッピングなどによる区切り情報入力を行っている．これらにより音節区切り情報の効果は示されているが，Voice-to-MIDIの用途には各区間の音高判定処理が必要となる．

人間と計算機が協調して採譜するシステムとして，半田らは発音時刻の候補を音楽情景分析器で求めて表示し，人間が音の有無や音高の上行・下行の情報を入力するシステム[34]を提案した．しかし，視覚情報による協調であり，聴覚情報を用いた本研究とは協調の方法が異なる．

声とマウスなどのデバイスを併用したMIDIデータの入力インタフェイスとして，ボイスコマンドを用いてMIDIデータを入力するシステムが提案されている[61]．しかしながら，マウスなどで行う操作を声で代行するにとどまるため，特に楽譜情報がない場合の音高やリズム，音価の入力には，音楽的な知識が必要となる．また商品[62]に搭載されているStep Entryモードでは，声から音高を取得する間に音価をマウス入力可能である．しかし，これは本質的にはステップ入力であり，リズムや音価を把握する必要がある．

Voice-to-MIDIの精度向上に関して，文献[63][22][64]では音程の外れた歌唱にも対応可能な手法について述べており，発声した個々の音が絶対音高から外れていても，相対音高としてはスケールを構成していることを利用して，補正を行うことが提案されている．また文献[12]のVoice-to-MIDIシステムでは，スケール上の音に優先して認識されるように重み付けを行うことが可能である．文献[65]では，突然大きく跳躍するような音は誤認識と判断し補正を行っている．これらの音高認識結果の補正手法は，TVMと組み合わせることによってさらに高精度な

Voice-to-MIDI システムを実現することが可能と考えられる.

2.2.1 既存 Voice-to-MIDI システムの問題点

既存の Voice-to-MIDI システムに歌詞歌唱を入力したときの問題点を示す. 市販の Voice-to-MIDI システムに童謡「赤とんぼ」(野ばら社刊「童謡」の変ホ長調版 [66] を使用: 図 2.1) を歌詞歌唱入力した結果を 2 例示す.

図 2.2 にタタタ歌唱入力を前提とするある市販システムにおける「(ゆうやけこや) けーのあかとんぼ」部分の変換結果を示す. 上段は入力された歌詞歌唱の音声波形を, 中段は音区切りの比較のために正解のメロディラインを手動入力したもの (正解データ), 下段はシステムによる認識結果をピアノロールで示す. このシステムは主に音量変化で音が区切られると推測されるが, 本来 1 音であるのに複数の音に認識されてしまったり, 逆に複数音存在する箇所が 1 音と認識されてしまう箇所が多数ある.

図 2.3 は, 別のシステムによる「おわれてみた」部分の変換結果である. このシステムでは主に音高変化によって音が区切られると推測されるが, 意図しない音高の変化にも反応してしまい,「お」と「て」の部分で余計な音が出力されてしまっている.

このように, 従来の Voice-to-MIDI システムは歌唱音声データを適切に 1 音ずつに区切れず, その結果個々の音の音高や音長の誤認識が起こっていると言える.

総じて, 以下のような箇所や条件において区切りミスがみられた.

- 同一音高の連続
- 激しい音量変化
- 大きい音高変動
- 不十分な音高変動
- 歌詞 (任意発音) 歌唱
- 環境音の誤入力

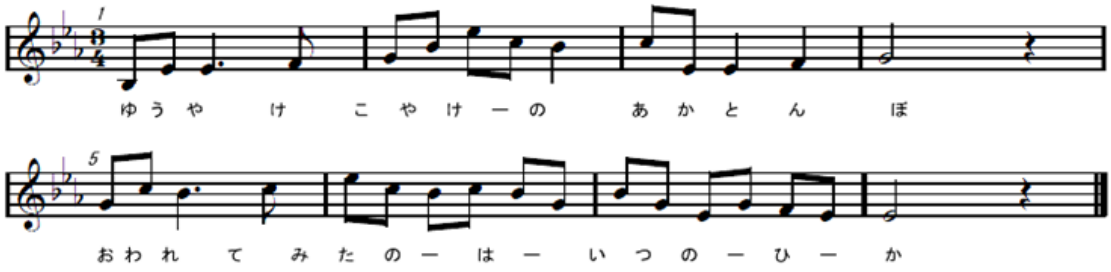

図 2.1: 赤とんぼの楽譜 作曲：山田耕作，作詞：三木露風

2.3 タップ併用型 Voice-to-MIDI システム

2.3.1 タップ併用型 Voice-to-MIDI（TVM）手法の概要

「はじめに」で述べたような問題に対処するためには，音量変化が乏しくて音が区切られない問題や音高変化などによる意図しない区切れの発生の抑止，不要区間の除去が必要となる．そこで TVM では，計算機が苦手とするが人間にとっては容易な音区間の区切りを人間が担当し，計算機は得意だが人間が苦手としやすい F0 推定を計算機が担当する，人間と計算機との協調的な処理機構を採用した．

具体的には，ユーザは，歌唱するメロディのリズムに併せて鍵盤楽器や PC キーボードなどのデバイスをタッピングし，メロディの各音を区切る情報（リズム区切り情報）を入力してゆく．それと同時にシステムは，歌唱から音高，リズム区切り情報からリズムと音長を取得し，最終的にマージして出力する（図 2.4）．

2.3.2 プロトタイプシステムの構成

上記の処理を実装したタップ併用型 Voice-to-MIDI システムのプロトタイプシステム（以下 TVM プロトタイプシステム）について述べる．入力は音声波形とリズム区切り情報，出力は D2-F5 までの半音単位の音高（A4 = 440Hz を基準とする）を持った MIDI データである．入力音声は 22050Hz，16bit, モノラルでサンプリングされる．リズム区切り情報には MIDI キーボードや PC キーボードの打鍵および離鍵の入力時刻情報を用いる．PC キーボードの場合は，タップに「,」および

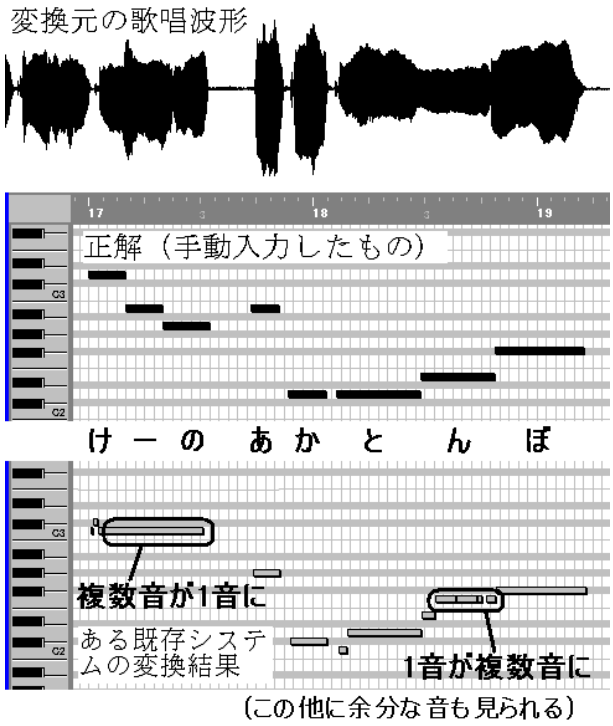

図 2.2: 音量によって区切られたと推測される，複数音が 1 音に，1 音が複数音に変換された例（赤とんぼの「けーのあかとんぼ」）

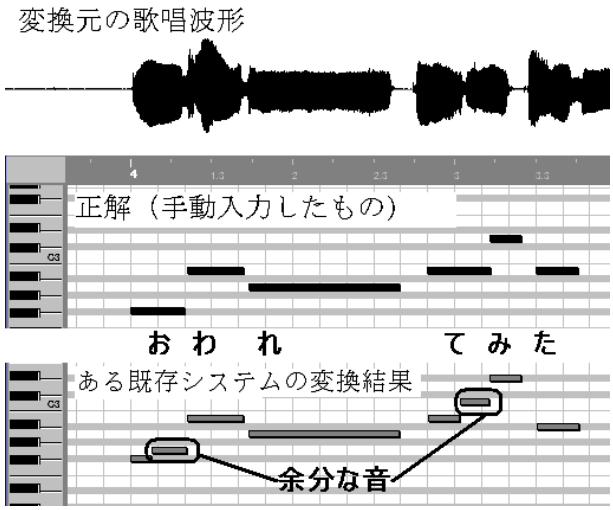

図 2.3: 音高変化によって区切られたと推測される，余分な音が出力された例（赤とんぼの「おわれてみた」）

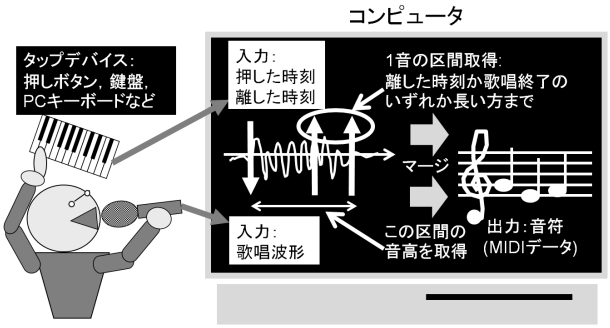

図 2.4: タップ併用型 Voice-to-MIDI の概要

「.」の 2 キーを使用し，1 キーのみ連打しても 2 キーを交互に打鍵してもよい仕様とした．以下に 1 音毎の区間検知と，各区間における音高判定の処理手順を示す.

1. キーが押下され，システムに押鍵情報が入力されたら，これをトリガーとしてマイクより入力される歌唱音声データに対して，後述する F0 推定処理を開始する．

2. キーが離されたら，その離鍵情報が入力された時点か，歌唱の途切れが検知された時点（これは後述する無発声検知機構によって決定される）の，いずれか時間的に後の方が 1 音の区間の終了となる．タップ開始から区間の終了までを音長として，その区間内で F0 推定処理を繰り返す.

3. 1 音の区間終了後，F0 時系列データから半音単位のヒストグラムを生成し，最頻音高の音名を求め，これをこの区間の音高として出力する．

   F0 推定は，入力波形に対する短時間フーリエ変換 (STFT，フレームサイズ =2048samples：約 100ms，フレーム移動間隔=128samples：約 6ms) から求めたパワースペクトルの D2-F5 相当の周波数間に存在するピークのうち，このパワースペクトルに対する IFFT から求めた循環自己相関の正の最大値近傍の周波数のものを用いる．更にスペクトルの内挿 [67] を用いて cent 単位で音高推定して F0 推定結果として出力する．これは周波数解像度不足を補うためである．

本システムでは，タップ開始時刻について，区切り情報と波形の同期が必要となる．PC キーボードのキーを叩いたときの Keypress イベントの時刻と打鍵音（パル

ス音）の録音時刻とのずれを調査したところ，試作システムでは，概ね1024sample（約50ms）分Keypressよりも遅れて録音されたため，1024sample分調整して同期精度を高めた．

2.3.3 無発声検知機構

予備実験において被験者のタップ方法を観察したところ概ね2通りとなった．1つは，1音の歌唱終了までキーを押下し続けるタップであり（図2.5のタップ法1），もう1つは，押下してすぐ離してしまうようなタップである（図2.5のタップ法2）．

過去に実施した実験[68][69]では，タップ法1のみに対応したシステムを用いたが，タップした時間がそのまま音長になるため，タップ法2が行われたときに音長が極端に短くなったり，十分な量のF0推定情報が取得できなくなる問題がみられた．そこで，歌唱区間の途切れを検知する機構によって，たとえタップが早期に終わってもそこで歌唱終了とみなされないようにした．

具体的には，循環自己相関の結果，タップ終了後でもD2-F5の音高範囲内に最大の正相関値が存在する限りフレーム移動間隔約6ms分区間が順次延長され，なくなれば歌唱の終了と判断するようにした．

この機構により，音長は，タップ終了と歌唱終了のタイミングで以下の3パターンに定められる．

1. タップ終了後に歌唱終了：歌唱終了時点
2. 歌唱終了後にタップ終了：タップ終了時点
3. 歌唱が終了しないまま次のタップ開始：次のタップ開始直前

ただし，タップ開始から200ms未満までは遅れて歌唱開始されても歌唱終了を誤って検知されないようにした．タップ開始時に歌唱がない場合，即座に歌唱が終了したとシステムが誤検知してしまうと，パターン（2）が適用されて，歌唱の有無に関わらず，必ずタップ終了時点までが1区間になってしまう．これを防ぐためである．

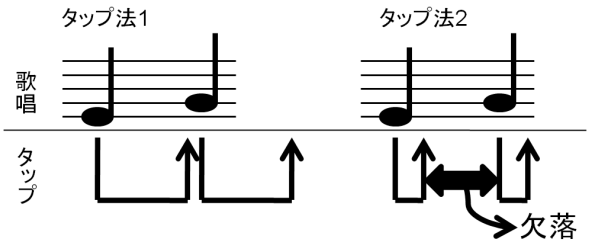

図 2.5: 2 種類のタップ方法

200ms 未満という値は，著者自身がどれぐらいまで自然に歌唱とタップをずらしうるかを実験で調査して経験的に得た値を基に，システムに慣れないユーザを考慮して余裕を持たせた値である．

また，F0 推定が上手くいかず，音があるのに音高範囲内に F0 が無いと判定されることを想定し，音量（パワースペクトルの合計値）が直前の FFT フレームの音量の 90 %以上であれば終了しない仕様とした．

この無発声検知機構によって，対象とする音高範囲内に他に目立つ音がなければ，音量閾値などの手法を用いずに有音／無音を判別可能となり，周期性がはっきりとした音が存在していなければ環境音の音量変化への動的対応や小音量下でも判別が可能となるなどのメリットがある．一方でこの手法では，タップ終了後でも，歌唱以外の音に反応したことによって範囲内に最大の正相関値が出現していれば消音されない可能性がある．しかし，著者自身が実使用において想定しているマイクである，比較的感度が低い PC 内蔵マイクやヘッドセットマイクなどのマイクで調査したところ，歌唱終了と推定できる位置から大きく外れることなく 1 音の区間が終了した．

2.3.4 TVM プロトタイプシステムの仕様的限界

TVM プロトタイプシステムが仕様として対応できる音域およびテンポ（タップ速度）の限界について述べる．

音域については，ポップス楽曲を想定し，A4 = 440Hz を基準として，下限を D2，上限を F5 とした．これは，おおよそバス歌手〜アルト歌手の音域に相当する（文献 [70]）．メゾソプラノやソプラノの音域には対応していないが，ポップス

等でよく使われる音域に対しては十分と考える．

テンポについては，FFT フレーム移動間隔が約 6ms なので，この間隔を 16 分音符とし，人間が 6ms 毎にタップできると仮定すれば，無発声検知機構の「歌唱が終了しないまま次のタップ開始」のパターンによって原理的には BPM=2500 程度まで対応できる．しかし実際の入力では，それほど早く歌唱やタップをすることはなく，BPM=250 程度まででよいと思われるため，本プロトタイプシステムは十分対応している．

2.4 評価実験

2.4.1 実験概要

提案手法の検証のため，前章で述べた TVM プロトタイプシステムを用いて，歌唱音声に対する音区切り（音数）と各区間の音高判定精度を評価するとともに，楽器経験のタップへの影響およびタップの有無の歌唱への影響を調査した．

なお，この実験の評価対象は，システム自体の性能であり，入力者の歌唱やタップの技術に依存する内容については評価の対象とせず，極力排除した．例えば，歌が下手で楽譜通りの変換結果にならなかったとしても，それだけではシステム自体の性能の良し悪しは言えない．この場合，楽譜通りの歌唱かどうかではなく，実際の歌唱の音高を割り出し，それとシステムの変換結果との比較を行うことによってシステム自体の性能の良し悪しが分かる．

また，システムの仕様として対応可能な音域やテンポ（タップ速度）の限界については 2.3.4 節「TVM プロトタイプシステムの仕様的限界」に記した．

評価では，TVM と同様に歌詞歌唱などの自由な発音による入力を許容し，歌唱スタイルを制限しない入力により近いと思われるシステムを比較に用いた．

評価項目は以下とした．

1. 任意発音歌唱に対して性能が向上したか？
2. 歌唱同期タップが可能であるか？
3. 評価項目 2 において楽器経験の影響はあるか？

4. タップの歌唱への影響はあるか？

（1）と（2）については，後述する 2.5.1〜2.5.3 節で曲および歌唱条件ごとに評価し,（3）は 2.5.4 節で TVM の結果を用いて楽器経験の影響について評価する.（4）については 2.5.5 節で比較 3 システムのタップあり歌唱の処理結果とタップなし歌唱の処理結果とを比較する.

2.4.2 楽曲

歌唱する楽曲は以下の 2 種類である.

1. 課題曲（赤とんぼ）
2. 各被験者が選んだ自由曲（歌詞のあるメロディを 1 コーラス程度）

赤とんぼは，音高の範囲が広く変化も激しいが，一方で同一音高が連続する箇所もあり，適度に難しい．そしてよく知られている曲であることから課題曲に採用した．歌唱テンポによって大きく 2 種類の歌唱条件を設定し,「テンポ自由」では，被験者の好みのテンポで歌唱させた．また，赤とんぼは通常遅いテンポで歌唱されるため,「BPM=120」で歌唱させ，速いテンポでも歌唱とタップの同期が可能かを検証した.

自由曲では，赤とんぼよりもリズムや音高変化が複雑でより実践的な曲への対応が可能かを検証するために，各被験者自身が選曲したポップスなどのメロディを歌唱させた.

2.4.3 比較に用いた Voice-to-MIDI システム

比較に用いた Voice-to-MIDI システムは，3 種類である.

