■2群（画像・音・言語）－ 9編（音楽情報処理）

2章 技術・アプリケーション

（執筆者：片寄晴弘）［2011年7月 受領］

■概要■

本章では，音楽情報処理の技術領域を，音楽分析系技術，楽音分析・合成技術，音楽生成系技術に大別して紹介する．

人間は音楽を聞いて，メロディやリズムやハーモニーをとらえ音楽的な心象を得ている．この部分を計算機処理によって実現できれば，自動採譜や音楽推薦やセッションシステムなどの幅広い応用が見込まれる．音楽分析系技術として，音響特徴抽出，基本周波数推定，リズム認識などについて紹介する．

電子音源の技術がなければ，現在の音楽プロダクションは成り立たないといっても過言ではない．この部分に関連する技術として，楽音分析・楽音合成と，近年その重要性が高まりつつある歌声分析・歌声合成について紹介する．

自動作曲に代表される生成系技術は，アート・デザイン系に関連した華やかな領域であるが，人工知能の実現対象としても古くから着目されており，1957年には，計算機を用いた音楽作品「イリアック組曲」が作られている．その後，計算機の能力の指数的な向上に支えられる形で，生成系の技術は，リアルタイム，インタラクティブに動作するシステムの開発へと進展してきた．自動作／編曲，演奏の表情付け，自動伴奏，インタラクティブパフォーマンスについて紹介する．

■2群－9編－2章

2-1 音響特徴抽出

(執筆者：北原鉄朗) [2011年6月 受領]

音楽音響信号に対する特徴抽出は，1970～80年頃の単旋律に対する基本周波数（F0）推定がその発端である．その後，混合音を扱う研究が現れ，自動採譜という分野を形成した（詳細は2-3節参照）．一方，1990年代から，楽器音から様々な特徴量を抽出して楽器の同定を行う研究が行われるようになった．はじめは単一音を対象としたものがほとんどであったが，徐々に混合音を扱うものも増えつつある．しかし，複数の楽器が同時になった音響信号に対して個々の楽器を同定するのは難しく，現状では数種類の楽器を扱うのがやっとといったところである．2000年代に入り音楽情報検索の需要が高まってくると，混合音全体から直接，低次の特徴量を抽出するアプローチがとられるようになり，その後の主流となる．しかし低次の特徴量の限界も指摘されるようになり，音響信号から高次の特徴量を抽出する試みも行われている．

本節では，上述の流れに従い，単一音に対する楽器音特徴抽出・楽器同定，混合音に対する楽器音特徴抽出・楽器同定，混合音からの低次特徴抽出，高次の特徴抽出の順に，研究動向を解説する．

2-1-1 単一音に対する楽器音特徴抽出・楽器同定

楽器音特徴抽出の基本となる概念は調波構造である．調波構造とは，打楽器を除く多くの楽器音がもつ性質で，聴覚上は一つに聞こえる音には，基本となる周波数成分（基音あるいは基本周波数成分）のほかに，その整数倍の周波数の成分（倍音あるいは高調波成分）が含まれているというものである．古くは倍音の振幅比によって音色が決まると信じられてきたが，現在では，その時間的変化や発音直後の非調波成分も関係していることが，音響心理学などの実験を通じて明らかになっている．

単一音からの特徴抽出では，こういった各倍音の振幅比のほか，様々な時間的変化に関する特徴量も用いられる．例えば，Martin[1]は，倍音振幅比関連としてSpectral Centroid*や奇数次倍音と偶数次倍音の振幅比†など，時間変化関連として，周波数変調や振幅変調，立ち上がり時間‡など計31個の特徴量を用いて，14楽器から1 023音に対して楽器同定実験を行った．Eronenら[2]，Essidら[3]，北原ら[4]なども同様の特徴量を用いている．また，メル周波数ケプストラム係数（MFCC）など，音声分析で用いられる特徴量も使われている[5]．

多くの研究では，高次元特徴空間の冗長性を排除するため，何らかの次元圧縮を行っている．次元圧縮は，主成分分析のように全特徴の線形結合で新たな特徴軸を作る方法と何らかの基準で特徴量を選択する方法とに分けられる．