1. CMP:音高変化に基づいて区切る先行研究システム
2. RYN:先行研究のシステム [23]
3. BP2:商用で市販されているシステム [13]

CMP は，この実験を行うにあたって音区切りの手動・自動の比較のために著者が作成した．F0 推定法などは TVM と同様とし，タップによる区切りの代わりに音高の変化で区切る．音高を区切る基準については，文献 [22] を参考に 50cent 以上の差があるときとした．無発声検知機構は，判定精度が低下したので実装しなかった．また，約 70ms 以上の音長のみ変換するようにした．これは予備調査により，速いテンポへの対応，できるだけ多い認識音数，不要な音の誤変換の少なさのバランスを考慮した値である．16 分音符換算で BPM=213 程度までの歌唱テンポに対応可能である．

RYN は，先行研究との比較のため用いた．文献 [23] の著者らからシステムの Linux バイナリの提供を受け，そのまま使用した．これは楽曲中からメロディーライン等を抽出し，MIDI データへの変換を行うシステムであり，文献 [17] 等，Ryynanen らが保有する技術を応用して構築されたシステムである．音の区切りは，“ Accent Signal ”と呼ばれる FFT フレーム中のスペクトルエネルギーの量を用いて行っている．

BP2 は，KAWAI: Band Producer 2 に付属の鼻歌入力機能である．この機能は，予め設定した音量閾値を超過したときと半音単位の音高閾値を超えたときに音符が区切られる仕様であると，変換結果から推測される．音高変化があれば区切られるため，歌詞歌唱にも対応していると考えられる．

2.4.4 機材設定

TVM においてタップに用いたデバイスは，HP: 2710p ノート PC のキー「,」および「.」である．これらのキーは隣接して存在し，被験者はこれらのキーの両方あるいは片方のみを好みに応じて用いる．また，歌唱収録用マイクは Shure: SM87A を用いた．

次に各種情報の記録および処理手順について述べる．2 台の PC を用意し，PC1 では，被験者に試唱させて BP2 の録音音量閾値を設定した後，BP2 に伴奏なし歌唱をリアルタイムで入力し，MIDI データに変換する．同時にその歌唱は Wave 波形として BP2 上で記録される．

PC2（2710p ノート PC）では，TVM のために，歌唱と同時に行ったタップ区

切りの情報を自作ソフトで記録する．このタップ情報と PC1（BP2）で記録した波形とを組み合わせてオフライン処理で MIDI データに変換する．実験では全システムで完全に同じ歌唱波形を使用するために便宜上，本来オンライン処理である TVM をオフライン処理にした．

また，PC1 で記録した歌唱波形と TVM のタップ情報の同期が必要となるが，PC2 で歌唱波形をタップと同期させて記録しており，その波形と PC1 の波形を目視して同期位置を探した．具体的には，PC1 と PC2 の両波形に共通する特徴的な形状の箇所を複数探し，それらの箇所の間隔が両波形で一致するかを評価して同期位置を決定した．なお相互相関などで自動同期推定を行っても，最終的に目視による確認が必要であると考えて自動処理は行わなかった．CMP と RYN は，いずれも BP2 で取得した波形を，必要があれば Adobe: Audition 1.0 で対応サンプリングフォーマットに変換した後，オフライン処理した．

2.4.5 被験者

被験者は，筆者らが所属する大学院の男子学生 8 名と女子学生 1 名である．TVM の支援対象は，主に音感を持たないユーザであるが，実験では様々なデータを得るために和音楽器やリズム楽器の経験者，音感があると思われる学生にも参加をお願いした．

どのような被験者が参加したかの傾向を知るために，予備調査により被験者の音楽知識や能力，楽器経験を調べた．項目を以下に示す．

1. 「鍵の音名」: ピアノ上で指差された鍵を見て音名を回答
2. 「音高聴取」: ピアノで弾かれた単音の音名を回答
3. 「音の高低」: ピアノで弾かれた 2 音の高低を回答

各項目はいずれも全 6 問ある.「鍵の音名」では基礎的知識,「音の高低」では基礎的な知覚能力,「音高聴取」では高度な学習経験・技能を調査した．実験では，被験者は最低限歌唱が可能であればよく（タップは，全くできないようなレベルでなければ問題ない），被験者 9 名が歌唱に問題がないことは確認している．

表 2.1: 各被験者の予備調査項目 1〜3 の正解数と楽器経験

被験者	音名	音高聴取（正解）	音高聴取（半音差）	音の高低	楽器経験
A	6	0	1	5	なし
B	3	0	0	2	なし
C	6	1	0	5	なし
D	3	1	0	6	なし
E	0	1	0	6	太鼓, ムックリ 1 カ月
F	5	0	0	5	和太鼓 2-3 年
G	6	0	0	6	電子オルガン 2 年
H	6	0	4	6	電子オルガン 3 年, ピアノ 5 年
I	6	5	1	6	ピアノ 10 年以上

注 1. 被験者 A〜D は「楽器経験なし」と回答した被験者

注 2. 予備調査項目（2）「音高聴取」は，正解した個数と正解から半音差だった個数を示す.

これらの結果より，楽器経験なし 4 名と経験あり 5 名に分類した．各被験者の正解数と楽器経験を表 2.1 に示す．表 2.1 より，安定した歌唱が可能と考えられる「音高聴取」の成績がよい被験者がいる一方で，Voice-to-MIDI の支援対象となりうる，基礎的な「鍵の音名」や「音の高低」の正解数が少ない比較的音楽に詳しくない被験者も含まれており，経験の有無だけでは測れない様々なレベルの被験者がいることが分かる.

表 2.2: 各曲の歌唱条件

（A）赤とんぼ

テンポ	タップ
自由	あり
	なし（BP2 のみ使用）
BPM = 120	あり
	なし（BP2 のみ使用）

（B）自由曲

テンポ	タップ
自由	あり

2.4.6 実験手順

実験は大学院内の防音室を用いて 1 名ずつ行った．まず Voice-to-MIDI の練習および歌唱しながらタッピングする練習を 5 分ずつ行った後，以下の順序で実施した．最初に被験者に課題曲の童謡「赤とんぼ」の 1 番（全 31 音符: 図 2.1 参照）を，歌詞を見ながら 3 回聴取させ，メロディをできるだけ覚えるように指示し，

1. 赤とんぼ : テンポ自由
2. 赤とんぼ : BPM=120
3. 自由曲

の順に歌唱させた．各曲の歌唱条件を表 2.2 に示す．課題曲ではタップありなしをランダムな順番で指示して歌唱させた．赤とんぼについては，それぞれ 3 回ずつ歌唱を入力させた.「BPM=120」で歌唱する場合は，メトロノームに合わせて歌唱するよう依頼した．自由曲については，被験者の負担を考えて 1 コーラス程度を 1 回歌唱させた．各被験者の自由曲を表 2.3 に示す．実験は全て歌詞歌唱（途中で歌詞が分からなくなった場合は適当な発音でもよい）で行い，実験中は，歌詞

表 2.3: 各被験者の自由曲

被験者	歌手名	曲名
A	Mr. Children	Over
B	井上あずみ	さんぽ
C	フォーククルセダース	11 月 3 日
D	スピッツ	チェリー
E	Acid Black Cherry	愛してない
F	ブルームオブユース	ラストツアー
G	チャーリー・コーセイ	ルパン三世　その 1
H	SMAP	世界で一つだけの花
I	高橋洋子	残酷な天使のテーゼ

カードは見てもよいが楽譜は一切呈示しなかった．また，全ての歌唱は無伴奏で行った．

2.4.7　評価方法

被験者が必ずしも楽譜通り，あるいはそれを移調した音高通りに歌唱できたとは限らない．ゆえに正しく各システムの音高判定性能を評価するために，楽譜上に記載されている音高ではなく，実際に歌唱された音高から正解の音高データを作成した．BP2 で記録した実験中の歌唱音響波形から，著者自身[1]が 1 音毎に音高の特定を行った．また，正解の音高データと各システムの出力結果との時間同期や欠落音などの判定のために発音開始時刻と終了時刻の特定も同時に行った．これらを「正解データ」とした．作成された音列は必ずしも楽譜通りの音高列とはならないが，被験者の歌唱誤りをシステムの誤りとみなしてしまうことを回避し，純粋にシステムの性能を評価できる．

[1]高校時代に男性合唱部に 3 年間所属した経験があり，また単音の音高を判定できる程度の絶対音感を保有している．

歌唱からの音高および発音開始時刻と終了時刻の特定の方法（正解データの求め方）は以下の通りである．

1. 各音のおおよその区切りを試聴や波形の目測で割り出し，発音開始時刻および終了時刻とする．

2. 波形編集ソフト（Adobe: Audition1.0）上で各音の発音開始〜終了までをループ再生させながら，ピッチベンドホイールつきのキーボード（Ensoniq: MR-76）を同時発音してうなりを聴き，音高特定を試みる．

3. 1音中で音高変化がある場合は，2〜4箇所程度の区間に分けて（歌い始め直後と歌い終わり付近は除く），局所的に音高特定を行う．

4. 適宜波形編集ソフト上で目視計測した1波長の時間から周波数を逆算して用いた．

あまりにも音高の変化が大きい音や音高の特定が困難な音は評価から除外した．この作業により各音を，

1. 音高が一意に決まる音

2. 2音高の間で決めがたい音

3. 分類（2）よりも明確に音高が変化する音

の3種類に分類した．また，(2）と（3）に分類される音は，可能性のある音すべてを正解データとみなした．正解音高は1音につき1音高に定まるのが最良だが，音高のゆれが大きい場合など，1音中でどの音高が優勢であるかを割り出すのは困難であるため，候補全てを正解とした．

なお，2音から生じるうなりがなくなる周波数は客観的に一意に決まるため，作業者の違いによる正解データの大きな違いは生じにくいと考えられ，よって作業者が1名であることは妥当性を有すると考える．

次に個々の音について正解データと認識結果とを対応づけ，両者の音高を比較して正否を判定した．分類（2），(3）に該当する音との比較では，複数ある正解

表 2.4: 認識結果の分類

カテゴリ	サブカテゴリ	説明
正解音	—	正解と一致した音
誤り音	—	正解と一致しなかった音
	全数	誤り音の全体数
	結合音による誤り音	他の音との結合で生じた誤り音
欠落音	—	欠落した音
	全数	欠落音の全体数
	結合音による欠落音	他の音との結合で生じた欠落音数
余分音	—	余分な音

データのうちいずれかの音高と一致すれば正解とした．最終的に表 2.4 のように分類された．

「結合音」とは，正しく区切られずに前後の音と結合した音を意味する．結合音の区間に一致する正解音列と比較したとき，先頭の音と結合音の音高が一致すれば結合音は「正解音」，不一致ならば「結合音による誤り音」に分類される．そして，残りの音は「結合音による欠落音」となる．

「誤り音」は，誤り音の全体数と，結合音によって生じた誤り音数に分けて示す．誤り音の全数と結合音による誤り音の差分は，F0 推定のミスによる誤り音数と考えてよい．

「欠落音」は，出力されなかった音の全体数と，結合音によって生じた欠落音数に分けて示す．これらの音数の差分は，そもそもシステムが認識しなかった音数となる．

「余分音」は，本来 1 音だが複数音に認識されたとき，必要な 1 音分を除いた残りの音，そして歌唱中における咳などのノイズである．1 音分については，複数音のいずれかの音が正解と一致すれば正解音，全くなければ誤り音に加算される．

各メロディの全歌唱音数（赤とんぼの場合正しく歌唱されれば 31 音）は，以下の式のように（1）～（3）の合計で求まる．

全歌唱音数（音）＝ 正解音数+誤り音数+欠落音数

最後に上記の分類結果を用いて変換精度を求める．例えば，正しく音高が変換された音数は多いが余分な音も多く出力された場合，よいシステムとは言い難い．そこで，歌唱された音数に対して正しく音高が変換された音数の割合を測る再現率，およびシステムが認識した全音数に対して正しく音高が変換された音数の割合を測る適合率の 2 つの尺度で評価する．また再現率と適合率を総合して評価する指標として F 値も求める．それぞれ以下の計算で求められる．

1. 再現率（%）＝ 正解音数 ／ 全歌唱音数*100
2. 適合率（%）＝ 正解音数 ／（正解音数+誤り音数+余分音数）*100
3. F 値 =（2*再現率*適合率）／（再現率+適合率）

2.5 評価実験結果および考察

評価実験結果および考察について述べる．

2.5.1 赤とんぼ:テンポ自由

「テンポ自由，歌詞歌唱，タップあり」の歌唱条件による入力 3 回分計 93 音について被験者ごとに集計を行った結果を表 2.7 に示す．

いずれの被験者とも TVM が最もよい再現率・適合率・F 値であった．5 名が再現率・適合率ともに 100 %であり，欠落音・余分音が十分抑制されていることが分かる．誤り音については，全て F0 推定ミスが原因であった．過不足のないタップによって欠落音・余分音が共に抑制され，タップ位置の大きなズレによる誤り音の発生もほとんど見られなかったことから，音数の切り出しや音高の判定に必要な歌唱同期タップができていると言える．

CMP・RYN・BP2 のいずれも正解音数自体は比較的多いが，TVM より欠落音が多く，また，欠落音中の結合音が，CMP（95 音中 58 音）と RYN（42 音中 23

音）では半数以上を占めた．赤とんぼでは同一音高の連続箇所が楽譜上 4 箇所存在しており，それらがロングトーンに誤変換されやすいことが影響したと見られる．

CMP・RYN・BP2 は，余分音も多かった．余分音が多い原因は歌唱中の音高変動や揺れが多いためである．例えば 3 小節目の「あか」のような落差の大きい箇所では，音高が大幅なアンダーシュートを起こし，本来の音高に戻るまでに複数の音高に掛かる．また 3-4 小節にかけての「とーんーぼー」のようなロングトーンは意図しない音高変動が起きやすい．

総じて，TVM は欠落音や余分音等の問題を解決し，任意発音歌唱に対して高い性能を実現可能と言える．

2.5.2 赤とんぼ:テンポ BPM = 120

「テンポ BPM = 120，歌詞歌唱，タップあり」の歌唱条件による入力 3 回分計 93 音について被験者ごとに集計を行った結果を表 2.8 に示す．

全体傾向としては，自由テンポ時よりも正解音数が減少が見られる．変化がないように見える RYN についても，正解音数に極端に差がある被験者 E を除くと減少している．

TVM では歌唱テンポの上昇に伴い負荷が高まるとともに誤り・欠落・余分の各音数も自由テンポ時より増加しているが，これは妥当な結果と言える．中でも被験者 E は欠落音・余分音が大きく増加しているが，音長をある程度保ったタップ間隔ではなく，区切るべき箇所から全く外れた音の途中でタップされた例が見られたことから，テンポが速く追いつかなかったというよりもタップするべき位置を把握できずに混乱したと見られる．しかし，全体では比較 3 システムよりも欠落音・余分音が十分に抑制されており，テンポが速くなっても音の切り出しや音高判定に必要なタップが可能な被験者が多いことが分かった．

比較 3 システムについては，余分音が自由テンポ時よりも減少している点が特徴として挙げられる．これは，テンポが速くなると 1 音当たりの歌唱時間が短くなり音高変動が減るためと考えられる．

総じて，タップ位置のミスが音高判定精度を落とすのは TVM の性質上避けがたく，テンポ自由時よりは多少劣るものの，再現率・適合率・F 値いずれもほとん

どの被験者についてTVMが高い結果となり，特に2名において再現率・適合率ともに100％であったことから任意発音歌唱に対して性能が向上したと言える．

2.5.3　自由曲

各被験者が選択した自由曲について「テンポ自由，歌詞歌唱，タップあり」で入力した結果を表2.9に示す．表2.9より，合計値ではTVMが比較3システムよりも再現率・F値のほとんどにおいて上回り，総合的にみるとTVMは,「タップしながら歌唱する」という負荷の高さにも関わらず，より実践的なポップスなどのメロディの入力においても高い音数・音高判定が実現可能であることが分かる．

ただし，被験者A, E, Fは，1音ごとに正しくタップされなかったため結合音が多い．そして，A, Fは結合音に起因する誤り音も多い．TVMでは，結合音の音高は，結合音区間に含まれる音のうち，もっとも頻度の高い音高が採用される．また同一音高の連続箇所に限らずタップ区切りをしなければ結合音が発生するため誤り音と結合音が同時に発生しやすくなる．よって再現率あるいは適合率の精度低下が見られた．

しかしF値で評価したところ，各被験者ともTVMが高いかあるいは同等となったため，TVMはより良好な性能を達成していると言える．

A，E，F以外の被験者における誤りの発生原因は，タップ開始位置のズレにより音区切りがうまくいかなかったことにあると考えられる．テンポが速く追いつかなかったと想像される箇所と，タップするべき位置を把握できずに混乱したと想像される箇所がともに存在した．しかしながら，各被験者とも非常に高いと思われる負荷にも関わらず高い再現率を達成していることから,「タップしながら歌唱する」行為は，基本的に実施可能なものであったと言える．

2.5.4　楽器経験の有無のタップへの影響

提案手法（TVM）に必要なタップの能力が，楽器経験に影響されるかを評価した．まず楽器未経験者A〜Dおよび経験者F〜Iの2群に分けて，課題曲のTVMの結果比較を行う．被験者Eは楽器経験はあるがごく短く，どちらの群が妥当か判断し難いので除いた．

テンポ自由歌唱では，楽器未経験者は再現率 98.7 %, 適合率 98.7 %, 経験者は同 99.7 %, 99.7 %であった．これについて楽器未経験者と経験者の再現率および適合率について t 検定を行ったところ，どちらも有意な差は見られなかった．また，再現率・適合率ともに 100 %の被験者が 5 名いたが，未経験者も含まれており，このレベルの曲や歌唱条件に対しては楽器経験の有無は影響を及ぼしにくいと見られる．

BPM=120 の歌唱では，未経験者は再現率 97.8 %, 適合率 96.6 %, 経験者は同 98.1 %, 98.1 %であった．これについても楽器未経験者と経験者の再現率および適合率について t 検定を行ったところ，どちらも有意な差は見られなかった．また，再現率・適合率ともに 100 %の被験者が 2 名いたが，1 名が未経験者であった．これらより多少速いテンポの入力であっても楽器経験の有無は影響を及ぼしにくいと考えられる．

次に課題曲の TVM の結果について表 2.1 の予備調査の結果も交えて評価した．

まず，表 1 の全 4 項目 (音高聴取の結果は合計して使用) と全被験者のテンポ自由歌唱の正解音数とを重回帰分析した．楽器経験については，楽器経験があれば通常，リズムの知識や練習経験があると考えられるため，楽器に関係なく年数をそのまま用いることとした．複数の楽器経験がある場合は長い方を，範囲による回答の場合は長い方，1 年未満のものは月数を 12ヶ月で割った値を用いた．その結果，求められた重回帰式に有意性は認められなかった．

同様に BPM=120 の歌唱についても，重回帰式には有意性が認められなかった．これらの結果より，楽器経験とタップ能力の間には相関がみられなかったことから，楽器経験はタップ能力に影響しないと思われる．

2.5.5　タップの有無の歌唱への影響

タップによって歌唱が不安定になるなどの影響があれば，判定精度にも何らかの影響が出る可能性がある．そこで，タップなしの歌唱による変換結果が得られる TVM 以外の 3 システムの課題曲の結果を用いて，タップあり（タップしながら歌唱したが，3 システムともタップ情報は処理に用いていない）とタップなしとで比較し，タップの歌唱への影響を調べた．

表 2.5: タップの有無による赤とんぼの被験者全体の再現率・適合率・F 値の比較（テンポ自由）

	CMP		RYN		BP2	
	タップ有	タップ無	有	無	有	無
再現率	85.6	85.4	87.2	88.9	92.7	94.7
適合率	84.0	83.3	79.5	75.4	88.7	86.4
F 値	84.8	84.3	83.2	81.6	90.6	90.4

単位：%

表 2.5 にテンポ自由歌唱の結果を示す．各システムについてタップの有無に分けて，全被験者の合計値を示す．全被験者の歌唱音数（母数）はタップありで 836 音，タップなしで 830 音であった．CMP, RYN, BP2 いずれも再現率, 適合率ともにタップの有無によらず同等の判定精度であった．よって，タップの有無はほとんど影響しないと考えられる．

表 2.6 に BPM=120 の歌唱の結果を示す．全被験者の歌唱音数（母数）はタップありで 837 音，タップなしで 835 音であった．BP2 は，タップの有無に関わらず再現率・適合率ともに大きな差は見られなかった．CMP では，自由テンポ時には同等だった再現率が，タップありの方がやや低くなった．RYN はタップありで再現率 87.5 %, 適合率 84.7 %, タップなしで同 81.7 %, 77.1 %であり，タップありが再現率・適合率ともにタップなしを上回った．これは，被験者 E のタップなし歌唱時の誤り音が 35 音でタップあり歌唱時の 11 音に対して大きく増えているのが主因である．

以上から，総じて赤とんぼのような曲やテンポでは，タップの有無は歌唱にほとんど影響しないと言える．なお，BPM=120 の場合にタップの有無が若干影響する可能性が見られたが，必ずしもタップありの場合に悪影響が出るわけではない．

表 2.6: タップの有無による赤とんぼの被験者全体の再現率・適合率・F 値の比較（BPM=120）

	CMP		RYN		BP2	
	タップ有	タップ無	有	無	有	無
再現率	83.8	86.1	87.5	81.7	78.5	79.0
適合率	86.9	86.7	84.7	77.1	92.1	92.1
F 値	85.3	86.4	86.1	79.3	84.8	85.0

単位：%

2.5.6 全体考察

TVM システムは，歌唱時の負荷や速いテンポなどでタップと歌唱のズレの発生はあるものの，既存の歌詞歌唱などの任意の発音の歌唱を許容するシステムに比べて，欠落する音や不要な音の発生が抑制され，音数および音高判定精度が向上することが示された．

楽器経験の有無のタップへの影響については，赤とんぼレベルの曲であれば，多少速いテンポの入力であっても大きく影響しないと見られることが分かった．また，タップの有無の歌唱への影響についても，赤とんぼレベルの曲の場合，入力テンポが速くなると多少影響が出る可能性があるものの，必ずしもタップが悪影響を及ぼすわけではなく，総じてタップの有無の影響は小さいことが分かった．