* パワーを重みとした周波数の重み付き平均．音のかん高さを表す．
† クラリネットなどの閉管楽器は偶数倍音が小さいという特徴がある．
‡ 音が鳴りはじめて振幅が最大になるまでの時間．

2-1-2 混合音に対する楽器音特徴抽出・楽器同定

混合音に対する特徴抽出・楽器同定が難しい一つの要因は，異なる音源の倍音成分が同じ周波数で重なることで，特徴が変化してしまうことである．もしも音源分離技術で完全に音源ごとの音響信号を取り出せるのであれば，混合音に対する特徴抽出・楽器同定は，単一音に対するそれに帰着する．しかし，実際には倍音成分の重なりにより歪みは不可避であるため，それによる特徴量の変化に対処する必要がある．

この問題に対する解決策として，各特徴量が周波数成分の重なりの影響を受けているかどうかを判定し，受けていれば無効化したり低い重みを与えるといった試みがなされた．木下ら[6]は，特徴量を，音の重なりによる影響の受け方に従って加算特徴量，優先特徴量，崩壊特徴量に分類し，この分類に基づいて特徴量を再計算あるいは無効化する処理を提案した．

Egginkら[7]は，ミッシングフィーチャ理論による特徴量のマスキング（無効化）を行った．この方法では，同定対象音と同時になるほかの音の F0 を推定し，その最小公倍数の周波数に由来する特徴量をマスキングしている．北原ら[8]は，混合音から学習データを得た場合に，音の重なりによる影響を受けるほどクラス内分散・クラス間分散比が高くなることに着目し，これが最小化するように特徴量の重みを決定するアプローチをとった．

なお，柏野らは，特徴抽出をせずにテンプレートマッチングを行うことで，特徴変化の問題を回避した[9]．この手法では，楽器の個体差などによる音響特性の違いに対処するため，テンプレートを適応する処理も行われている．

2-1-3 混合音からの低次特徴抽出

前節のような，個々の音源に対して特徴抽出・楽器同定を行う方法は非常に難しく，現状でも 3〜5 種類の楽器を扱えるにすぎない．そこで，混合音全体の音響的特徴を一つの音色と考え，混合音全体から直接特徴抽出する試みがされている．このような音響特徴は polyphonic timbre と呼ばれることもある．以下，よく用いられる特徴量を紹介する．なお，$M_t[n]$ は時刻 t における n 番目の周波数ビンのパワー，$N_t[n]$ は正規化パワー，N は周波数ビン数を表す．

- Spectral Centroid（SC）— 2-1-1 節でも述べたとおり，振幅を重みとした周波数の重み付き平均，すなわち，

$$C_t = \sum_{n=1}^{N} nM_t[n] \Big/ \sum_{n=1}^{N} M_t[n].$$

- Spectral Rolloff（SR）— ある周波数 R_t 以下の周波数帯域におけるパワーの累積値が全体のパワー累積値の例えば 85%になるとき，すなわち，

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^{N} M_t[n]$$

を満たすときの周波数 R_t.

- Spectral Flux（SF）— 隣り合うフレーム間における各周波数ビンの正規化パワーの二乗差の合計値，すなわち，

$$F_t = \sum_{n=1}^{N} \left(N_t[n] - N_{t-1}[n] \right)^2.$$

・Zero Crossing Rate（ZCR）— 時間領域信号 $x(t)$の符号が変わる回数（直線 $x = 0$ をまたぐ回数）.

このほかに，MFCC や，パワーなどの時間変化に関する特徴量などが用いられる．これらの特徴量を用いて，ジャンル識別 [10]，ムード検出 [11]，映像との調和度計算 [12]，楽曲間の類似度計算 [13]などが行われている．

マルチメディアコンテンツに対するメタデータ記述の枠組みである MPEG-7 でも，様々な音響特徴量が定義されている．ただし，SC は採用されているものの，SR，SF，ZCR，MFCC などは含まれていない．以下，MPEG-7 で定義されている低次音響特徴量を列挙する．詳細は文献 14) などを参照されたい．

- Basic descriptors: Audio Waveform, Audio Power.
- Basic spectral descriptors: Audio Spextrum Envelope, Audio Spectrum Centroid, Audio Spectrum Spread, Audio Spectrum Flatness.
- Basic signal parameters: Audio Harmonicity, Audio Fundamental Frequency.
- Temporal timbral descriptors: Log Attack Time, Temporal Centroid.
- Spectral timbral descriptors: Harmonic Spectral Centroid, Harmonic Spectral Deviation, Harmonic Spectral Spread, Harmonic Spectral Variation, Spectral Centroid.
- Spectral basic representations: Audio Spectrum Basis, Audio Spectrum Projection.

2-1-4 高次の特徴抽出

SCなどの低次特徴量は，音響的な特性は表すが，音楽的な特徴を表すわけではない．我々が音楽を聴くとき，通常は音響的な特性に興味があるのではなく，その楽曲の内容に興味がある．そのため，楽曲の特徴量はその音楽的内容を適切に表すべきである．このような観点から，音楽音響信号から高次の特徴量を抽出する研究が行われている．具体的には，メロディ・ベースラインのF0 軌跡，ビート，繰り返し構造やサビ [15]，コード進行 [16~18]，楽器構成 [19]，ドラムパターン [20]，歌手の声色 [21] などの抽出に取り組まれている．また，こういった特徴量を用いた音楽情報検索の研究も取り組まれている [19, 21, 22].

■参考文献

1) Martin, K.D., “Sound-Source Recognition: A Theory and Computational Model,” PhD Thesis, MIT, 1999.
2) Eronen, A. and Klapuri, A., “Musical Instrument Recognition using Cepstral Coefficients and Temporal Features,” Proc. ICASSP, pp.735−756, 2000.
3) Essid, S., Richard, G. and David, B., “Musical Instrument Recognition by Pairwise Classification Strategies,” IEEE Trans. Audio, Speech, Lang. Process., vol.14, no.4, pp.1401-1412, 2006.
4) 北原鉄朗，後藤真孝，奥乃博，“音高による音色変化に着目した楽器音の音源同定：F0 依存多次元正規分布に基づく識別手法，” 情処学論，vol.44, no.10, pp.2448-2458, 2003.
5) Brown, J.C., “Computer Identification of Musical Instruments using Pattern Recognition with Cepstral Coefficients as Features,” J. Acoust. Soc. Am., vol.103, no.3, pp.1933-1941, 1999.
6) 木下智義，坂井修一，田中英彦，“周波数成分の重なり適応処理を用いた複数楽器の音源同定処理，” 信学論, no.J83-D-II, no.4, pp.1073-1081, 2000.
7) Eggink, J. and Brown, G.J., “A Missing Feature Approach to Instrument Identification in Polyphonic Music,” Proc.