著者は，これまでに文献 [69] において市販の「タタタ歌唱」システムに自由歌唱を入力して比較実験を行っている．文献 [69] では，今回提案した手法と比べて無発声検知機構や精度の高い F0 推定法を採用していない，性能が劣る手法を用いたが,「タタタ歌唱」を必要とするシステムに対する優位性を示している（提案手法は再現率 65.9 %，適合率 70.2 %，比較システムは同 24.0 %，36.8 %）．この結果と合わせて，総じて TVM は十分な有用性があると考えられる．

2.6 おわりに

本章では，3 章および 4 章に先立って，それらの基盤となる，人間のタップによる音区切り情報の入力と計算機の音高抽出を用いた協調的な音高判定技術を提案した．これを，計算機を用いた音楽制作に用いられている MIDI シーケンスデータ入力法の一つである Voice-to-MIDI において，音高ならびに音数の判定精度を向上させる課題に適用するため，タップ併用型 Voice-to-MIDI システムとして実装し，評価を行った．

評価では，歌唱されたフレーズから音高・音数を取得することによって，歌詞歌唱などの任意の発音の歌唱を許容する既存の Voice-to-MIDI システムとの判定精度を比較した．また，楽器経験の有無やタップの有無の判定精度への影響の評価を行った．

その結果，タップの付加により音数抽出の正確さが増し，それが音高判定の精度向上にも寄与したことを示した．また，楽器経験の有無やタップの有無は判定精度への影響に大きく影響しないことが分かった．これらのことから，人間によるリアルタイムの音区切り情報の入力が可能であることを確認した．

今後，タップへの依存度を減らすために必要なタップか否かを判定する機構を開発することや歌詞先作曲における実践的な使用評価を行っていく予定である．

2.7 謝辞

文献 [23] について，プログラムの提供および比較評価への使用を快諾いただいた，Matti Ryynanen 氏および Anssi Klapuri 博士に感謝の意を表します．

また，多忙な中，評価実験に参加いただいた，被験者の皆様に感謝の意を表します．

表 2.7: 赤とんぼの変換結果　[歌唱条件：テンポ自由，歌詞歌唱，タップあり]

被験者	全歌唱音(音)	正解(音)				上:誤り(音)/下:結合(音)				上:欠落(音)/下:結合(音)				余分(音)				再現率(%)				適合率(%)				F値			
		TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2
A*	93	93	76	78	87	0	5	12	0	0	12	3	6	0	13	29	14	100	81.7	83.9	93.5	100	80.9	65.5	86.1	100	81.3	73.6	89.7
						0	5	0	0	0	10	3	1																
B*	93	93	82	77	80	0	3	3	1	0	8	13	12	0	7	8	6	100	88.2	82.8	86.0	100	89.1	87.5	92.0	100	88.6	85.1	88.9
						0	2	1	0	0	8	3	2																
C*	92	88	81	86	73	4	1	2	1	0	10	4	18	0	18	14	4	95.7	88.0	93.5	79.3	95.7	81.0	84.3	93.6	95.7	84.4	88.7	85.9
						0	1	0	0	0	4	1	2																
D*	93	92	75	88	90	1	4	3	0	0	14	2	3	0	6	5	13	98.9	80.6	94.6	96.8	98.9	88.2	91.7	87.4	98.9	84.3	93.1	91.8
						0	4	0	0	0	14	2	3																
E	93	91	75	51	88	2	4	38	0	0	14	4	5	2	18	11	9	97.8	80.6	54.8	94.6	95.8	77.3	51.0	90.7	96.8	78.9	52.8	92.6
						0	0	0	0	0	2	4	2																
F	93	93	88	88	90	0	0	2	1	0	5	3	2	0	29	41	28	100	94.6	94.6	96.8	100	75.2	67.2	75.6	100	83.8	78.6	84.9
						0	0	0	0	0	4	3	2																
G	93	92	84	87	90	1	0	3	1	0	9	3	2	0	11	6	14	98.9	90.3	93.5	96.8	98.9	88.4	90.6	85.7	98.9	89.4	92.1	90.9
						0	0	0	0	0	4	2	2																
H	93	93	77	88	87	0	3	0	0	0	13	5	6	0	2	4	2	100	82.8	94.6	93.5	100	93.9	95.7	97.8	100	88.0	95.1	95.6
						0	3	0	0	0	5	2	3																
I	93	93	78	86	90	0	5	2	0	0	10	5	3	0	7	5	5	100	83.9	92.5	96.8	100	86.7	92.5	94.7	100	85.2	92.5	95.7
						0	5	0	0	0	7	3	2																
合計	836	828	716	729	775	8	25	65	4	0	95	42	57	2	111	123	95	99.0	85.6	87.2	92.7	98.8	84.0	79.5	88.7	98.9	84.8	83.2	90.6
						0	20	1	0	0	58	23	19																

注 1.“＊”付きの被験者は「楽器経験なし」と回答した被験者（表 2.8 も同様）

注 2. 欠落音の下段は欠落音中の結合音数，誤り音の下段は誤り音中の結合音に起因する誤り音数を示す．また，誤り音と結合音由来の誤り音の差分は F0 推定ミス由来の誤り音数を示す．(表 2.8 も同様)

注 3. 黒地白文字：タップあり歌唱で 4 システム中最もよい値を示す．但し誤り・欠落音の下段の結合音と結合音由来の誤り音は対象外とする．(表 2.8 も同様)

表 2.8: 赤とんぼの変換結果　[歌唱条件：BPM=120，歌詞歌唱，タップあり]

被験者	全歌唱音(音)	正解(音)				上:誤り(音)/下:結合(音)				上:欠落(音)/下:結合(音)				余分(音)				再現率(%)				適合率(%)				F値			
		TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2
A*	93	93	71	80	76	0	5	10	0	0	17	3	17	2	8	16	13	100	76.3	86.0	81.7	97.9	84.5	75.5	85.4	98.9	80.2	80.4	83.5
						0	4	0	0	0	8	3	3																
B*	93	93	76	77	76	0	3	2	0	0	14	14	17	3	2	11	5	100	81.7	82.8	81.7	96.9	93.8	85.6	93.8	98.4	87.4	84.2	87.4
						0	3	0	0	0	13	3	3																
C*	93	85	83	78	54	7	2	10	2	1	8	5	37	1	20	9	0	91.4	89.2	83.9	58.1	91.4	79.0	80.4	96.4	91.4	83.8	82.1	72.5
						0	0	0	0	0	3	1	1																
D*	93	93	67	75	88	0	9	16	2	0	17	2	3	0	3	1	7	100	72.0	80.6	94.6	100	84.8	81.5	90.7	100	77.9	81.1	92.6
						0	8	0	0	0	16	2	2																
E	93	73	79	77	62	5	1	11	1	15	13	5	30	11	17	3	6	78.5	84.9	82.8	66.7	82.0	81.4	84.6	89.9	80.2	83.2	83.7	76.5
						1	0	0	0	1	3	3	1																
F	93	90	81	80	67	3	3	5	0	0	9	8	26	0	15	18	3	96.8	87.1	86.0	72.0	96.8	81.8	77.7	95.7	96.8	84.4	81.6	82.2
						0	1	1	0	0	6	4	2																
G	93	90	77	88	80	1	2	3	2	2	14	2	11	2	7	6	11	96.8	82.8	94.6	86.0	96.8	89.5	90.7	86.0	96.8	86.0	92.6	86.0
						0	0	0	0	0	3	2	2																
H	93	93	81	89	71	0	2	0	0	0	10	4	22	0	3	3	1	100	87.1	95.7	76.3	100	94.2	96.7	98.6	100	90.5	96.2	86.1
						0	2	0	0	0	4	1	3																
I	93	92	86	88	83	1	2	2	0	0	5	3	10	0	2	6	3	98.9	92.5	94.6	89.2	98.9	95.6	91.7	96.5	98.9	94.0	93.1	92.7
						0	2	0	0	0	5	2	3																
合計	837	802	701	732	657	17	29	59	7	18	107	46	173	19	77	73	49	95.8	83.8	87.5	78.5	95.7	86.9	84.7	92.1	95.8	85.3	86.1	84.8
						1	20	1	0	1	61	21	20																

表 2.9: 自由曲の変換結果　[歌唱条件：テンポ自由，歌詞歌唱，タップあり]

被験者	全歌唱音(音)	正解(音)				上:誤り(音)/下:結合(音)				上:欠落(音)/下:結合(音)				余分(音)				再現率(%)				適合率(%)				F値			
		TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2	TVM	CMP	RYN	BP2
A*	120	**85**	82	**97**	92	15	12	14	5	20	26	9	23	0	10	15	9	70.8	68.3	**80.8**	76.7	**85.0**	**78.8**	**77.0**	86.8	77.3	73.2	78.9	**81.4**
						13	9	5	3	16	18	8	7																
B*	63	58	50	46	44	5	2	3	1	0	11	14	18	0	3	4	2	92.1	79.4	73.0	69.8	**92.1**	**90.9**	**86.8**	93.6	92.1	84.7	79.3	80.0
						0	1	0	0	0	7	3	7																
C*	61	52	43	37	14	9	4	3	3	0	14	21	44	0	9	0	0	85.2	70.5	60.7	23.0	**85.2**	**76.8**	92.5	**82.4**	85.2	73.5	73.3	35.9
						0	1	1	3	0	9	7	14																
D*	122	120	83	92	99	2	4	10	0	0	35	20	23	0	14	16	20	98.4	68.0	75.4	81.1	98.4	**82.2**	**78.0**	**83.2**	98.4	74.4	76.7	82.2
						0	2	1	0	0	30	19	16																
E	98	80	65	79	65	10	10	7	0	8	23	12	33	8	27	20	4	81.6	66.3	80.6	66.3	**81.6**	**63.7**	**74.5**	94.2	81.6	65.0	77.5	77.8
						1	1	2	0	7	9	12	11																
F	172	155	125	134	134	8	9	7	1	9	38	31	37	2	60	39	31	90.1	72.7	77.9	77.9	93.9	**64.4**	**74.4**	**80.7**	92.0	68.3	76.1	79.3
						5	3	0	0	9	26	27	21																
G	90	90	71	78	66	0	5	4	1	0	14	8	23	0	27	13	12	100	78.9	86.7	73.3	100	**68.9**	**82.1**	**83.5**	100	73.6	84.3	78.1
						0	2	0	0	0	8	8	14																
H	198	193	139	149	140	3	7	5	1	2	52	44	57	0	5	3	0	97.5	70.2	75.3	70.7	**98.5**	**92.1**	**94.9**	99.3	98.0	79.7	83.9	82.6
						2	6	1	1	2	36	42	43																
I	209	197	176	179	166	12	5	11	2	0	28	19	41	1	22	19	7	94.3	84.2	85.6	79.4	**93.8**	**86.7**	**85.6**	94.9	94.0	85.4	85.6	86.5
						0	3	1	1	0	17	16	17																
合計	1133	1030	834	891	820	64	58	64	14	39	241	178	299	11	177	129	85	90.9	73.6	78.6	72.4	93.2	**78.0**	**82.2**	**89.2**	92.0	75.7	80.4	79.9
						21	28	11	8	34	160	142	150																

注 1.“＊”付きの被験者は「楽器経験なし」と回答した被験者

注 2. 欠落音の下段は欠落音中の結合音数，誤り音の下段は誤り音中の結合音に起因する誤り音数を示す．また，誤り音と結合音由来の誤り音の差分は F0 推定ミス由来の誤り音数を示す．

注 3. 黒地白文字：4 システム中最もよい値を示す．

第 3章

環境からの触発を受けて音情景を再構成するための楽器

本章では，第 2 章で提案した技術が持つ，人間による音の取捨選択が可能な特長を用いて，提案技術を自然音などの環境音を主な入力音とする課題に提案技術の適用を試みる．具体的には，環境に応じて得られるあらゆる音の時系列に対して，ユーザ自身の心象風景を音の取捨選択という形で織り込んで環境の音情景を再構成するための新しいリズム楽器“ EnvJamm ”の提案を行い，評価を行う．

3.1　はじめに

世界は音楽で満ちている．街角に流れる「人工的な音楽」のみならず，そよぐ風の音や人混みのざわめき，交通の喧噪など，我々の身の回りの環境が内包する音情景（soundscape）[71] は，すべて潜在的に音楽的側面を有している．ただ，それがあまりにありふれた存在であるが故に，我々は多くの場合そこに音楽があることに気づかない．また，仮にそこに音楽があることに気づいたとしても，それを音楽として抽出し，記録し，奏でることは，音楽的な才能や技能が乏しい者にとっては，非常な困難を伴う．

もっと誰もが自分を取り囲む環境の中の音楽に気づき，それを自分なりの音楽的感性に基づいて編み直し（reweave）,「音楽」として表現を外在化することができないだろうか．本章で提案する“ EnvJamm ”は，このような夢を叶える新しい

リズム楽器である．

EnvJammは，あたかもジャムセッションをするかのように，街や自然などの環境との協調的な曲作りを楽しむことを可能とする．EnvJammは，マイクから取り込んだ環境音を音高ジェネレータとして音高情報を取得し，これにユーザがキーボードなどのタップデバイスで弾いたリズム情報を組み合わせることにより，音情景を再構成した演奏情報を出力する．なお出力される音の音色は，サンプリングしたままの環境音を用いるのではなく，電子音源の任意の音色を用いる．このように，音情景から取得するのは音高情報のみであるので，出力される音は，音情景を抽象化したものとなる．

EnvJammer（EnvJammを用いる作曲家・演奏家）は，身を置く環境から五感への多様な刺激を受け，これに触発されて様々に変容する自らの内的心象を反映したリズムを刻む．こうして，環境から得られる音に自らの心象風景を織り込んで環境の音情景を再構成する．このように，EnvJammによって環境と人の表現を融合した新たな音楽の作成が可能となる．加えて，創作を通じて環境のありように対する新たな気づきや視点の変化が促進されることも期待される．

なお，環境の中の音楽に気づきという点において，本研究と関連する分野・概念にサウンドスケープがある．サウンドスケープでは，調性音楽は排除され，自然界や街の喧噪のような音だけが注目されることもあるが，街の喧噪の中でBGMが聞こえてくる場合がある．これをサウンドスケープとみなすかどうかは個々の主義や哲学，立場があるため議論しないが，本研究では，その音空間において調性音楽が強い個性や支配力を得ているかどうかの程度の問題とみなし，サウンドスケープは調性音楽（歌唱や楽器音）も包含するという立場をとる．

3.2 先行研究

サウンドスケープは，音楽家であるR. マリー・シェーファーが提唱した[71]，音の環境を示す言葉であり，我々の周りにある音に注意を向け，理解することによって音環境を考える（サウンドスケープ・デザイン）ことがその根底にある．

R. マリー・シェーファーは，サウンドスケープの分析においてまずやることは個性や支配力など重要な特徴を発見すること[72]と述べている．また，サウンド

スケープ・デザインでは，環境を巨大な音楽作品とみなし，作曲家がどの音を残し，どの音を広め，どの音を増やしたいかということを意識することを促している [73] ．このサウンドスケープを音楽に取り込もうとする試みがなされている．しかし，その出力が聴きなじみのあるリズム構造を持った音楽や調性音楽になるとは限らず，多分に実験音楽的要素を含んでいる．

ピエール・シェッフェルによるミュージック・コンクレートでは，動物の鳴き声などを録音し，音素材として音高などを変化させて楽曲を作成する [74] ．素材に対する加工を行っている点は本研究と同様であるが，素材を MIDI などに符号変換しない点が本研究と大きく異なる．

Karmen Franinovic らの Recycled Soundscape [75][76] は，音のサンプリングを行う Beludire と演奏を行う SonicBowls からなり，人間が SonicBowls で演奏した楽器音と Beludire でサンプリングした環境音をミックスすることにより“ Public orchestration ”を行う楽器である．環境音を音楽表現に取り込もうとする目標は本研究と同様であるが，環境音を MIDI などに符号変換しない点，再生される環境音の時系列が保持されない点が本研究と大きく異なる．

宮下らが文献 [77] の中で提案した Sound Dust は，掃除機を楽器化する試みである．掃除機にカメラをつけ，画像入力に基づいてバックトラックにエフェクトを掛ける．また掃除機の吸込音も積極的に変化させることで演奏を行っている．演奏と出力の関係性を見出すのが困難な点が本研究と類似している一方で，画像の入力に対してエフェクトの変化を出力としている点が異なる．

提案システムでは，マイクから入力した音そのものを用いるのではなく，MIDI シーケンスデータに変換することによって音色を自由に割り当てられる．音高やリズムレベルの再構成であるため，再構成という観点では，ミュージック・コンクレートのような入力された音から切り取った音の断片を楽曲に用いる方が本義に近い．しかし著者は，音断片を直接出力音にすることを表現上の「制約」と捉え，音符のようなより抽象化された状態にして出力音も作曲者に委ねることによってこの制約を取り除けると考えている．特に最近は，サンプラー [78][79][80] の普及によりさまざまな音を作曲家自ら録音して出力音にすることも容易になっており，音色選択に自由を求める作曲家にとっては有用な手法である．ゆえに本手法は，純粋な再構成による創作とは違った，再構成と音色的創造性を兼ね備えた新

たな音楽創作手法と言える．

3.3 提案システムの概要

3.3.1 音情景の再構築手順

EnvJamm は，第 2 章で提案した技術を用いて，マイクから取り込んだ環境音を音高ジェネレータとして音高情報を取得し，また人間によるキーボードなどのタップからリズム情報を取得し，これらを統合することによって演奏情報として音情景を再構成し，出力する．

再構成の手順は以下のようになる．

- セッションしたい環境に身を置きながら，EnvJamm（図 3.1）で環境音の入力を開始する．
- その環境の発するあらゆる情報（聴覚的，視覚的，触覚的 etc.）に触発されつつ，その触発をタップリズムとして表現してセッションを行う．具体的には，MIDI キーボードなどの MIDI 対応楽器や PC キーボードのキーをタップして音の開始を，またタップし続けることで音長を表現する．
- 入力終了後，EnvJamm は，タップされている区間それぞれについて入力された環境音から音高を取得し，音長とともに最終的に MIDI ノートイベント化し，SMF ファイルとして出力する．

EnvJamm は，演奏者がセッションしたい環境で直接セッションすることを想定している．しかし，タップ直後にリアルタイムで音を出力するとその音が音情景に大きく影響を与えてしまうため，リアルタイムでは音を出力しない．

3.3.2 システムデザイン

EnvJamm は，基本的に第 2 章で実装したタップ併用型 Voice-to-MIDI システムを踏襲するが，音高検知範囲についてより広い範囲の音高を扱えるように拡張を行い，さまざまな音高が存在する環境音への対応を行ったシステムである．

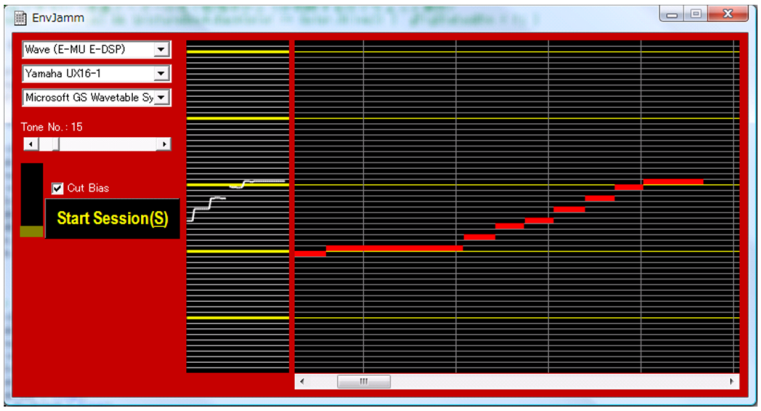

図 3.1: EnvJamm の概観

つまり，入力は環境音波形とリズム区切り情報，出力は D2-C#7 までの半音単位の音高（A4 = 440Hz を基準とする）を持った単音の MIDI データである．入力音声は 22050Hz，16bit, モノラルでサンプリングされ，リズム区切り情報には MIDI キーボードや PC キーボードの打鍵および離鍵の入力時刻情報を用いる．PC キーボードの場合は，タッピングに‘ < ’と‘ > ’の 2 キーを使用し，1 キーのみ連打しても 2 キーを交互に打鍵してもよい．短時間フーリエ変換は，フレームサイズ =2048samples：約 100ms，フレーム移動間隔=128samples：約 6ms で行われる．また，無発声検知機構で対象とする音高範囲も出力する音高範囲と同じ D2-C#7 までとなる．

システムの開発には Microsoft Visual C # 2005 を用い, 短時間ピッチ算出部については Visual C++2005 の DLL で作成した．

3.3.3 タップに対する音のマッピング

入力された環境音から切り出した区間の音高をどのように決定するか，という問題は出力される作品の質にも影響を与える．今回は 3.3.2 節で述べたようにシステムの初期的デザインとして，タップで切り出した区間に単音を割り当てることとした．

音高決定に用いている自己相関法は，本来単音のピッチ算出に有用な手法であり，複雑に混ざり合った音に対して注目した特定の音のピッチを抽出するような

表 3.1: 被験者の楽器経験

被験者	作曲経験	楽器経験
A	なし	なし

目的には向かず，全く無関係のピッチが算出されることもある．しかしEnvJammでは，これをその瞬間の音情景が総合的に織り込まれて表現された代表値であると捉えている．

3.4 システムの試用と評価

1名の被験者に実際にEnvJammを用いたセッションを行わせた．本節では，その概要と結果について述べる．

3.4.1 概要

被験者は，システムの使用方法などの説明を受けた後，北陸先端科学技術大学院大学の敷地内を散策し，セッションしたいと思った場所で1分程度の時間でセッションを行った．セッション終了後，被験者は作成されたMIDIデータの試聴およびアンケートへの回答を行った．

使用機材について述べる．Dell：Lattitude XTノートPC上でシステムを稼働し，タップデバイスには同PCのキーボード上の’＜’キーまたは’＞’キーのいずれかあるいは両方を用いた．集音マイクは同PC内蔵のマイクとした．

3.4.2 被験者

被験者は，筆者らが所属する大学の女子学生1名である．表3.1に被験者Aの作曲経験と楽器経験，今回セッションを行った場所を示す．

3.4.3　アンケート設問項目

実験後に行うアンケートの設問項目は以下とした．項目 1〜3 はセッション終了直後，項目 4・5 については，作成された MIDI データを試聴後に回答させた．いずれも自由記述である．

1. セッションしたシチュエーション
2. セッションの説明（セッション中に意識した音などの説明）
3. セッションの感想（システムの使用感など）
4. システムの問題点・気になった点
5. 作成された MIDI データを聴いて思ったこと
    - 評価に使った音色の Program No.

「セッションの感想」「システムの問題点・気になった点」

3.4.4　アンケートの結果と考察

被験者 A に対して行ったアンケート結果を示す（原文のまま記載）．

セッションを行ったシチュエーションは，大学構内のバス停付近（図 3.2）であった．周りにはバス停の他，大学の施設や駐車場，一般道路が存在している．当日は晴れの暖かい日であった．表 3.2 に各設問項目に対する回答コメントを示す．

設問 2「セッションの説明」のコメント「JAIST バス停で，車が走る音や機械の音やモーターの音や鳥の声，また，偶にカエルの音も取りました.」より，人工物から自然音まで，あらゆる音に注意を向けてセッションを行ったことがわかる．これらのコメントより EnvJamm を用いることによって，身の回りの音を注意深く聴いたり，周辺環境において注目するモノを視覚的に探索する等の行為，つまり環境からの刺激を十分に受け取るために必要な「感覚の研ぎ澄まし」の行為を自然に醸成できる可能性が伺えた．

次に，設問 5「作成された MIDI データを聴いて思ったこと」のコメント「連続的にいる変わってない音（モーターの音）が，思い通り連続的に変わっていると

図 3.2: 被験者 A が選んだシチュエーション

思いました.」より，被験者 A の意図が上手く織り込まれたフレーズが記録できたことが伺え，環境から得られる音に被験者の心象風景を織り込んで環境の音情景を再構成できたと考えられる[1].

3.5　システムを熟知した被験者による試用と評価

次に提案システムを熟知している著者が，テーマ別に 4 か所のロケーションで試用した結果について記す．試用を行ったロケーションは以下の通りである．括弧内はセッションのテーマを表す.

1. 海岸（自然音 1）
2. 林（自然音 2）
3. レストラン（屋内）
4. 交差点付近（騒がしい屋外）

[1]実際のところ，Envjamm ではユーザの意図を完全に記録できるとは限らない．また，例え意図と全く違ったフレーズであってもユーザが気に入るかもしれず，出力されたフレーズを第三者が聞いただけでは，ユーザ自身がフレーズに対して行った評価を推測するのは難しい．その点において，被験者 A のコメント内容と「思い通り」というコメントから，今回はリズムと音高について共に被験者 A の心象風景がとても上手く織り込まれたよい例と推測できる.

表 3.2: 被験者 A のアンケート結果

設問	コメント
1. セッションしたシチュエーション	構内バス停付近
2. セッションの説明	JAIST バス停で，車が走る音や機械の音やモーターの音や鳥の声，また，偶にカエルの音も取りました.
3. セッションの感想	バスが止まった時，ずっとモーターの音を取った.
4. システムの問題点・気になった点	長い時間（1 分間ほど）で音を取るので，あと違う音色を聞く時，先きいた音色のリズムがわからなくなっちゃったことが気になりました.
5. 作成された MIDI データを聴いて思ったこと	（試聴音色：ダルシマー） 連続的にいる変わってない音（モーターの音）が，思い通り連続的に変わっていると思いました.

図 3.3: セッションを実施した海岸の情景

使用機材について述べる． Dell : Lattitude XT ノート PC で本システムを稼働し，タップデバイスには同 PC のキーボード上の ’ < ’ キーまたは ’ > ’ キーのいずれかあるいは両方を用いた．集音マイクは同 PC 内蔵のマイクとした．

3.5.1 各セッションの詳細

以下にセッション別に詳細を述べる．いずれのロケーションも石川県小松市内に存在する．

海岸

安宅の関付近の海岸（図 3.3）で行った．ここでは，人工音や生物の声や鳴き声が極力ないような自然とのセッションを行った．ほぼ波の音のみの環境であった．

ここでは主に海岸に打ち寄せる波の周期を聴覚的，視覚的に感じながらセッションした．波のタイミングによって大小や周期性がさまざまに変容する様子を意識することでき，また，何気ない波音への興味の深まりを覚えた．

林

安宅の関にある松林に囲まれた休憩スペース（図 3.4）で行った．波の音などに小動物の鳴き声を加えたセッションを行った．図 3 において，林の後方には海岸

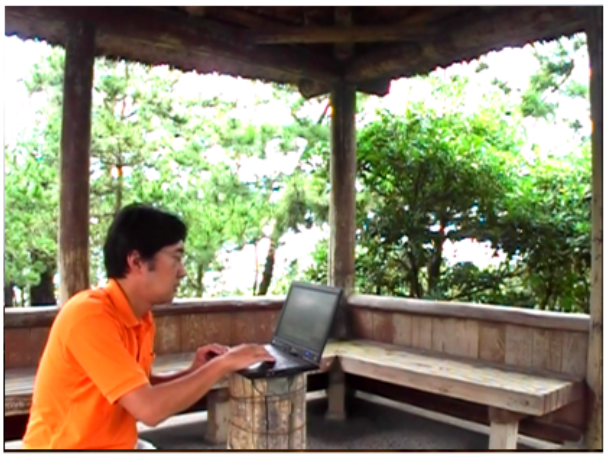

図 3.4: 林の中でのセッション風景

があり，波の音が聞こえる他，蝉の鳴き音や鳥のさえずりが聞こえた．主に蝉の声と鳥のさえずりのリズムに触発された．しかし，波の音も同時に取得したいという欲求も生まれた．今後，和音や複数のリズムを同時入力できるようなシステムに拡張することが検討事項として浮かび上がった．

レストラン

今回の試用における唯一の屋内セッションかつ比較的ピッチがはっきりと表れやすい人声が存在するロケーションである．会話や子供達の声，食器のあたる音など様々な音が存在した（図 3.5）．ここでは主に人の声や食器の音に注目してセッションを行った．騒がしい中で注目する音を切り替えていく一方で，非常に多くの音ソースが存在していたため，とりこぼした音ソースに対して「もったいなさ」を感じた．

交差点付近

交通量のある交差点付近（図 3.6）という比較的騒がしいロケーションにおけるセッションを行った．雨が降っており，車の走る音や付近の店の店外放送などが存在した．ここでは車の走行音やすれ違い音に注目してセッションを行った．車両

図 3.5: レストランでのセッション風景

図 3.6: セッションを行った交差点付近の情景

サイズや速度によって音が大きく変わることを積極的に活用しようという意識が生まれた.

3.5.2 考察

セッションの結果，単発の音やリズムを持った音，高い音や低い音など様々な音が紡ぎ出した音風景を活かしたフレーズを作成し，音風景を再構成できたと感じる．また，環境から得られる音に対して音区間の取捨選択によって自らの心象風景を織り込む行為から，特定の音やモノに注目し，気付きや興味を持つ効果を実感できた．これは，与えられた環境から注目する音やモノを探索する行為や選択した音区間がどれぐらい続くかを判断する等の行為，つまり 3.4 節の被験者 A と同様，聴覚に限定しない「感覚の研ぎ澄まし」が行われた結果と考えられる．それに加えて,「交差点付近」のセッションにおいて,「車両サイズや速度によって音が大きく変わることを積極的に活用しようという意識が生まれた.」という点は，環境からの「触発」を受けたと考えられる．

以上より，環境と人の表現を融合すること，および環境のありように対する新たな気づきや視点の変化の促進という目的は達成できたと考える．一方で，単音のみの入力である現在のシステムでは場面により不満を感じた.

3.6 全体考察

1 名の被験者を用いた試用評価では，被験者のコメントより，被験者の意図が上手く織り込まれたフレーズが記録できたことが伺え，環境から得られる音に被験者の心象風景を織り込んで環境の音情景を再構成できたと考えられる．また，EnvJamm を用いることによって，身の回りの音を注意深く聴いたり，周辺環境において注目するモノを視覚的に探索する等の行為，つまり環境からの刺激を十分に受け取るために必要な「感覚の研ぎ澄まし」の行為を自然に醸成できる可能性が伺えた.

システムに詳しい著者自身による評価では，単発の音やリズムを持った音，高い音や低い音など様々な音が紡ぎ出した音風景を活かしたフレーズを作成し，音

風景を再構成ができた．また，先述の被験者の試用評価と同様，身の回りの音を注意深く聴くなどの「感覚の研ぎ澄まし」の行為が見られ，加えて，環境からの「触発」も感じられた．

以上のことから，EnvJamm によって環境の音情景の再構成を通じた環境と人の表現を融合した音楽の作成，および環境のありように対する新たな気づきや視点の変化の促進という目的は達成できたと考える．

3.7 おわりに

本章では，第 2 章で提案した技術が持つ，人間による音の取捨選択が可能な特長を用いて，自然音などの環境音を主な入力音とする課題に提案技術の適用を試みた．

環境に応じて得られるあらゆる音の時系列に対して，ユーザ自らの心象風景を音の取捨選択という形で織り込んで環境の音情景を再構成するための新しいリズム楽器“ EnvJamm ”を提案し，被験者 1 名による試用評価および著者による 4 か所の環境における試用と評価を行った．今後，更なる作曲の試み，複数音対応などの新たな音マッピング方略の検討を行う予定である．

第 4章

認知症患者に対する音楽療法支援システムへの応用

本章では，2 章で提案した技術が持つ，人間による音の取捨選択が可能な特長を用いて，他者が自由に発する非音楽的な音声が入力音となる課題に提案技術の適用を試みる．認知症患者が常同言語の繰り返しなどの状態から抜け出すことや落ち着かせることを目的として，音楽的には初心者である介護者が音楽療法に携わることを支援する伴奏システム“ MusiCuddle ”を提案し，評価を行う．

4.1 はじめに

認知症は，器質性の障害による知能や認識スキルなどの低下であり，その症状は総じて BPSD（Bihavioral and Psychological Symptoms of Dementia）と呼ばれ，徘徊癖，幻覚，暴言などがあるが，適切な介護によって緩和したり進行を遅らせると考えられている．中でも音楽療法は，認知症の症状を和らげる手法の一つとして知られており，通常は，音楽的な知識や演奏技術を持った専門家が行い，専門家は，要請があれば介護ホームなどを訪れて BPSD の緩和のために音楽療法を行う．

しかし，専門家が必要ゆえに，その有用性にも関わらず患者が音楽療法を受ける機会を得ることは容易とは言い難い．そのため，音楽的な知識が乏しい介護者でも音楽療法に携わることを支援するシステムの開発が望まれる．

そこで本章では，音楽療法を支援するシステムの構築を試みる．認知症患者の不安などの心的状態から生じる症状の一つに,「常同言語」と呼ばれる何度も同じ言葉を繰り返す症状がある．この常同言語の繰り返しなどの状態から抜け出すことや落ち着かせることを目的として，音楽的には初心者である介護者が音楽療法に携わることを支援する伴奏システム“ MusiCuddle ”を提案する．

MusiCuddle は，患者が何らかの発声をしているときに，システム操作者が鍵盤などをタップすることによってトリガーを入力すると，システムがそのトリガーのタイミング付近の患者の声の音高を取得して，予めその音高に割付けられたカデンツなどの音楽フレーズを再生する．カデンツは「同質の原理（Iso-principle）[81]」にならって作成することによって，最初は常同言語を繰り返している患者の現在の感情に寄り添い，カデンツの最後には患者の症状を落ち着かせる効果が期待される．

MusiCuddle を用いて音楽フレーズを演奏することによって患者の行動がどう変わったかを評価するため，ケーススタディとして患者 1 名に対する調査を行った．

4.2 先行研究

音楽療法に関する研究として Clark[82] は，入浴中に録音された音楽を演奏すると患者が攻撃的になる機会を減らせることを示した．Ragneskog ら [83] は，介護者が音楽を使って患者をなだめた幾つかの事例を示し，Svansdottir ら [84] は，音楽療法によって活動の妨害や攻撃性，不安が顕著に減少することを示した．また，Park ら [85] の実験が，認知症患者に好みの音楽を聴かせると，聴かせる前よりも落ち着くことを示した一方で，Nair ら [86] は，柔らかくゆったりとしたバロック音楽では，落ち着かせる効果がないことを示した．これらのことから，音楽療法の効果を得るためには，音楽は個々の患者に合わせたものを用意すべきと考えられる．

これについて，アメリカの音楽療法士にはよく知られている「同質の原理（Iso-principle）[81]」では，患者の感情を早期に落ち着かせるために，その患者の状況に合わせた音楽を使用する [87]．例えば，患者が興奮しているときには，まず患者の感情に合う音楽を提示するべきであり，エネルギッシュで親しみのある音楽が

良いと考える [88]．そして，この原理は音量やリズムについても適用できるとされている [81]．

このような音楽療法は，BPSD の症状の 1 つである常同言語症状に対して効果があると考えられる．MusiCuddle では，Iso-principle を利用し，最初は常同言語を繰り返している患者の感情に寄り添い，徐々に患者の症状を落ち着かせるような音楽を提示することを目指している．

4.3 常同行動や発声を繰り返す患者について

ここで，常同行動や発声を繰り返す患者の例について示す．1ヶ月に渡って精神的に不安定なときの発声を記録したものをまとめ，表 4.1 に患者の症状が出ている状態，認知症の原因，日頃の様子，録音時のシチュエーションを示す．表中の「HDS-R（長谷川式簡易知能評価スケール）[89]」は，9 つの簡単な質問を含んだ知能テストの 30 ポイント中の最高スコアである．なお，21 ポイント以上では認知症ではないとされ，軽度の患者の平均が 19.1 ポイント，中程度で 15.4 ポイント，重度では 10.7 ポイントとされている．

表中のスコアより，ほとんどの患者が重度であることがわかる．患者 C や I のような軽度の患者であっても精神の不安定さや発声の繰り返しが見られた．また，そこにはいない誰かや介護者を呼ぶように感情のままに奇声を発する患者がいたり，患者 A のように，介護者が患者の呼びかけに応えた後も同じ呼びかけを繰り返す患者もいたが，多くのケースにおいて，患者が何と言っているのかを聞き取れなかった．医師や看護師，介護者は患者と対話したり他の行動へと誘導しているが，それでも患者を落ち着かせることは容易とはいえない．従って，患者の感情に寄り添ったり症状を落ち着かせるような音楽は，介護者が患者をケアすることを支援するものとして期待される．

表 4.1: 患者のデータおよび常同行動や発声の例

患者	認知症の原因	性別	年齢	HDS-R	日ごろの患者の様子
	収録されたデータの内容				
A	アルツハイマー型	女	89	7	大声で頻回に「トイレしたい」と人を呼ぶが空振りが多い.
	高音，裏声，大声で「ねー」の連続発音.				
B	脳血管性	男	58	3	易怒性あり．日ごろは威嚇するような粗暴な言動がある.
	やや泣くような声で「動きます」「迷惑かけないから」「お願いします」と発話.				
C	脳血管性	女	77	20	記憶障害は目立たないが，大声・常同言語が目立つ.
	高音・裏声で「＊ったか」と同じ言葉を繰り返す.				
D	アルツハイマー型	女	69	施行不能	意味のある発話はない.
	かすれた声で「あー」と叫ぶ.				
E	アルツハイマー型	女	78	施行不能	重度.
	大声で「たすけて」などの発話を繰り返す.				
F	アルツハイマー型	女	90	施行不能	奇声を出しながらの徘徊が目立つ.
	だみ声で意味不明な叫びを繰り返す.				
G	アルツハイマー型	男	78	施行不能	大声・易怒性が目立つ.
	意味不明な発話をする.				
H	アルツハイマー型	女	82	施行不能	常に大声．疎通性は不良.
	怒ったように発話．泣きながら発話．発話内容は不明.				
I	前頭側頭型	女	70	17	疎通性良好．独り語・常同言語が目立つ.
	トイレに入りこみ，扉を閉めて正座して,「＊ったですよ」の繰り返し.				
J	レビー小体型	女	83	13	易怒性が目立つ.
	怒った声で「ばかやろう」と発話.				