ICASSP, vol.V, pp.553-556, 2003.
8) 北原鉄朗，後藤真孝，駒谷和範，尾形哲也，奥乃博，“多重奏を対象とした音源同定：混合音テンプレートを用いた音の重なりに頑健な特徴量への重み付け及び音楽的文脈の利用，”信学論, vol.J89-D, no.12, pp.2721-2733, 2006.
9) 柏野邦夫，村瀬　洋，“適応型混合テンプレートを用いた音源同定，”信学論，vol.J81-D-II，no.7, pp.1510-1517, 1998.
10) Tzanetakis, G. and Cook, P., “Musical Genre Classification of Audio Signals,” IEEE Trans. Speech Audio Process., vol.10, no.5, pp.293-302, 2002.
11) Lu, L., Liu, D. and Zhang, H.-J., “Automatic Mood Detection and Tracking of Music Audio Signals,” IEEE Trans. Audio, Speech, Lang. Process., vol.14, no.1, 2006.
12) 西山正紘，北原鉄朗，駒谷和範，尾形哲也，奥乃　博，“マルチメディアコンテンツにおける音楽と映像の調和度計算モデル，”情処研報, 2007-MUS-69, pp. 31-36, 2007.
13) Aucouturier, J.-J. and Pachet, F., “Improving Timbre Similarity: How High’s the Sky?,” Journal of Negative Results in Speech and Audio Sciences, 2004.
14) Kimi, H.-G., Moreau, N. and Sikora, T., “MPEG-7 Audio Beyond,” Wiley, 2005.
15) 後藤真孝，“リアルタイム音楽情景記述システム：全体構想と音高推定手法の拡張，”情処研報，2000-MUS-37, pp.9-16, 2000.
16) Fujishima, T., “Realtime chord recognition of musical sound: a system using common lisp music,” Proc. ICMC, pp. 464−467, 1999.
17) Sheh, A. and Ellis, D.P.W., “Chord segmentation and recognition using EM-trained hidden Marcov models,” Proc. ISMIR, 2003.
18) Yoshioka, T., Kitahara, T., Komatani, K., Ogata, T. and Okuno, H.G., “Automatic Chord Transcription with Concurrent Recognition of Chord Symbols and Boundaries,” Proc. ISMIR, pp.100-105, 2004.
19) Kitahara, T., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G., “Instrogram: Probabilistic Representation of Instrument Existence for Polyphonic Music,” IPSJ Journal, vol.48, no.1, pp.214-226, 2007. (also published in IPSJ Digital Courier, vol.3, pp.1−13).
20) Yoshii, K., Goto, M. and Okuno, H.G., “Drum Sound Recognition for Polyphonic Audio Signals by Adaptation and Matching of Spectrogram Templates with Harmonic Structure Suppression,” IEEE Trans. Audio, Speech, Lang. Process., vol.15, no.1, pp.333-345, 2007.
21) Fujihara, H. and Goto, M., “A Music Information Retrieval System based on Singing Voice Timbre,” Proc. ISMIR, pp.467-470, 2007.
22) 土橋祐亮，北原鉄朗，片寄晴弘，“音響信号を対象としたベースラインからの音楽ジャンル解析，”情処研報，2008-MUS-74, 2008-SLP-70, pp. 217-224, 2008.

■2 群－9 編－2 章

2-2 基本周波数推定（歌声研究に関する視点から）

（執筆者：森勢将雅）［2011 年 6 月 受領］

ピッチは音の高さに対応する心理量であり，ピッチに相当する物理量の基本周波数（f_0）の推定は，楽器音や音声の音色に相当するスペクトル包絡推定と並び，古くから研究されてきた．基本周波数の抽出技術には，自動採譜，歌声の分析[1)] や合成[2)]，演奏の表情の分析，あるいは，音響に反応するインタラクティブシステムの制作など，幅広い応用領域がある．

基本周波数の推定は，単音楽器や音声単独を対象とした研究と，楽曲中の歌唱など多重音を対象とした研究により方針が大幅に異なる．本節では，基本周波数推定に関して積極的な研究がなされてきた音声領域の研究事例に焦点を当てて技術紹介を行う．

2-2-1 音声に対する基本周波数の定義と推定に対する問題点

音声は，本来声帯振動を伴う有声音と伴わない無声音とに区別されるが，本節では有声音に限定して議論する．基本周波数は，声帯振動が生ずる時間間隔の逆数と定義され，基本周波数の高低はピッチの高低と対応する．一般的に，人間の発話を長期的に観測すると特性が大きく変化するため，厳密な意味での基本周波数は定義が困難である．通常の基本周波数推定は，音声波形を短時間のフレームとして切り出し，その区間に存在する周期を推定する．しかしながら，短時間の音声を観測した場合も声帯振動の時間間隔や声帯振動の波形は微細に変化しているため，正確な基本周波数の推定は容易ではない．

基本周波数推定法では，音声の時間波形に対する周期性に着目した分析法と，パワースペクトルの調波構造に着目した方法とに大別される [3)]．基本周波数分析について提案されてきた従来法の位置づけを整理するため，**図 2･1**，**2･2** に音声波形とその音声波形のパワースペクトルを示す．音声波形は，基本周期 T_0 で声帯振動が繰り返され，パワースペクトルは，f_0 Hz の基本波を示すピークに加え，その整数倍にもピークをもつ調波構造となる．したがって，基本周波数を推定する場合，音声波形に着目すると図 2･1 における T_0 を求める問題として扱われ，パワースペクトルに着目すると図 2･2 における基本波の周波数 f_0 を求める問題として扱われる．