4.4 MusiCuddle システムについて

4.4.1 概要

MusiCuddle システムの動作概要について述べる（図 4.1）．MusiCuddle では，第 2 章で提案した技術を用いて，患者の発声から操作者（例えば，介護者）が任意に選択した区間の音高を取得する．次にその音高に紐付いた MIDI フレーズ等を出力する仕組みとなっている．

コンピュータ上の MusiCuddle ソフトウェア（図 4.2）を介して，MIDI 出力が可能な電子キーボードとスピーカ，マイクが接続されている（電子キーボードはなくてもよい）．電子キーボードの鍵盤のいずれか 1 つあるいは MusiCuddle ソフトウェアのトリガーボタンを操作者が押すと，MusiCuddle ソフトウェアは予め用意された音楽フレーズデータベースの中の 1 曲を演奏し，スピーカから出力する．

演奏する音楽フレーズの選択方法について述べる．MusiCuddle ソフトウェアはマイクから患者の声を取り込んでおり，操作者のトリガー入力によって，その入力時刻付近の声から音高を取得する．次に，この音高を基にしてデータベースから音楽フレーズを一つ選び出す．なお，MusiCuddle ではトリガーデバイスを押し

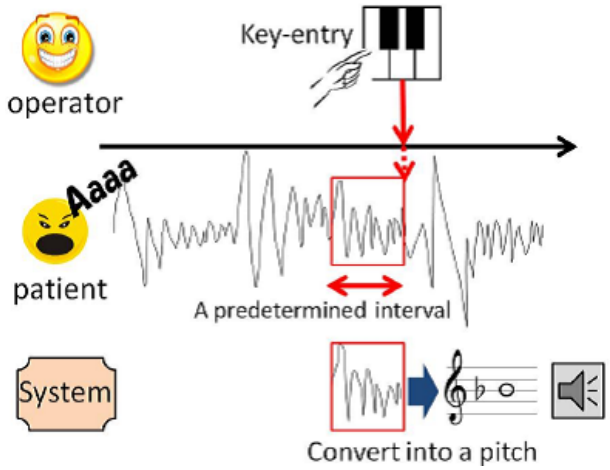

図 4.1: 音高取得の流れ

たときのタイミングのみを使用し，離したときのタイミングは使用しない．そのかわり，音高を取得する時間範囲を予め調整することができる．通常は患者の声を聴いてからトリガーを入力するため，トリガー入力時点よりも数 100ms 程度前～トリガー入力時点まで範囲指定すればよい．

例えば音楽フレーズがカデンツ[1]であった場合，患者の発声から取得した音高を最初のコードの最高音に持つカデンツが選択される．また，操作者は患者の状態に応じて，2 タイプのカデンツを選択できるようになっており，電子キーボードの場合は黒鍵を押すか，白鍵を押すかで使い分ける．

システムの利便性を考慮すると，患者の声と音楽の演奏が同時に存在していても，音楽が処理に与える影響を取り除いて患者の声から音高を取得できるようにする必要がある．そこで通常は，音楽は左右同程度，声は必ず左右いずれか一方のチャンネルがより大きく録音されるようにスピーカ（モノラル）とステレオマイクを配置することが望まれる．MusiCuddle ソフトウェアには，マイクのステレオ信号から差分信号を生成する処理が実装されているため，センターキャンセルによって声の成分のみを残すことで，F0 推定の精度を高めている．

[1]カデンツとは，楽曲の中間部や最後によく見られる和音構造を意味し，楽曲の続きを予期させたり，楽曲の終わりを感じさせることができる．

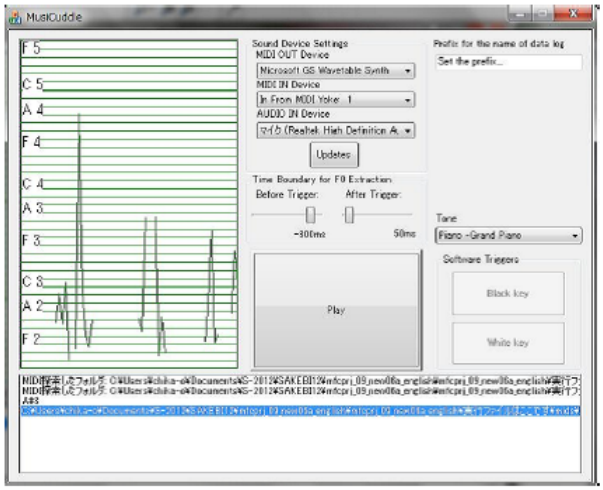

図 4.2: MusiCuddle のユーザインタフェイス

4.4.2 データベースに用意した音楽フレーズ

表 4.2 は，4.5 節の調査のために用意した MusiCuddle で使用できる音楽フレーズの内訳である．

一般的に，不協和音と協和音の組み合わせによって身体の緊張と弛緩を促せるとされており [90]，常同言語を繰り返す患者は，感情的であるなどの精神の不安定さがよく見られることから，これを落ち着かせることを考えて，不協和音から始まって協和音で終わるカデンツを用意した．また，同じ和音が続くフレーズも不協和音および協和音のそれぞれについて用意した．

その他，回想法による音楽療法が患者の抑鬱性を緩和する可能性が言われているため [91]，いくつかの唱歌の一部を抜き出したフレーズを用意した．また，患者の発する声を予めフレーズに変換したものも用意した．

4 つ連ねた和音

同じ和音を 4 回繰り返したフレーズを用意した．図 4.3 にその 1 例を示す．和音は 2 音で構成し，患者の発声から得た音高を基にフレーズを決めるため，上の音が C3〜C6 になる 37 種類を用意した（図 4.4）．リズム 4 種類，2 音の音程 3 種類，4 つの和音の音量変化の有無で 3 種類のバリエーションがあるため，最終的に 1332

種類の MIDI ファイルとなった．リズムは文献 [90] を参考に準備した．

カデンツ

不協和音から始まり協和音で終わるカデンツフレーズを 96 個用意した．これらはバッハ, J. S. 作曲の「平均律クラヴィーア曲集　第 1 巻」の全 24 曲およびショスタコーヴィチ, D. D. 作曲の「24 のプレリュードとフーガ」の全 24 曲の各曲の一部を抜き出したものである．多くは終盤から抜き出したが，不協和音が見られないものは曲の途中から抜き出している．

いずれの曲集とも 24 の調性（ド〜シの 12 音を主音とする長調と短調）の楽曲が揃っているため，患者の発声した音高に対応したフレーズを十分用意できる．また，いずれも 1 曲が「プレリュード」と「フーガ」で構成されるため，それぞれ 48 個のカデンツを抜き出し，合計で 96 個カデンツを用意した．各カデンツの最初の和音の最高音を用いてド〜シの 12 種類に分類され，よって各音高に対して複数のカデンツが用意される．

唱歌

高齢者がよく知っていると思われる唱歌の冒頭部分を抜き出して，MIDI ファイルにしたものを用意した．各曲とも開始音が C3〜C6 の 37 種類を作成した．図 4.5 は，唱歌「赤い靴」の一部について開始音を変えて表示した楽譜である．上段は開始音が C4 であり，下段は G ♯ 4 であり，同じ曲であっても，開始音が異なると全体の雰囲気が変わる．

常同言語

患者の常同言語として発声していた「いませんですよ」を音楽フレーズに変換して，開始音が C3〜C6 となる 37 種類の MIDI ファイルを作成した．

表 4.2: データベース上のフレーズ

Type	Details
4 つ連ねた和音	リズム：4 分音符，8 分音符，16 分音符，ランダム 音程：完全 1 度，完全 5 度（協和音程），長 7 度（不協和音程） 音量変化：変化なし，Crescendo，Decrescendo
カデンツ	Bach：Das Wohltemperierte Klavier Band 1, BWV846-869 Shostakovich：24 Preludes and Fugues, Op.87
唱歌	「雪」「赤い靴」「花」「月の砂漠」
常同言語	「いませんですよ」を「レレレシララシ」に変換

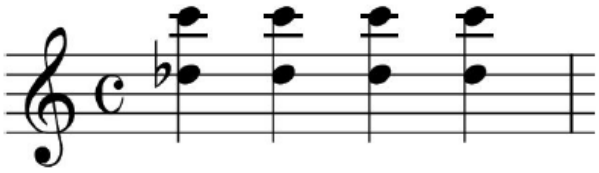

図 4.3: 4 つ連ねた和音の例

4.5　ケーススタディによるシステムの試用と評価

MusiCuddle によって患者の習慣がどう変わったかをケーススタディとして調査した．認知症の症状や患者の性格などが様々であるため，現段階ではまず 1 名の重症患者を対象とし，ごく短期間の調査であったため，MusiCuddle の効果を推定するために MusiCuddle の使用・不使用における患者の発声の記録や比較を行った．なお，この調査は，評価も含めて，システムを開発した筆者ではなく，共同研究者の大島が行った．

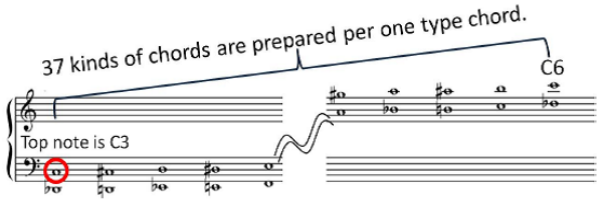

図 4.4: C3-C6 までの 37 種類を用意

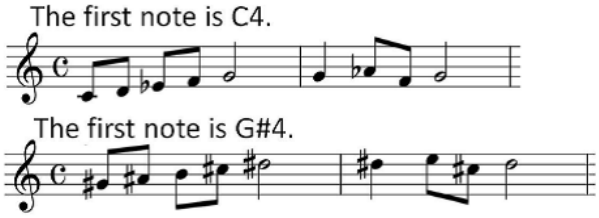

図 4.5: 開始音が異なることによって曲の雰囲気が変化

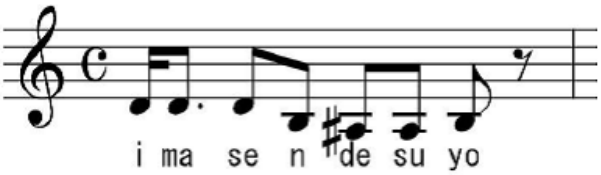

図 4.6: 患者の常同言語を音楽フレーズに変換したもの

4.5.1　調査にあたって

調査は，佐賀大学の研究倫理委員会の承認を得て行い，また，協力者となる患者やその家族に対してこの調査の意図および個人情報の取り扱いについての説明を行った．さらに，途中で調査をやめてもよいことを伝え，これらについて同意書を得た．

調査中は，主治医と看護師は同じフロアで活動し，参加者の様子をチェックできるようにした．もしMusiCuddleによる音の提示が適切ではなかったり，患者がより感情的になったときは，調査を即座に中止することとした．

4.5.2　調査協力者

この調査への協力者は，72歳の女性1名である．この患者は，毎日数時間に渡って常同言語を繰り返し，また，感情的になると常同言語を繰り返しながらトイレに鍵をかけてこもるといった行動が見られた．しかし，看護師の名前を覚えていたりしっかりとした挨拶をする面もあり，曜日や時間も答えられる．HDS-Rのスコアは17であった．以下に患者がよく発する言葉の例を示す．30秒程度の発声である．

- 「入ったよ（8回繰り返す）　ましたよ　いません　ません　いませんです　ません　（3回繰り返す）」

様々な種類の言葉を発したが，その多くはリズミカルで拍に沿っており（図4.7），リズムを抽出すると，言葉は違っても4/4拍子であることがわかった．

協力者は，特に空腹時には常同言語を繰り返し，午前11時ごろから昼食までトイレに閉じこもることが多かった．また，午後1時ごろからおやつの時間までは絶え間なく常同言語を繰り返していた．しかし，看護師が話しかけるとそれに反応することもあった．

4.5.3　予備調査

予備調査として，協力者が同じ発話を繰り返しているときに，協力者の近くで同じメロディとリズムで一緒に繰り返したときの反応を調べた．

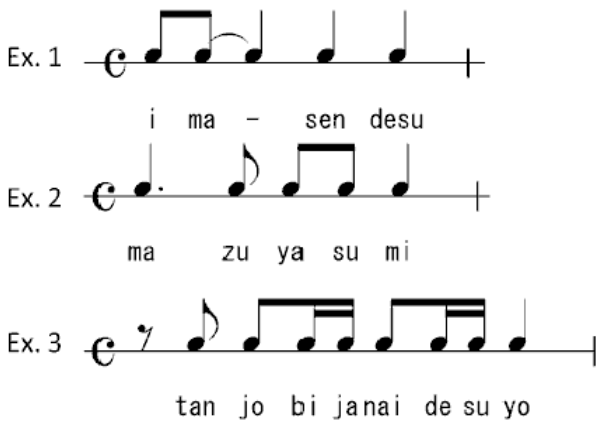

図 4.7: 調査協力者である患者が発する言葉のリズム

最初は，協力者が繰り返している発話と全く同じ発話を同じメロディ，リズムで協力者と一緒に繰り返した．次に，途中から協力者と同じメロディ，リズムのまま，別の発話内容に変えて一緒に繰り返した．すると，協力者と一緒に同じ発話を繰り返したときは，実験実施者に注意を向けつつも協力者は同じ発話を繰り返すことを続けていた．実験実施者が異なる発話に変えると，協力者もその発話に切り替えて一緒に繰り返すようになった．

4.5.4 調査方法

午前 10 時から正午までと午後 1 時から 2 時半までを対象に 2 日間行った．音楽の有無による 2 通りの協力者の声を比較する必要があるため，MusiCuddle の使用・不使用の 2 通りの環境で声の収録を行った．

MusiCuddle を使うときは，協力者が常同言語を繰り返し始めたら MusiCuddle にトリガーを送って音楽を再生した．

4.5.5 機材設定

調査は協力者が入院している病院内で行った．図 4.8 に機材の配置を示す．音楽は，Bluetooth による無線接続のスピーカから流し，協力者の声も Bluetooth 接続のマイクから取得する．これらはトイレのドアに設置した．

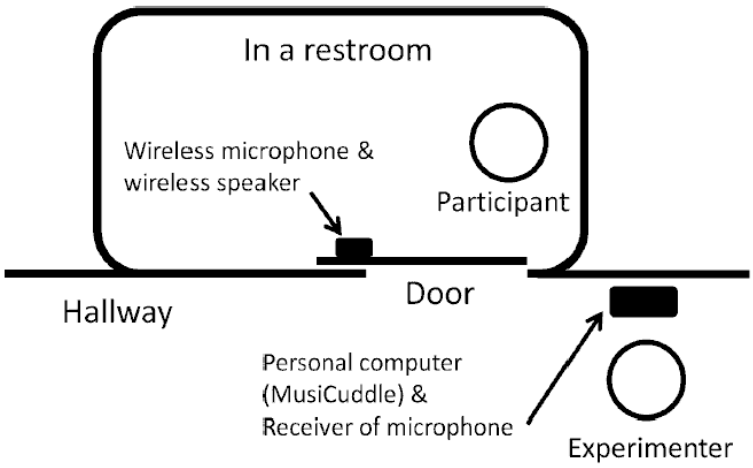

図 4.8: 機材の配置

MusiCuddle では，音楽が鳴っている中でも協力者の声を取得しやすいように本来ステレオマイクを使って収録する必要があるが，この調査では，設置場所の制約や無線接続の必要性などからステレオマイクを使用していない．従って，音楽が演奏されているときには，次の音楽を演奏するためのトリガーを入力しないこととした．

4.5.6 MusiCuddle の使用方法

この調査における MusiCuddle の使用方法について述べる．協力者がトイレ内で常同言語を繰り返しているとき，トイレの前にいる MusiCuddle の操作者が，MusiCuddle を使って音楽を流す．そのために，操作者はトイレの外で協力者の声を聴き，声の音高が安定したポイントを見つけたら，MusiCuddle ソフトウェア上のトリガーボタンを押すようにした．

トリガーが入力されたら，MusiCuddle は取得した協力者の声から音高を割り出し，その音高から開始される楽曲を 1 曲選びだし，演奏する．演奏中は，協力者の声と音楽が重なることになる．

表 4.2 に挙げた音楽フレーズを使用し，全楽曲が 3〜30 秒の間に短い繰り返しを含むものであるため，協力者の反応や状態に応じてこれらから選択した．

4.5.7　分析方法

この調査では，MusiCuddle が協力者の常同言語の発声にどう影響を与えるかを調査する．もし，MusiCuddle の音楽提示によって協力者の常同言語への注意が逸れたら，発声は止まるか言い淀むと考えられる．そこで，音楽の有無による協力者の発声について比較することとし，特に，注意が逸れたときを見つけるために協力者の言い淀みに着目した．

協力者の発声の繰り返しパターンを使って，発声を意味内容によって分割することとし，この分割の変化を分析する．以下に元の発声および分割結果の一例を示す．

- 発声：「いませんですいませんですひとやすみまずやすみいませんですよ」
- 分割結果：「いませんです　いませんです　ひとやすみ　まずやすみ　いませんですよ」

分析の手順については，まず，以下の基準を設定して発声されたときの音楽の有無によって分類する．

1. 音楽再生中に発声されたとき：　音楽があるときの発声
2. 音楽が終わった直後に発声が始まったとき：　音楽があるときの発声開始
3. 発声が始まってから音楽が再生されたとき：　音楽がないときの発声
4. それ以外：　音楽がないときの発声

次に，協力者の言い淀みからセンテンスを見つけ出す．協力者は，4.5.2 節で記したように，わずかな変化もなくいくつかの常同言語を繰り返すことがある．しかし，常同言語の発声への注意が逸れると言い淀み，常同言語の一部のみを発声するか，常同言語ではない発声をする．そこで，直前に発声した語の一部を含んだ語が発見されたら，そこを言い淀みとした．以下の例では，「いま」は直前の「いませんです」の繰り返しにおける言い淀みと判定する．

- 「いませんです　いま」

4.6　ケーススタディの結果

4.6.1　収録結果

収録を行った結果，協力者の声が明瞭に録音されていた3か所（16分，8分，3分）について分析することとした．計27分間のうち，音楽が提示されていたのは6分54秒であった．1回目と2回目の収録では，MusiCuddleを使用し，3回目では使用していない．

4.6.2　提示した楽曲

表4.3に曲の種類，時間，提示時間を示す．

1回目では6個の四和音を提示した．四和音は，協力者に不安定なイメージを伝えられる．ここではメジャー7th，クォーターノート，音量変化なし　からなる四和音のうちの一つを使用した．次に「雪」を4回提示した．このとき協力者は，最後の「ずんずんつもる」の部分をメロディに合わせてロずさんだ．続いて「赤い靴」を5回提示した後，5つのカデンツを提示した．どのカデンツを演奏するかは，協力者の発声から取得した音高に基づく．その後，協力者が再びロずさんでいたので，もう一度「雪」を提示し，最後に「花」を提示した．

2回目は,「月の砂漠」を4回,「雪」を9回提示したが，協力者は一緒に歌うことはなかった．次に協力者の「いませんですよ」の発声から作ったフレーズを5回流した．最後に「花」を2回流した．

フレーズの開始音，つまり患者の発声した音高は，49回がD3〜D4の収まっていたが，1回だけG＃4であった．このとき，患者は空腹によっていらだっていた．

4.6.3　発話の分類

協力者は収録中の多くの時間，発声を行っており，計27分間で680の発話に分割された．もっとも多かったのは「いませんです」の201回であり，言葉は多少違うが内容は似ている発話が多くあった.「いま」や「まず」などの短いものを除くと，多くの発話が4／4拍子に則ったリズミカルなものであった．

表 4.3: 演奏したフレーズ

曲のタイプ	曲	時間（秒）	回数	開始時の音高	合計時間（秒）
4 つ連ねた和音	長 7 度 かつ 四分音符 かつ 音量変化無し	3	6	D3(2),D#3,F3,A3,C4	18
カデンツ	No.22 Fugue (B)	8	1	C4	8
	No.1 Fugue (B)	8	1	F3	8
	No.15 Fugue (B)	11	1	F#3	11
	No.22 Prelude (S)	15	1	D4	15
	No.3 Fugue (S)	25	1	F#3	25
唱歌	「雪」	9	21	D#3, E3(3),F3(7),G3(2),G#3,A3(3),A#3,B3,C4,D4	189
	「赤い靴」	10	5	D3,F3,F#3,G#3,A#3	50
	「花」	8	4	D3,G#3,A#3,G#4	32
	「月の沙漠」	11	4	F#3,F3, C4,C3	44
常同言語のメロディ	い-ま-せ-ん-で-す-よ	3	5	G3,G#3,A3(2),A#3	15
合計	–	–	50	–	414

注 :「カデンツ」の欄の”(B)” は，Bach 作曲の曲集の 1 曲であることを示し，”(S)” は，ショスタコーヴィチ作曲の曲集の 1 曲であることを示す.「開始時の音高」で，C, D, E〜A, B は，ド，レ，ミ〜ラ，シを意味する. ”A4” は，約 440Hz の音高である. 括弧内の数字は，その音高で開始した回数を示す.

表 4.4 は音楽の有無による結果の比較である．直前の発話と異なる発話がなされた数は，音楽有りで 114 回，無しで 179 回であり，時間に占める割合では，音楽有りで 16 %，無しで 9 %であった．しかし，直前の発話と異なるがその一部を含んでいたものについて調べると，音楽有りで 114 回中の 94 回（82.5 %），音楽無しで 179 回中の 74 回（41.3 %）となり，割合で 2 倍の差となった.

以下は，音楽の提示無しのときに見られた，直前の発話と全く異なった発話に切り替わった例である.

・まず　いませんです　おやつじゃ　まずやすみ

また，以下は発話途中で音楽の提示をしたときに見られた，直前の発話の一部を含んだ発話に切り替わった例である（「ん」の位置で音楽を提示）.

・まずごはんです まず...

4.7　考察

MusiCuddle は，Iso-principle に基づいて，患者の気持ちに寄り添った音楽の演奏がされることを目標としている.

このケーススタディでは，提示された多くの楽曲の開始音が D3〜D4 であり，協

表 4.4: 変化した発話の数

	音楽有り	音楽無し
変化した発話（すべて）	114 個	179 個
直前の発話の一部を含んだもの	94 個	74 個
割合	82.5 %	41.3 %

力者の発声から得た音高は様々かつ広い音域であった．開始音の音高は，曲調全体に影響や変化を与え，また，調性や音高は曲調を決定する大事な要素であるため，MusiCuddle は患者の声から音高を取得し，Iso-principle に基づいた患者の精神状態に寄り添った音楽の演奏が可能であると考えられる．

また，MusiCuddle が提示する音楽は，患者が常同言語をやめるきっかけを与える可能性が示唆された．調査協力者の常同言語は，歌であるかのように 4 ／ 4 拍子に則ったリズミカルなものであった．しかし，音楽の提示によって常同言語に言い淀みが起こる傾向が見られた．演奏される音楽フレーズは，協力者の声から取得した音高に基づいて決まるが，通常，そのフレーズは協力者の常同言語とは異なる．一方で，予備調査で述べたように，調査実施者が協力者が繰り返す言葉と同じ言葉を一緒に発声したとき，協力者は調査実施者に注意を向けつつもなお発声を繰り返していた．

MusiCuddle は，患者の常同言語の一部に合致している音楽を提示しつつ，同時にその音楽は，患者の常同言語パターン全体とは異なったものである．従って，患者は，常同言語とその音楽との同質性によって注意を惹きつけられると同時に，異質性によって今までの常同言語パターンから注意をそらされてしまう，と推察される．

この作用を用いて，MusiCuddle は，常同言語を繰り返す患者の精神状態への寄り添いを行いながらも，発声を繰り返す患者を膠着状態から脱却させられる可能性がある．

4.8 おわりに

本章では，2 章で提案した技術が持つ，人間による音の取捨選択が可能な特長を用いて，他者が自由に発する非音楽的な音声が入力音となる課題に提案技術の適用を試みた．認知症患者の常同言語の繰り返しなどを落ち着かせることを目的として，音楽的には初心者である介護者が音楽療法に携わることを支援するシステム“ MusiCuddle ”を提案した．

MusiCuddle を用いて音楽フレーズを演奏することにより，患者の行動がどう変わったかを評価するため，ケーススタディとして患者 1 名に対する調査を行った．患者の発話の変化を調べたところ，音楽が提示された時は，言い淀む傾向が見られ，直前の発話の一部を含んだ新しい発話を発声する傾向が見られた．患者は，常同言語とその音楽との同質性によって注意を惹きつけられると同時に，異質性によって今までの常同言語パターンから注意をそらされた，と推察されることから，MusiCuddle が，常同言語を繰り返す患者の精神状態への寄り添いを行いながらも，患者が常同言語の繰り返しから抜け出すきっかけを与える可能性を示している．

「認知症」の症状やさまざまであり，患者の性格などもさまざまである．従って，ケーススタディを蓄積しながら開発を進めることが課題となる．

第 5章

本論文のまとめ

本論文では，音の時系列の中から人間が「価値がある」と判断した区間について情報，特に音楽で重要な音高を取得するための，人間と計算機との協調的な音高判定技術の構築と評価を行った．まず，人間による音区間の取捨選択を反映させる仕組みを実装した基盤システムを構築し，評価した．次に，人間による音の取捨選択が可能な特長をもって解決可能な課題として，Voice-to-MIDI システムの操作者自身の声以外を入力音とする事例を 2 例採り上げ，提案技術の適用を試みた．

第 2 章では，第 3 章と第 4 章に先立って，人間のタップによる音区切り情報の入力と計算機の音高抽出を用いた協調的な音高判定技術を提案した．また，提案技術をタップ併用型 Voice-to-MIDI システムとして実装し，評価を行った．このシステムは Voice-to-MIDI を応用し，マイクからの入力音と同時に，人間がコンピュータや電子楽器のキーボードなどからリアルタイムに音区間の区切りをタップ入力し，計算機はその区間の音高を取得する仕組みを持つ．このシステムは，従来の計算機の自動処理による Voice-to-MIDI システムと比較すると，人間の介入を増やすことによってさらに多くの情報を与えることができるため，より人間と計算機が協調して音符への変換を行うことができる Voice-to-MIDI システムと言える．評価実験では，Voice-to-MIDI 技術の課題であった音高と音数の判定精度向上の課題について評価した．また，楽器経験の有無やタップの有無の変換精度への影響の評価を行った．9 名の被験者に歌唱しながらそのフレーズのリズム区切りを入力させ，歌詞歌唱などの任意の発音の歌唱を許容する既存 Voice-to-MIDI システムとの変換精度の比較によって評価を行った．

その結果タップ併用型 Voice-to-MIDI システムは，既存の歌詞歌唱などの任意の発音の歌唱を許容するシステムに比べて，欠落する音や不要な音の発生が抑制され，音数および音高判定精度が向上することが示された．楽器経験の有無のタップへの影響については，赤とんぼレベルの曲であれば，多少速いテンポの入力であっても大きく影響しないと見られることが分かった．また，タップの有無の歌唱への影響についても，赤とんぼレベルの曲の場合，入力テンポが速くなると多少影響が出る可能性があるものの，必ずしもタップが悪影響を及ぼすわけではなく，総じてタップの有無の影響は小さいことが分かった．我々は，これまでに文献 [69] において市販の「タタタ歌唱」システムに自由歌唱を入力して比較実験を行っており,「タタタ歌唱」を必要とするシステムに対する優位性を示していた．この結果と合わせて，提案手法は，Voice-to-MIDI 技術の課題であった音高と音数の判定精度向上の課題に対して有用であると考えられる．また，マイクからの入力音と同時に，人間がコンピュータや電子楽器のキーボードなどからリアルタイムに音区間の区切りをタップ入力可能であることを確認した．

第 3 章では，環境に応じて得られるあらゆる音の時系列に対して，ユーザ自身の心象風景を音の取捨選択という形で織り込んで環境の音情景を再構成するための新しいリズム楽器システム“ EnvJamm ”の提案と評価を行った．EnvJamm は，第 2 章で実装したタップ併用型 Voice-to-MIDI システムに対して，検知できる音域が拡張されている．ユーザが身を置く環境から五感への多様な刺激を受け，これに触発されて様々に変容する自らの内的心象を反映したリズムを刻む行為によって，環境から得られる音に自らの心象風景を織り込んで環境の音情景を再構成し，環境と人の表現を融合した音楽を作成することを目的とする．加えて，創作を通じた環境のありように対する新たな気づきや視点の変化の促進を目的とする．

評価実験では，被験者 1 名による野外における試用とシステムをよく理解している著者による試用による評価を行った．その結果，コメントより，被験者自身の意図が上手く織り込まれたフレーズが記録できたことが伺え，環境から得られる音に被験者の心象風景を織り込んで環境の音情景を再構成できたと考えられる．また，EnvJamm を用いることによって，身の回りの音を注意深く聴いたり，周辺環境において注目するモノを視覚的に探索する等の行為，つまり環境からの刺激を十分に受け取るために必要な「感覚の研ぎ澄まし」の行為を自然に醸成できる

可能性が伺えた．システムに詳しい著者による評価では，単発の音やリズムを持った音，高い音や低い音など様々な音が紡ぎ出した音風景を活かしたフレーズを作成することによって，音風景を再構成できた．また，先述の被験者の試用評価と同様，身の回りの音を注意深く聴くなどの「感覚の研ぎ澄まし」の行為が見られたのに加えて，環境からの「触発」も感じられた．

以上から，EnvJamm によって環境の音情景の再構成を通じた環境と人の表現を融合した音楽の作成，および環境のありように対する新たな気づきや視点の変化の促進という目的は達成できたと考える．課題として，単音のみの入力であるなど現在のシステムでは場面により不満があるため，今後改良を検討したい．また PC キーボード以外のタップ入力デバイスによるタップ入力評価を行いたい．

第 4 章では，他者が自由に発する非音楽的な音声に対して音を取捨選択する必要がある課題として，認知症患者に対する音楽療法を支援するシステム“MusiCuddle”を提案し，評価を行った．患者が，不安などの心的状態から生じる症状の一つである，何度も同じ言葉を繰り返す行為（「常同言語」と呼ぶ）から抜け出させたり，落ち着かせるために，MusiCuddle は患者の発声から得た音高に応じた音楽フレーズを操作者の指示によって自動演奏し，患者に聴かせるという仕組みを持つ．操作者は，タップ入力によって音高を取得する区間を指定する．音楽フレーズには，Iso-principle に基づいた，患者の精神状態に寄り添えると思われるものなどを使用する．評価は，認知症患者 1 名に対して試用を行うケーススタディの形で行った．

このケーススタディの結果，MusiCuddle は，患者の声から取得した音高を使用して，Iso-principle に基づいた患者の精神状態に寄り添った音楽の演奏が可能であると考えられる．

また，調査協力者の常同言語は，拍子に則ったリズミカルなものであったが，MusiCuddle による音楽の提示によってその常同言語に言い淀みが起こる傾向が見られた．従って，MusiCuddle が提示する音楽は，患者が常同言語をやめるきっかけを与える可能性が示唆された．

MusiCuddle は，患者の常同言語の一部に合致している音楽を提示しつつ，同時にその音楽は，患者の常同言語パターン全体とは異なったものである．そのため患者は，常同言語とその音楽との同質性によって注意を惹きつけられると同時に，異質性によって今までの常同言語パターンから注意をそらされてしまう，と推察

される．

これらの結果から，MusiCuddle は，常同言語を繰り返す患者の精神状態への寄り添いを行いながらも，発声を繰り返す患者を膠着状態から脱却させられる可能性が考えられる．

以上，各章の成果を通じて，本論文で提案した人間と計算機との協調的な音高判定技術は，Voice-to-MIDI の従来からの課題である音高と音数の判定精度向上に対する有用性が示された他，Voice-to-MIDI の応用範囲拡大への寄与が示された．

次に，本論文の成果を人間と計算機との協調処理の観点から俯瞰する．本論文は，人間と計算機がそれぞれの得意・不得意を相補えるような役割分担を構築することが，協調処理を有効に作用させる一つの方策であることを示した．もしこの役割分担が，人間が音高取得、計算機が音区切りであったとすると，互いの不得意としやすい処理の組み合わせとなり，協調処理は破綻していたと推察される．今後は，人間と計算機がそれぞれの得意・不得意を相補えるような役割分担が人間と計算機との協調処理課題全般について有用であるか検討してゆきたい．

最後に，本論文の成果を知識科学の観点から俯瞰する．

まず，本論文で提案した人間と計算機との協調的な音高判定技術は，第 2 章と第 3 章で示されたように，創造的活動のためのユニバーサルメディアの一つとして，より多くの人間に対して容易かつ的確な音楽の表出や創造を支援可能にすると考えられる．従って，音楽というメディアに対する知識創造の生産性向上に貢献したと言える．また，より多くの介護者が音楽療法に携わることおよび認知症患者が音楽療法を受ける機会の増加を目的とした研究の端緒として，第 4 章で示されたように，認知症患者が常同言語の繰り返しから抜け出す過程や方法の解明を支援する技術と捉えられる．今後この過程や方法の解明と知識体系化および当初の目的の達成によって，提案技術の更なる貢献が望まれる．

以上のような知識創造や認知科学などの分野への貢献を足掛かりとして，本論文の提案手法ならびに成果がより広く浸透し，知識社会の拡大・質の向上の一助となることを願い，この論文を終えたい．

謝辞

本研究を行うに当たり, 長期に渡って終始丁寧な御指導を賜わった北陸先端科学技術大学院大学　ライフスタイルデザイン研究センター　西本 一志 教授に深謝致します.

知的クラスタをはじめ各種方面で大変お世話になりました，北陸先端科学技術大学院大学　知識科学研究科　國藤 進 教授，藤波 努 准教授，ならびに北陸先端科学技術大学院大学　ライフスタイルデザイン研究センター　金井 秀明 准教授に感謝の意を表します.

忙しい中，本博士論文の審査を引き受けて下さいました神戸大学大学院工学研究科電気電子工学専攻　寺田 努 准教授，北陸先端科学技術大学院大学　知識科学研究科　宮田 一乗 教授，北陸先端科学技術大学院大学　知識科学研究科　永井 由佳里 教授に感謝の意を表します.

私の副テーマ指導を引き受けて下さり，指導していただきました，ATR メディア情報科学研究所の苗村 昌秀 研究員（現　NHK 放送技術研究所）に感謝の意を表します.

私の副テーマ指導ばかりではなく，私の提案技術を大変素晴らしい研究へと昇華して下さいました本学 OG でもある大島 千佳 研究員（現　佐賀大学）に感謝の意を表します.

本学入学へのきっかけを作って下さったばかりか，大学時代から今に至るまでずっと私のことを気遣っていただいた神奈川工科大学　徳弘 一路 准教授に深謝の意を捧げます.

日頃から有益な御助言をいただき, 多面に渡って励ましていただいた西本研究室の諸兄に厚く御礼申し上げます. 中でも，OB の宮下 芳明 准教授（現　明治大学），OG の小倉 加奈代 助教（現 北陸先端科学技術大学院大学），所属当時から現在に至るまでずっと一緒であり，苦楽をともにした千葉 慶人，小林 智也 両君

そして，林研究室 OB の小野 泰正 君，橋本研究室 OB の小林 重人 博士（現 北陸先端科学技術大学院大学），池田研究室 OB の小川 泰右 博士（現 北陸先端科学技術大学院大学）に感謝致します.

研究についての有益な意見をいただいた産業技術総合研究所 後藤 真孝 博士に心よりお礼を申し上げます.

学科が違うにも関わらず会うたびに私のことを気遣っていただいた本学 OB の齋藤 毅 博士（現金沢大学）に心よりお礼を申し上げます.

忙しい中時間を割いて実験に協力していただいた被験者の方に感謝致します.

2 年に渡り私の活動を支援していただきました石淵 耕一 代表取締役社長をはじめ，インターメディアプランニング株式会社の皆様に心よりお礼を申し上げます.

最後に，今日まで私を辛抱強く支えてくださった父と母に深謝の意を捧げます.

参考文献

[1] MIDI Manufacturers Association Incorporated http://www.midi.org/（2010年5月現在）

[2] 加藤 充美，“MIDI規格誕生の背景と規格の概要：電子音楽をとりまく環境の変化 (小特集 MIDI規格がもたらしたものと今後の展望)，”日本音響学会誌 64(3), pp.158-163, 2008.

[3] 高橋 信之，コンプリートMIDIブック，株式会社リットーミュージック，ISBN4-845-61151-1，2005.

[4] Avid, Protools.

[5] Steinberg Media Technologies GmbH, Cubase 7.

[6] Sony Creative Software, ACID Pro 7.

[7] Celemony Software GmbH, Melodyne.

[8] 川野 洋，芸術・記号・情報，勁草書房，pp.94-95，1982.

[9] 宮崎 謙一，“「絶対音感」はどこまで分かったか?(小特集 音楽音響における最近の話題) ”，日本音響学会誌 60(11), pp.682-688, 2004.

[10] 株式会社インターネット, Singer Song Writer 9 Professional.

[11] MakeMusic, Inc., Finale 2010, 2009.

[12] YAMAHA株式会社，XGworks ST，浜松，2003.

[13] 株式会社河合楽器製作所，Band Producer 2，浜松，2008.

[14] 新原 高水，今井 正和，井口 征士，“歌唱の自動採譜，”計測自動制御学会論文集，20，10，pp.68-73，1984.

[15] L. Prechelt and R. Typke, "An interface for melody input," ACM Trans. on Computer-Human Interaction (TOCHI), Vol.8, No.2, pp.133-149, 2001.

[16] C. C. Toh, B. Zhang, Y. Wang, "Multiple-Feature Fusion Based Onset Detection for Solo Singing Voice," Proc. of ISMIR 2008, 2008.

[17] M. Ryynanen, A. Klapuri, "Modelling of Note Events for Singing Transcription," Proc. of ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio, 2004.

[18] Lutz P.，Rainer T.: "An Interface for melody input," ACM Trans. on Computer-Human Interaction（TOCHI），Vol.8，No.2，pp133-149，2001.

[19] Alexandra U.，Justin Z.: "Melodic matching techniques for large music databases," Proc. of the seventh ACM int. conf. on Multimedia，MULTIMEDIA ’99，pp57-66，1999.

[20] A. Ghias, J. Logan, D. Chamberlin and B. C. Smith: "Query by humming: Musical Information Retrieval in an Audio Database," Proc. of ACM Multimedia'95, San Francisco, California, Nov. 1995.

[21] A. Uitdenbogerd and J. Zobel, "Melodic matching techniques for large music databases," Proc. of the 7th ACM intl. conf. on Multimedia, pp.57-66, Orlando, Florida, 1999.

[22] 清水 純，丸山 剛志，三浦 雅展，柳田 益造，“ハミングによる単旋律の自動採譜，”音響学音楽音響研資，Vol.23，No.5，pp.95-100，2004.

[23] M. Ryynanen, A. Klapuri, "Automatic Transcription of Melody, Bass Line, and Chords in Polyphonic Music," Computer Music Journal, 32(3), pp.73-86, 2008.

[24] 波多野 誼余夫　編，音楽と認知，東京大学出版会，pp.42-43，ISBN4-13-013062-5.

[25] SongTapper.com, http://www.songtapper.com/public_html/（2010 年 5 月現在）

[26] Air Guitar World Championships 2010,

http://www.airguitarworldchampionships.com/en/event/agwc-2010（2010 年 5 月現在）

[27] エアギタージャパン，http://airguitar.jp/news/（2010 年 5 月現在）

[28] 大島 千佳，宮川 洋平，西本 一志: “Coloring-in Piano : 表情付けに専念できるピアノの提案,” 情処研報 2001-MUS-42，Vol.2001，No.103，pp.69-74，2001.

[29] 大島 千佳，宮川 洋平，西本 一志: “Coloring-in Piano による 2 ステップ打ち込みの提案,” 情処研報 2001-MUS-43，Vol.2001，No.104，pp.21-26，2001.

[30] 山田 真司，南野 千鶴: “安室奈美恵の「あとノリ」 : Never End の分析結果，” 日本音響学会春季研究発表会講論集，pp.865-866, 2003.

[31] 奥平 啓太，平田 圭二，片寄 晴弘: “ポップス系ドラム演奏の打点時刻及び音量とグルーブ感の関連について，” 情処研報 2004-MUS-55，pp.21-26，2004.

[32] 藤原義章，“機械的リズム，” リズムはゆらぐ，pp.74-76，株式会社白水社，ISBN4-560-03687-X，1990.

[33] 藤原義章，“自然リズム，” リズムはゆらぐ，pp.76-78，株式会社白水社，ISBN4-560-03687-X，1990.

[34] 半田伊吹，木下智義，武藤誠，坂井修一，田中英彦，“ マン・マシン協調による採譜システム，” 情処学音楽情報科学研報，MUS-34，pp.21-26，1999.

[35] Z. Obrenovic, “ A flexible system for creating music while interacting with the computer,” Proc. of the 13th annual ACM intl. conf. on Multimedia, pp.996-1004, 2005.

[36] 原田 宏美，DTM で学ぶオーケストレーション入門，株式会社音楽之友社，ISBN4-276-10615-X，2002.

[37] 松浦 あゆみ，ブラス&ストリングスアレンジ自由自在，株式会社リットーミュージック，ISBN4-845-61086-8，2004.