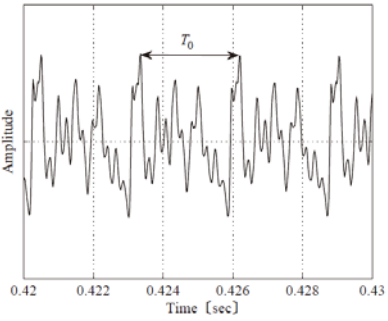

図 2・1　時間波形における基本周期．T_0 の逆数が f_0 となる．

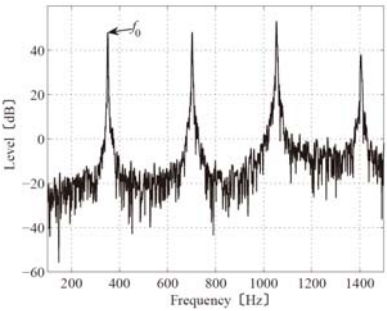

図 2・2　図 2·1 のパワースペクトル．最も低い周波数のピークが f_0 となる．

(1)　時間波形に着目した方法

時間波形における性質を利用した方法では，信号の相関を用いる方法 [4]が一般的である．音声の基本周期が T_0 であれば，音声波形の自己相関，あるいは相互相関を求めると，T_0 の整数倍で高い相関値を示し，それ以外の時刻では低い相関値を示す．そのため，時刻 0 のピークを除いた最も早い時間のピーク時刻が T_0 となる．

相関に基づく方法を用いる場合，誤ったピークを選択することによる推定誤りが問題となる．一般的には，T_0 の整数倍以外に生ずる不要なピークを除外するための閾値を設け，閾値を上回るピークの中から最も早い時間のピークを T_0 とする．しかしながら，声帯振動は，生ずる時間間隔，及び声帯振動波形が例え短時間であっても微細に変化しているため，目的とする時刻のピークが検出されない，あるいは T_0 以外，特に T_0 の整数倍の時刻のピークを誤検出する問題が起こり得る．

近年では，音声の相互相関関数を基準とし，不要なピークの低減や不要な演算を削減することにより，高速かつ高精度に基本周波数を推定可能な推定法 YIN [5] が提案されている．YIN による基本周波数推定では，YIN の提案された 2002 年以前の従来法と比較して誤差を 1/3 以下に低減できることが文献 5)により示されている．

(2)　パワースペクトルに着目した方法

周期信号のパワースペクトルは，f_0 の整数倍にピークをもつため，パワースペクトルの最も低いピークの周波数を抽出することで，基本周波数が推定できる．あるいは，f_0 の整数倍にピークを有する調波構造に着目し，調波構造のピーク間隔を推定することでも基本周波数を推定できる．パワースペクトルを基準とした推定法では，調音フィルタに起因する影響がパワースペクトルに混在するため，その影響を除去するための方法が要求される．

ケプストラム [6, 7] は，対数パワースペクトルを逆フーリエ変換することで得られ，ケフレンシーと呼ばれる時間を単位とするパラメータである．調音フィルタによる影響を分離可能であることから，ケプストラムは，基本周波数推定だけではなくスペクトル包絡推定にも利用される音声分析の代表的なパラメータといえる．音声のケプストラムを求めた場合，調音フィルタに起因するケプストラムは低次のケフレンシーに集中し，基本周波数に起因するピ

ークが，調音フィルタに起因する成分よりも高次のケフレンシーである時刻 T_0 に生ずる．そのため，高次ケフレンシーに存在するピークを抽出し，その逆数を求めることで基本周波数を推定可能である．

基本周波数が高い場合，基本周波数に起因するピークがケプストラムの低次に生ずることとなる．相関を用いた方法と同様にケフレンシー軸における T_0 以外のピークを誤検出することに加え，調音フィルタのケプストラムが原因で正しいピークを抽出できないことが問題点となる．

2008 年に提案された SWIPE' [8] は，パワースペクトルの調波構造に着目し，誤差を低減する様々な工夫を施すことで高い推定精度を達成する方法である．誤差を低減させる処理に複雑な演算が必要であることから計算コストは大きいが，YIN と比較した場合推定誤差を更に低減することができる．