[38] 高橋 信之，コンプリート MIDI プログラミング・ブック，株式会社リットーミュージック，ISBN4-845-61326-3，2006.

[39] 米谷 知己，XGworks・SOL ユーザーへ捧ぐ　リアルな MIDI の作り方教えます!!，株式会社ヤマハミュージックメディア，ISBN4-636-16025-8，2005.

[40] RENCON, http://www.renconmusic.org/（2010 年 5 月現在）

[41] Vocaloid2，http://www.vocaloid.com/（2010 年 5 月現在）

[42] クリプトン・フューチャー・メディア株式会社，初音ミク，2007, http://www.crypton.co.jp/mp/pages/prod/vocaloid/（2010 年 5 月現在）

[43] 中野 倫靖，後藤 真孝: “VocaListener : ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案”，情処研報, MUS (50), pp.49-56, 2008.

[44] MIDI Espressivo，http://magarchive.halfmoon.jp/vector/midiexp/（2010 年 5 月現在）

[45] M. Droettboom, I. Fujinaga, K. MacMillan, G. S. Chouhury, T. DiLauro, M. Patton, T. Anderson: “Using the Gamera framework for the recognition of cultural heritage materials,” Proc. of 2nd ACM/IEEE-CS joint conf. on Digital libraries, pp.11-17, Portland, Oregon, USA, 2002.

[46] MusicLab, Inc., RealGuitar 2, 2006.

[47] MusicLab, Inc., RealStrat, 2007.

[48] Arturia, BRASS, 2005.

[49] Applied Acoustics Systems DVM Inc., Strum Electric GS-1, http://www.applied-acoustics.com/strumelectric/overview/（2010 年 5 月現在）

[50] 長嶋 洋一: “マルチメディアコンテンツのための自動作曲システム”，静岡文化芸術大学研究紀要，6, pp.49-58, 2005.

[51] 平野 砂峰旅: “メディアインスタレーション ”Movement ”のインタラクションについて，” 情処研報，ヒューマンインタフェース研究会報告 2000(94), pp.39-42, 2000.

[52] 安藤 大地 , Dahlstedt Palle , Nordahl Mats , 伊庭 斉志: “対話型進化論的計算による作曲支援システム : CACIE,” 情処研報. 音楽情報科学 2005(14), pp.55-60, 2005.

[53] 荒谷 綾太，蓮井 洋志: “感性データベースを利用した自動作曲システムの実現,” 情処研報，音楽情報科学 2008(78), pp.173-178, 2008.

[54] 株式会社河合楽器製作所, スコアメーカー FX4.

[55] 野口 義修，“ 詞先・メロ先作曲術，” 作曲本，pp.109-118，株式会社シンコーミュージック・エンタテインメント，東京，2005.

[56] 奥平 ともあき，“ 詞先・曲先，” 誰にでもできる作曲講座，pp.20-21，株式会社ドレミ楽譜出版社，東京，2003.

[57] 齋藤 毅, 鵜木 祐史, 赤木 正人: “歌声の基本周波数変化に含まれるオーバーシュートの知覚への影響に関する検討”，日本音響学会，聴覚研資 36(7), pp.611-616, 2006.

[58] 齋藤 毅, 後藤 真孝, 鵜木 祐史, 赤木 正人: “SingBySpeaking : 歌声知覚に重要な音響特徴を制御して話声を歌声に変換するシステム”，情処研報. MUS 36(7), pp.25-32, 2008.

[59] 番弘光，伊藤克亘，武田一哉，板倉文忠，“ タッピングを利用した音声認識の検討，” 情処学音声言語情報処理研報，SLP-47，pp71-76，2003.

[60] 岩田憲治，渡邉康司，中川竜太，篠田浩一，古井貞熙，“音声とペンの準同期入力に対するマルチモーダル認識，”音響学2006年秋季講論集，1-2-23，2006.

[61] L. A. Smith, E. F. Chiu and B. L. Scott, "A Speech Interface for Building Musical Score Collections," Proc. of the 5th ACM conf. on Digital libraries, San Antonio, Texas, 2000.

[62] Wildcat Canyon Software Inc.: Autoscore 2.0, 1999.

[63] G. Haus and E. Pollastri, "An Audio Front End for Query-by-Humming Systems," Proc. of the 2nd Intl. Symp. on Music Information Retrieval, pp.65-72, Indiana, USA, Oct. 2001.

[64] 来海 大輔，江村 伯夫，三浦 雅展，柳田 益造: “音高・音価テンプレートを用いた単音節歌唱の採譜精度の向上，”日本音響学会，音響研資 MA2007-73, Vol.26, No.6, pp.99-104, 2007.

[65] N. Kosugi, Y. Nishihara, T. Sakata, M. Yamamuro and K. Kushima, "A practical query-by-humming system for a large music database," Proc. of the 8th ACM intl. conf. on Multimedia, pp.333-342, Marina del Rey, California, 2000.

[66] 野ばら社編集部，童謡，野ばら社編集部，p68，株式会社野ばら社，東京，1994.

[67] 原裕一郎，井口征士，“複素スペクトルを用いた周波数同定，”計測自動制御学会，pp718-723，1983.

[68] 伊藤直樹，西本一志，“MIDI シーケンスデータの2step打ち込み法への鼻歌による音高入力の適用，”情処学エンタテインメントコンピューティング研報，2006-EC-5，Vol.2006，pp.43-48，2006.

[69] N. Itou and K. Nishimoto, "A voice-to-MIDI system for singing melodies with lyrics," Proc. of the intl. conf. on ACE'07, pp.183-189, Salzburg, Austria, 2007.

[70] 金澤正剛　監修,“記号表,”新編　音楽小辞典，p.439，株式会社音楽之友社，東京，2004.

[71] R. マリー・シェーファー，世界の調律　The Tuning of the World，鳥越 けい子，小川 博司，庄野 泰子，田中 直子，若尾 裕 訳，平凡社，ISBN4-582-76575-0，2006.

[72] R. マリー・シェーファー,“サウンドスケープの諸特徴,”世界の調律　The Tuning of the World，鳥越 けい子，小川 博司，庄野 泰子，田中 直子，若尾 裕 訳，平凡社，p36, ISBN4-582-76575-0，2006.

[73] R. マリー・シェーファー,“音響生態学とサウンドスケープ・デザイン,”世界の調律　The Tuning of the World，鳥越 けい子，小川 博司，庄野 泰子，田中 直子，若尾 裕 訳，平凡社，p414, ISBN4-582-76575-0，2006.

[74] 大蔵 康義,“現代の音楽,”音と音楽の基礎知識，pp.219-221，株式会社国書刊行会，1999.

[75] Karmen Franinovic, Yon Visell: Recycled soundscapes, Proc. of the 5th conf. on Designing interactive systems, p.317, 2004.

[76] Recycled Soundscapes: http://www.zero-th.org/RecycledSound.html（2010年5月現在）

[77] 宮下芳明, 西本一志:「音楽の条件」とは何か？, 日本バーチャルリアリティ学会論文誌, Vol.10, No.1, pp.11-20, 2005.

[78] プロ・オーディオ・ジャパン株式会社，MPC シリーズ，http://www.akai-pro.jp/products.php（2010 年 5 月現在）

[79] ローランド株式会社，SP-404SX，http://www.roland.co.jp/products/jp/SP-404SX/index.html（2010 年 5 月現在）

[80] Native Instruments GmbH., KONTAKT 4, http://www.native-instruments.com/#/jp/products/producer/kontakt-4/

[81] I. M. Altshuler, "The past, present and future of musical therapy," Music therapy, Podolsky, E. Eds. Philosophical Library, pp.24-35, 1954.

[82] Clark, M.E.: "Use of Music to Decrease Aggressive behaviors in People with Dementia," J. Gerontological Nursing, pp.10-17, July, 1998.

[83] Ragneskog, H., Kighlgren, M.: "Music and Other Strategies to Improve the Care of Agitated Patients with Dementia," Scand J. Caring Sci., 11, pp.176-182, 1997.

[84] H. B. Svansdottir and J. Snaedal, "Music therapy in moderate and severe dementia of Alzheimer's type: a case-control study, International Psychogeriatrics," Int Psychogeriatr. vol. 18(4), pp.613-621, 2006.

[85] H. Park and J. K. Pringle Specht, "Effect of individualized music on agitation in individuals with dementia who live at home," J Gerontol Nurs, Vol. 35(8), pp.47-55, 2009.

[86] B. K. Nair, C. Heim, C. Krishnan, C. D' Este, J. Marley and J. Attia, "The effect of Baroque music on behavioural disturbances in patients with dementia," Australasian Journal on Ageing, vol. 30(1), pp.11-15, 2001.

[87] D. E. Michel and J. Pinson, "Music Therapy In Principle And Practice," Charles C Thomas Publisher, 2005.

[88] D. Grocke, and T. Wigram, "Receptive Methods in Music Therapy: Techniques and Clinical Applications for Music Therapy Clinicians, Educators and Students," Jessica Kingsley Publishers, 2007.

[89] Y. Imai and K. Hasegawa, "The Revised Hasegawa's Dementia Scale (HDS-R) - evaluation of its usefulness as a screening test for dementia," Journal Hong Kong Coll Psychiatr, vol. 4, pp.20-24, 1994.

[90] 若尾，岡崎：音楽療法のための即興演奏ハンドブック，音楽之友社，1996.

[91] S. Ashida, “The effect of reminiscence music therapy sessions on changes in depressive symptoms in elderly persons with dementia,” Journal of Music Therapy, American Music Therapy Association, vol. 37(3), pp.170-182, 2000.

[92] BEST SERVICE, CHRIS HEIN GUITARS.

[93] ローランド株式会社，GK シリーズ，http://www.roland.co.jp/GK/

[94] ヤマハ株式会社, G50, http://www.yamaha.co.jp/product/syndtm/p/cont/g1b1g50/

[95] Godin，http://www.jes1988.com/godin/index.html

[96] Jordi J.: Voice-controlled plucked bass guitar through two synthesis techniques, Proc. of the 5th int. conf. on NIME2005, pp.132-135, Vancouver, Canada, 2005.

[97] Karjalainen M., Maki-Patola T., Kanerva A., Huovilainen A., Janis P.: Virtual Air Guitar, Proc. of the 117th Audio Engineering Soc. Conv. San Francisco, CA, USA, October 28-31, 2004.

[98] Junichi K., James G., Dr. Laurent M.: Mountain Guitar: a Musical Instrument for Everyone, Proc. of the 7th int. conf. on NIME2007, pp.396-397, 2007.

[99] ヤマハ株式会社: EZ-AG, http://www.yamaha.co.jp/product/epiano-keyboard/ez-ag/index.html

[100] ヤマハ株式会社: EZ-EG, http://www.yamaha.co.jp/ez/product/ez-eg/index.php

本研究に関する発表論文

学術誌掲載論文（査読あり）

[1] 伊藤 直樹，西本 一志：“Voice-to-MIDI のためのメロディリズムタップを用いた音数・音高の判定手法の提案”，電子情報通信学会論文誌 D Vol.J96-D No.4 pp.965-977, 2013.

[2] Chika Oshima, Naoki Itou, Kazushi Nishimoto, Naohito Hosoi, Kiyoshi Yasuda, and Koichi Nakayama: “An Accompaniment System for Healing Emotions of Patients with Dementia Who Repeat Stereotypical Utterances,” Lecture Notes in Computer Science, B. Abdulrazak et al.(Eds.), Springer, Vol. 6719, pp. 65-71, 2011.

[3] Chika Oshima, Naoki Itou, Kazushi Nishimoto, Kiyoshi Yasuda, Naohito Hosoi, Hiromi Yamashita, Koichi Nakayama, Etsuo Horikawa: “A Music Therapy System for Patients with Dementia who Repeat Stereotypical Utterances,” Journal of Information Processing, Vol. 21, No. 2 pp. 283-294, 2013.

国際会議発表論文（査読あり）

[4] Naoki Itou, Kazushi Nishimoto: “A voice-to-MIDI system for singing melodies with lyrics,” Proc. of the int. conf. on ACE’07, pp.183-189, Salzburg, Austria, 2007.

[5] Chika Oshima, Naoki Itou, Kazushi Nishimoto, Kiyoshi Yasuda, Naohito Hosoi, Hiromi Yamashita, Koichi Nakayama, and Etuso Horikawa: “A Case Study of a Practical Use of “ MusiCuddle ”that is a Music Therapy System

for Patients with Dementia who Repeat Stereotypical Utterances," Global Health 2012, October 21-26, Venice, Italy, 2012.

査読付き国内学会論文

[6] 伊藤 直樹，西本 一志: "歌詞歌唱による入力が可能な Voice-to-MIDI 手法の提案"，インタラクション 2007 論文集，pp.71-72, 2007.

[7] 伊藤 直樹，西本 一志: "メロディリズムのタップを併用する Voice-to-MIDI 変換手法の音高変換精度評価"，インタラクション 2010 論文集，pp.143-150，2010.

その他

[8] 伊藤 直樹，西本 一志: "MIDI シーケンスデータの 2step 打ち込み法への鼻歌による音高入力の適用"，情処研報 2006-EC-5，Vol.2006，pp.43-48，2006.

[9] 伊藤 直樹，西本 一志: "歌唱とリズムタップによるハイブリッド入力型 Voice-to-MIDI システム"，デモセッション，情処研報 2007-MUS-71，Vol.2007, 2007.

[10] 伊藤 直樹，西本 一志: "吟たぁ：ギター型インタフェイスを用いた鼻歌作曲家のためのフレーズ入力システム"，デモセッション，エンタテインメントコンピューティング 2008 予稿集，pp.169-170，2008.

[11] 伊藤 直樹，西本 一志: "吟たぁ：ギター型インタフェイスによる弾弦併用型 Voice-to-MIDI システム"，情処研報 2008-MUS-78，Vol.2008，No.127，pp.53-58，2008.

[12] 伊藤 直樹，西本 一志: "EnvJamm：環境からの触発を受けて音情景を再構成するための楽器"，エンタテインメントコンピューティング 2009 予稿集，pp.83-86，2009.

[13] 大島 千佳, 伊藤 直樹, 西本 一志, 細井 尚人, 安田 清, 中山 功一："常同言語や叫びを伴う認知症患者の感情を静穏化する音楽療法支援システムの実現に向けて"，ヒューマンインタフェース学会，第 72 回研究会，2011.

[14] 大島 千佳, 安田 清, 伊藤 直樹, 西本 一志, 細井 尚人, 中山 功一: “MusiCuddle: 常同言語で示される精神症状の緩和を目指したシステム”, 計測自動制御学会, システム・情報部門学術講演会 2011 講演論文集, 3D4-1, pp.491-496, 2011.

[15] 大島 千佳, 伊藤 直樹, 西本一志, 細井 尚人, 安田 清, 中山 功一 : “認知症患者の常同言語や発声に伴奏づけして患者の感情を静穏化するシステムの提案”, インタラクション 2011（Paper ID: 132）, IPSJ Symposium Series Vol. 2011, No. 3, pp.625-628, 2011.

[16] 大島 千佳, 中山 功一, 安田 清, 伊藤 直樹, 西本 一志, 細井 尚人, 奥村 浩 : “認知症者のための音楽療法システムの提案”, 第 25 回人工知能学会全国大会, 1A1-NFC1a, 2011.

[17] 大島 千佳, 伊藤 直樹, 西本 一志, 安田 清, 山下 ひろみ, 細井 尚人, 中山 功一, 奥村 浩, 堀川 悦夫 : “音楽により常同言語を緩和するシステムの構築に向けて”, ヒューマンインタフェース学会研究報告集, Vol.14, No.2, pp.7-12, 2012.

[18] 大島 千佳, 伊藤 直樹, 西本 一志, 安田 清, 細井 尚人, 中山 功一, 堀川 悦夫 : “常同言語を音楽により緩和するシステムの構築に向けた試用”, 人工知能学会全国大会, 2A1-NFC-6-7, 2012.

[19] 大島 千佳, 中山 功一, 伊藤 直樹, 西本 一志, 安田 清, 細井 尚人, 堀川 悦夫 : “MusiCuddle の試用による認知症者の発話の変化”, 計測自動制御学会, システム・情報部門学術講演会 2012, 2012.

本研究に関連する受賞

ICOST2011 Best Multi-Disciplinary Paper Award

[1] C. Oshima, N. Itou, K. Nishimoto, N. Hosoi, K. Yasuda and K. Nakayama, "An Accompaniment System for Healing Emotions of Patients with Dementia Who Repeat Stereotypical Utterances," Lecture Notes in Computer Science, B. Abdulrazak et al.(Eds.), Springer, Vol. 6719, pp. 65-71, 2011.

本研究に関して受けた助成金

財団法人　情報科学国際交流財団，研究者海外派遣助成，

GrantNo.2007.1.2.950，2007.

[1] N. Itou, K. Nishimoto: “A voice-to-MIDI system for singing melodies with lyrics”, Proc. of the int. conf. on ACE’07, pp.183-189, Salzburg, Austria, 2007.

第 A 章

付録．ギターフレーズ入力のための弾弦併用型 Voice-to-MIDI システム

3 章や 4 章で示した以外に，人間と計算機との協調的な音高判定技術を応用した例を付録として 1 例示す．

ギターの音高の指定と発音タイミングの指定を個別に行える特徴に着目し，フレットによる音高の指定を声で代替することによって，ギター演奏の気分を味わえることを目的とした弾弦併用型 Voice-to-MIDI システム“ 吟たぁ（Vouitar）”を提案し，評価を行う．

A.1 はじめに

近年の DTM（Desk Top Music）システムの普及により，楽器を弾けない者でも様々な楽器を用いた音楽作品の制作を楽しむことができるようになった．しかし，個々の楽器にはその楽器特有の発音の特徴があり，これをシーケンサ上での編集によって的確に再現することは極めて困難である．特にギターパートについては，MIDI 音源による再現が難しい．

この難しさには，複数の要素があり，例えば本来弦ごとに音色が違うのに一種類の音色しか容易されていなかったり，音色がエフェクトを掛けられた状態でサンプリングされている，というような音色の問題がある．また，ギターには複数の弦（通常は 6 本の弦）があり，これを弾弦して発音する．演奏の際，個々の弦

の弾弦タイミングや弾弦の強さは一様ではなく，通常ばらつきがあり，これがギターらしさを生み出す重要な要因となっており，この発音のばらつき，ピッチや音量の変化を含めたイベントを弦数分入力する手間の問題がある．

音色の問題については，パーソナルコンピュータの処理速度の進化やHDDの大容量化とともに，VSTi などのソフトウェア音源の実用性が高まり，[92][49] などの高品位のギター音源が登場し，またソフトウェアでギター向けのエフェクトを掛けることも可能となってきた．

一方で，イベント入力の手間の問題については，例えば [46] や [47] では鍵盤楽器で言うところの高音域でコードを入力し，低音域でリズムを入力するだけでコードストロークを演奏させられるようになっている．しかし，入力は簡易であるがばらつきや音量のストロークごとの「揺れ」を考慮すると編集が必要である．

このバラツキを的確かつ簡易に入力するためには，ギターそのものを演奏し，そのデータを入力することが望ましい．このため，ギターの弾弦をMIDI情報に変換し，出力できるMIDIギターシステムが市販されている [93][94][95] ．しかし，これを使うためにはギターを演奏できることが前提となり，楽器を弾けない者でも音楽制作を楽しめるというDTMの最大の魅力が損なわれてしまう．

そこで，いわゆる鼻歌入力 (Voice-to- MIDI) 機能と市販のギター型電子楽器インタフェイスを組み合わせ，歌唱から音高を，弾弦からタイミングやベロシティを取得することによって，ギターを演奏できない者でも簡易にギターらしいフレーズを入力できるシステム「吟たぁ」を提案する．また，システムデザインの基礎的問題点を洗い出すために，ギターについてよく理解していると考えられるギター上級者1名に試用してもらい評価した．

A.2　関連研究

ギターやベース，ギター型インタフェイスに着目した研究がいくつか存在する．

Jordi Janer は，ユーザの声でBassギターの弾弦を演奏・制御するための技術 [96] を提案している．歌唱を弾弦トリガーとして用いるだけではなく，歌唱の特徴抽出に基づいてどのような音を出力するかの決定も行っている．吟たぁでは，歌唱の特徴には依存させず，また出力をMIDIイベントで行うことによって，ユーザ

が音色を自由に選べる仕様としており，この点で違いがある．

近年 Air Guitar が流行っているが，Virtual Air Guitar [97] は，画像処理を用いることにより Air Guitar のようにギターがなくても演奏しているふりをするだけで，指の位置に応じて実際に発音可能な楽器である．ギターは不要となるが，指の位置と音のマッピングを理解している必要がある．

Mountain Guitar [98] は，音楽経験がなくても簡単に演奏することのできる楽器インタフェイスであり，構える高さで音の高さを指定できる，弾く真似をすることによりコード演奏が可能であるなどの特徴がある．本研究とは機構や楽器かフレーズ入力インタフェイスかの違いはあるがフレーズ入力用途への応用可能性はある．

A.3　提案システムの概要

A.3.1　動作モードおよび操作方法

吟たぁは，ギター型電子楽器インタフェイスから出力された弾弦情報から NoteOn タイミングと On ベロシティを取得し，Voice-to-MIDI 機能により弾弦中の歌唱の音高を取得し，これらをマージして MIDI シーケンスを作成する，という動作を行う．音高を歌唱から取得するためユーザは出したい音とその楽器上のマッピングを覚えずとも入力が可能となる．よって，ユーザは歌唱しながら各音の発声開始に合わせて弾弦を行えばよい．

現在下記のような入力モードがある．

1. 単音入力モード (単弦ピッキング[1])
2. 単音入力モード (カッティング)
3. パワーコード入力モード

となる．以下，各モードごとに詳細を説明する．

[1]英語では「弾弦」は「Plucking」とすることが多いが，本論文では日本でよく使われている「Picking」に統一した．

単音入力モード (単弦ピッキング)

入力前にあらかじめ選択した1弦の弾弦情報を取得し，それに歌唱の音高を組み合わせ，単音フレーズを入力する．確実に特定の弦を弾く必要がある．

単音入力モード (カッティング)

単音入力モード の単弦ピッキングの代わりにカッティング (複数弦を鳴らす) による弾弦を用いる．弦がカッティングされたときの最初に弾弦された弦の情報を取得し，それに歌唱の音高を組み合わせ，単音フレーズを入力する．特定の弦を弾弦しなくてよい．

どの弾弦までがひとかたまりの弾弦であるかについての判定法については，全ての弾弦がカッティングでなされることを前提として，カッティングのダウン時には弦番号が単調減少し，アップ時には弦番号は単調増加することを利用し，弦番号がひとつ前に弾弦された弦の番号と同じか不連続，あるいは単調減少→単調増加，単調増加→単調減少への移行があったときに新規の弾弦群と判断するルールとした．

パワーコード入力モード

常にパワーコード (基本的にはルート音と完全5度上の音から成り立つギター特有の和音) フレーズのルートを歌っていると仮定し，ルート音と完全5度上の音を入力する．ルート音の弾弦タイミングとベロシティは6弦のものを，5度上の音の弾弦タイミングとベロシティは5弦のものを適用する．その他の弦の弾弦情報は出力しない．

ルート音（歌唱の音高）は常に6弦に割り当てられるため，仮に6弦を弾弦せずに5弦だけ弾弦しても，5弦に割り当てられる音高は歌唱されているルート音の5度上の音高となる．

A.3.2　その他の機能

その他以下のような機能を備えている．

- ピッチベンド付き MIDI データ保存機能

  本システムでは 1 音毎の音高はその区間の最頻値で決定しているが，本機能は最頻値の音高を基準としてその区間の短時間ピッチの変動をピッチベンド情報に変換し，保存する機能であり，単音モード，パワーコードモードとも対応する．なおピッチが基準音高より大きく離れた場合は無視する．歌唱によるスライドなどの表情入力が可能となる．

- 演奏機能

  弾弦後最初に E2-G5 内に存在した短時間ピッチを基準音高として MIDI NoteOn メッセージを生成し，MIDI 音源をリアルタイムで発音させることにより，本システムを楽器として用いることができる．これは単音モード，パワーコードモードとも対応している．

  弾弦と歌唱タイミングが合うとは限らず，意図した歌唱音高を取得できないことがありうるので，歌唱から得たピッチ情報をピッチベンド情報にし，発音後でもある程度音程調整できるようにした．

  現状では，録音機構に用いた DirectSound のレイテンシおよび FFT フレームサイズに起因する遅延の問題により 100ms 程度の発音遅延が発生する．

A.3.3 システムデザイン

第 2 章で実装したタップ併用型 Voice-to-MIDI システムを元に，2 弦分の入力/出力に対応するように拡張を行った．

入力は歌唱波形と弾弦によるリズム区切り情報およびベロシティ，出力は E2-G5（A4 = 440Hz とする）の半音単位の音高列，つまり MIDI データとなる．ギターの音域と若干異なるのは，人間の音域を考慮したためである．入力音声は 22050Hz, 16bit, モノラルでサンプリングされ，短時間フーリエ変換はフレームサイズ= 2048samples：約 100ms， フレーム移動間隔=128samples：約 6ms で行われる．

無発声検知機構で対象とする音高範囲は E2-G5 である．後述する A.3.6 節に列挙した挙動にあるように，EZ-AG では次の弾弦か弾弦中に弦スイッチや消音セン

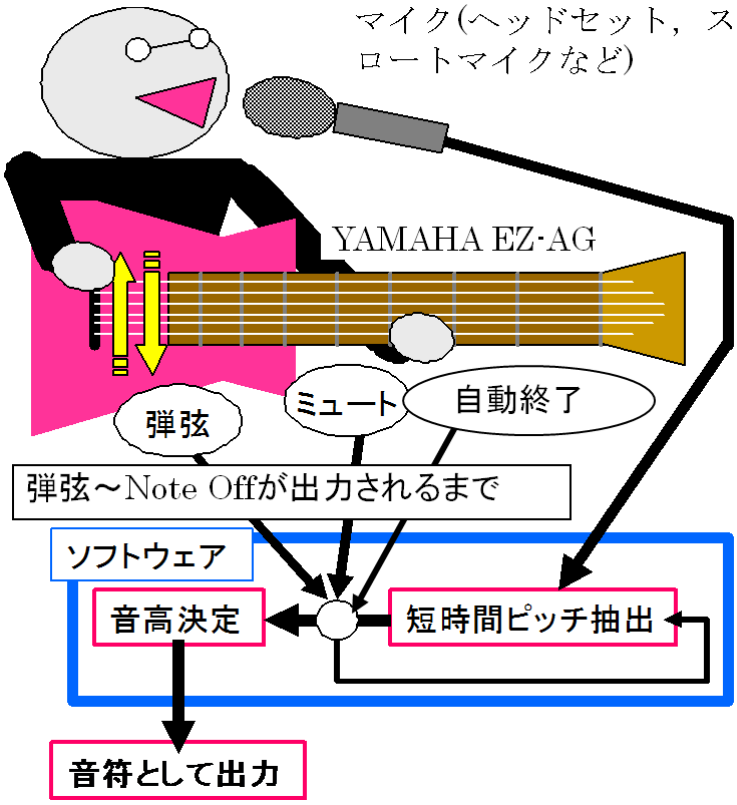

図 A.1: 処理の流れ：弾弦情報が入ると，短時間ピッチのヒストグラム化開始．自動終了 or 弦スイッチ or 消音センサ or 次の弾弦により，ヒストグラムの最頻値から音高を決定する．パワーコードを出力する場合は，5 度上の音高を付加して出力する．

サに触れないと Note Off メッセージが発行されない．よって弦スイッチや消音センサを駆使しなければ，短い音長で入力したいときでもテヌートになってしまったり，最後の音が消音されないことが起こりうる．そのため，無発声検知機構は吟たぁにとって重要な機能と言える．

システムの開発には Microsoft：Visual C # 2005 を用い, 短時間ピッチ算出部については Visual C++2005 の DLL で作成した．

A.3.4　動作概要

次に単音モード/パワーコードモードに共通する入力〜出力の処理の流れについて述べる（図 A.1）．録音が開始されると，入力されている音声波形に対して短時間フーリエ変換による短時間ピッチ算出処理が録音終了まで繰り返される．この間に弾弦情報が入力されたら，短時間ピッチを E2-G5 の範囲で半音単位で音高ヒストグラム化してゆく．これは次に同じ弦が弾弦されるかギター型ディバイスの指板上の弦スイッチや消音センサに触れる，あるいは無発声検知機構によって Note Off が発行されるまで続く．その後，弾弦された弦に応じた MIDI チャンネル，弾弦の強さをベロシティ，ヒストグラムの最頻音高を用いて音符情報を生成し，出力する．これを録音停止まで続ける．最後の音については，何らかの理由で録音停止時に終了していなければそこで終了させる．

A.3.5　歌唱入力用マイク

歌唱の入力には，ヘッドセットマイクか環境に応じてスロートマイクを用いる．通常は PC 用に市販されているヘッドセットマイクのようなものでもよいが，激しく弾弦すると弦を弾く音が混入したり，演奏機能使用時にスピーカの出力音を拾うことがあるためこのような場合にはスロートマイクを用いる方がよいと思われる．

A.3.6　弾弦情報入力装置

弾弦情報の入力には YAMAHA EG-AG [99] を用いた．本ディバイスは，

- 弾弦時に弦自体から音が出ない
- 音源のローカルオフ設定が可能
- 弾弦情報の取得が比較的容易

という理由により採用した．

以下に吟たぁ制作に当たって著者が調べた EZ-AG の出力可能な情報や挙動について記す．

- 弦ごとに独立した出力チャネルを持つ (1 弦=1ch,..., 6 弦=6ch)
- 弦ごとに独立したベロシティを出力
- ブリッジ付近にある消音センサで鳴っている全弦の Note Off を出力（Note Off は Velocity = 0 の Note On メッセージを使用）
- 次の弾弦か弦スイッチや消音センサに触れないと Note Off が発行されない
- 指板上の弦スイッチを押すと Velocity = 16 の Note On を発行される．そのときすでに弾弦中であれば先にその音を消音するための Note Off が発行される（今回は，歌唱中であれば指板上の弦スイッチを押したときの Note Off による消音は行うが，Note On は全く用いないので除去する）．
- Note On/Off イベントの出力と同時にそれらのイベントがエクスクルーシブイベントととしても出力される．
- カッティング奏法などによって短時間に大量の MIDI イベントを出力すると，イベントの出力順序が入れ替わることがある．

なお YAMAHA から過去市販されていた EZ-EG [100] とはトレモロアームの有無，弦の共振の再現の有無などの点で違いがある．EZ-EG では共振を再現した弾弦情報が出力されるが，本システムではそれらを除去する処理は行っていないため EZ-EG には対応していない．

A.4 評価実験

本システムについての評価のために実験を行った．

A.4.1 実験概要

A.3.1 に挙げた機能のうち，単音入力 (単弦ピッキング) およびパワーコード入力について，その作業負荷や使用感，音の区切りの入力をギターインタフェイスに変えた結果，出力された MIDI シーケンスデータにその楽器らしさが反映できたかなどを評価する実験を行った．

1. 課題メロディの単音入力（テンポ自由）

2. 課題メロディの単音入力（オルタネイトピッキングでテンポ自由）

3. 課題メロディの単音入力（BPM=120）

4. 被験者が用意したフレーズのパワーコード入力（テンポ自由）

単音入力時の課題メロディは，作曲は負荷が高く，また新たに曲を覚えるのは負荷があるため，被験者にとって既知と思われる童謡「赤とんぼ」とした．パワーコード入力に用いたフレーズは，被験者が用意したLUNA SEA「ROSIER」の中の冒頭などで聴くことのできる8小節程度のものである．

被験者は，主にビジュアル系ロックやUKロックを嗜好するギター上級者1名である．システムが本来想定している歌唱やリズムをとることはできるが，十分なギター演奏はできない人である．しかし，ギターは自由度の高い楽器であり，弦のマッピングなどシステム設計に無理がないかを調べ，また上級者ならではの工夫やコツなどの点からアドバイスを得るため，ギター上級者に評価を依頼した．

Dell : Lattitude XT ノートPC上でシステムを稼働し，MIDIインタフェイスを介して接続したEZ-AGから弾弦情報を，マイク（バッファローコクヨサプライ：USBヘッドセットマイクBMHUN01SVA）から歌唱を取得した．

実験前にEZ-AG自体の試奏時間を設けた．当初弾弦について弾きづらいとの意見を得たが，しばらく後，慣れてきたとのことであったのでEZ-AGのインタフェイス自体が実験に影響を及ぼすことはなかったと思われる．

評価はアンケートによって使用感などを訊いて行った．各条件に共通するアンケートの項目を以下に示す．

(a) 精神的負荷について

(7段階：非常に低い-低い-どちらかというと低い-普通-
どちらかというと高い-高い-非常に高い)

(b) リズム通りに弾弦できたと思うかについて

(6段階：非常に思う-思う-どちらかというと思う-
どちらかというと思わない-思わない-全く思わない)

(c) 歌唱について

(7 段階：非常に歌いやすい-歌いやすい-どちらかというと歌いやすい-普通-どちらかというと歌いづらい-歌いづらい-非常に歌いづらい)

(d) 意見や感想の自由記述

またパワーコード入力については，吟たぁの出力であるルート音に 6 弦の，5 度上の音に 5 弦の弾弦情報を用いて生成したフレーズ (最初のフレーズ) とルート音，5 度上の音とも 6 弦の弾弦情報を用いて著者が手動生成したフレーズ (後のフレーズ) をブラインド試聴によって比較させ，以下の評価を行った．試聴に用いた音色は Distortion Guitar である．

(e) 気に入った方を選択およびその理由を自由記述（2 択:最初のフレーズ，後のフレーズ)

そして実験の最後に以下の 2 項目について自由に記述させた．

(f) 使用してみて満足な点

(g) 使用してみて不満な点

A.4.2 実験結果および考察

結果および考察について実験条件ごとに述べてゆく．

1. 課題メロディの単音入力（テンポ自由）

   (a) 負荷は [非常に低い]

   (b) リズム通りに弾弦できたと [思う]

   (c) [非常に歌いやすい]

   (d) 「鍵盤を使ったときよりもスイッチ的感覚がなかった.」[2]

   被験者自身に合ったテンポであり，また普段から弾弦には慣れている部分もあるのか，歌唱と弾弦を同時に行っても余裕があると思われる．

2. 課題メロディの単音入力（オルタネイトピッキングでテンポ自由）

   (a) 負荷は [どちらかというと低い]

[2]被験者はタップ併用型 Voice-to-MIDI システムを使用して「赤とんぼ」を歌唱したことがあり，弾弦という動作をタップの代わりに音符を区切る手段として捉えている．

(b)　リズム通りに弾弦できたと [非常に思う]

(c)　[歌いやすい]

(d)　「どこをアップにするか考えてしまうときがある→精神的負荷となる.」

オルタネイトピッキング（ダウンピッキングだけでなくアップピッキングも用いる奏法）による弾弦では，設問 (a) の結果より，条件「課題メロディの単音入力」と比べて負荷が上がっている．この理由については設問 (d) で述べられている通りであるが，更に被験者は「ダウンピッキングは表拍，アップピッキングは裏拍のように決めている」と回答している．よってメロディのような音価がばらつきやすいフレーズの場合，ギタリストにとっては多少混乱を起こすことがあると考えられる．

3.　課題メロディの単音入力（BPM=120）

(a)　負荷は [非常に低い]

(b)　リズム通りに弾弦できたと [非常に思う]

(c)　[歌いやすい]

(d)　特になし

メトロノームに合わせて入力したが，テンポが速くなっても条件「課題メロディの単音入力」と比べて大きく悪くなることはなく，設問 (b) ではよりよい評価を得た．ただし更にテンポが速くなるか，短い音価のフレーズでは負荷が増すことが考えられるため，アンケート結果は変わると思われる．

4.　被験者が用意したフレーズのパワーコード入力（テンポ自由）

(a)　負荷は [非常に低い]

(b)　リズム通りに弾弦できたと [非常に思う]

(c)　[非常に歌いやすい]

(d)　「入力できたのかちょっと不安である」

(e)　気に入ったのは [最初のフレーズ]　　理由: 「雰囲気が入力しようとしていたものに近かったから.」

テンポは原曲より遅めで入力され，フレーズに細かなノートが少なかったこともあり，条件「課題メロディの単音入力」と同じように良好な評価を得た．しかし入力できたのか不安を感じている．その理由を尋ねたところ,「ロずさんだフレーズの調が正しかったか，不安だったから」とのことであった．これは直接的に「吟たぁ」の仕様などが影響したものではないと考えられる．しかしこのような不安を解消するために伴奏，CD などに合わせて弾くことはありうる．この点については，スロートマイクを使用するか伴奏をヘッドホンで聴くことで解決しうると思われる．

被験者は，気に入ったのは吟たぁの出力したフレーズであると回答した．タイミングとベロシティが独立していることがフレーズに表情を与え，また聴感上でも知覚できたと思われる．また，音の区切りの入力をギターインタフェイスに変えた結果，出力された MIDI シーケンスデータにその楽器らしさが反映できたと思われる．

最後に全体的なシステムへの意見や感想について示す（原文のまま記載）．

- (f) 使用してみて満足な点

  「速い旋律の入力にアップピッキングが使えるのは良い.」

  実験条件「課題メロディの単音入力（BPM=120)」において，アップピッキングも用いることに若干の負荷を感じているが，同時に利点でもあると考えていると察せられる．ギターにそれほどなじみがない人は，拍の表裏を気にしないと思われるためアップピッキングの使用についての評価は異なる可能性がある．

- (g) 使用してみて不満な点

  「左手を動かしたくなるときがある．パワーコード入力はベースラインを口ずさんでもらうのを限定した方がよいかも.」

  実は，被験者はルート音ではなく 5 度上にあたる音を歌唱していた．著者は必ずルート音の歌唱であると仮定していたが，選択できるような仕組みが必要であろう．

全体として，おおむね良好な評価を得たと言える．また，ギター上級者ならでは意見やアドバイスも得ることもできた．

A.5　おわりに

本章では，歌唱から音高を，弾弦からタイミングやベロシティを取得することによって，パワーコードによるギターフレーズや単音フレーズの入力が可能なシステム「吟たぁ」を提案した．

ギター上級者 1 名に使用してもらい評価を行い，おおむね肯定的な意見を得た．パワーコードフレーズの MIDI シーケンスデータの比較結果より，音区切り入力デバイスをギター型インタフェイスに変えた結果，出力された MIDI シーケンスデータにその楽器らしさが反映できたと考えている．このことから，音区切りデバイスの選択も表現の一要素になりうることが分かった．また将来的に，あらゆる音環境の時系列から音を切り出し，音楽フレーズへと変換・出力するためのシステムの拡張を行うための基礎的な評価ができた．

今後，歌唱やリズムをとることはできるが十分なギター演奏はできない人を対象としたギターフレーズ入力評価を行う予定である．そして，“ EnvJamm ”の音区切り入力デバイスへの応用を考えている．

今後の予定として

・パワーコード入力モードにおいて，ミュート弦の弾弦も取得

・より弾きやすくするため，弾弦する弦を 6, 5 弦固定からそれぞれ最初に弾弦された弦および 2 番目に弾弦された弦に変更

・パワーコードと単音入力のシームレスな切り替えが可能となるモードの追加

・単音入力のために，どの弦を弾弦しても入力可能なモードの追加

・弾弦する弦によって上下 1 オクターブトランスポーズして入力可能なモードの追加（女性による低音入力，男性による高音入力，ベースパート入力など）

・指板上の弦スイッチの活用：スライド奏法等

・歌唱からのエクスプレッション情報の抽出

といった機能について検討し，実装を行う．