(3) その他の特徴量を用いた方法

これらの方法以外にも，より高精度の基本周波数推定を達成するため，様々な方法が提案されている．例えば，基本波をフィルタリングで取り出す方法 [9]，瞬時周波数を用いた方法 [10] や，ウェーブレット変換を用いた方法 [11] などが提案されている．このほかにも，声帯振動時刻を直接検出する方法 [12] を用いることで，その間隔を直接求めることも可能である．2005 年に提案された NDF[13] は，既存の複数の特徴量を抽出し，基本周波数の時間軌跡が滑らかになるよう前後の基本周波数を修正する後処理により推定精度を大きく改善している．

これらの方法は，高い精度を追求するため，計算コストに関する検討が行われていないのが現状である．近年では，音声や歌唱を加工することが可能なソフトウェアも実用化されているが，それらの音声分析においては，計算時間を可能な限り低減する必要がある．そこで本節では，歌唱の分析や合成への応用を目的とした高速かつ高精度な基本周波数推定法 [14] を紹介する．

2-2-2 歌唱の分析・合成を目的とした基本周波数推定

歌唱分析・合成を目的としたコンテンツ制作を行う場合，そのコンテンツに用いられる音声は，防音室やレコーディングスタジオなど，背景雑音の少ない環境で録音される場合が多い．特に，歌唱合成などをコンテンツ制作現場で利用することを考慮すると，雑音がほぼ存在しない長時間の音声を対象として，正確な f_0 を可能な限り高速に推定することが望まれる．

文献 14) では，この目的を満たすため，分析対象とする音声を，低域雑音を含まない音声に限定し，高速かつ高精度な基本周波数推定法が提案されている．この方法は，図 2･2 における最も低い周波数の基本波を低域通過フィルタにより抽出し，基本波の周波数を時間波形から計算する簡素な方法である．基本波検出に基づいて基本周波数を推定する場合，低域通過フィルタのカットオフ周波数は，推定時には未知である f_0 以上，その整数倍のピークの最小値となる $2f_0$ 以下に設定することが要求される．基本波をフィルタリングで取り出す従来法 [9] では，事前に別の方法で基本周波数候補を推定し，その近辺を抽出するフィルタリングが行われていた．文献 14) の方法では，f_0 の仮定をせず，以下に示される三つのステップにより基本周波数の推定を行う．

ステップ 1：低域通過フィルタによるフィルタリング

基本波の抽出は，最適な一つのカットオフ周波数を有する単一のフィルタではなく，低域から高域まで様々なカットオフ周波数を有する複数のフィルタ群により行われる．ステップ 1 では，様々なカットオフ周波数を有する低域通過フィルタ群により信号全体を処理する．フィルタ数に応じて計算コストは増大するが，このフィルタリングは波形全体に対する処理であり，フレーム単位で推定を行う従来法よりも高速な処理が可能である．音声処理には高い時間分解能が必要なため，低域通過フィルタには，カットオフ周波数以上の周波数のエネルギーを十分に抑圧できることだけではなく，フィルタ長が短く有限の時間で振幅が 0 に収束することが要求される．

ステップ 2：基本波らしさの計算

フィルタリングにより基本波のみが抽出された場合，その時間波形は周期が T_0 の正弦波となる．不要なピークを含む，あるいは基本波を含まない場合は，正弦波とは異なる波形となる．したがって，フィルタリングにより基本波が得られているか否かは，時間波形がどの程度正弦波に近いのかを評価すればよい．

この方法では，波形が正弦波の場合，信号のピークの間隔，谷の間隔，正から負のゼロ交差の間隔，負から正のゼロ交差の間隔がすべて等しくなることに着目する．抽出された波形が正弦波に近いほど四つの間隔も等しくなるため，その標準偏差は 0 に近づくこととなる．基本波らしさは，四つの間隔の平均を f_{ave}，標準偏差を f_{std} とした場合，$\exp(-f_{std}/f_{ave})$ により与えられる．基本波らしさは，0 から 1 の値を示し，1 に近いほど高精度に基本波を検出されたといえる．また，f_{ave} がその信号，その時刻における基本周波数の候補となる．

ステップ 3：基本波らしさに基づく最終的な基本周波数の選定

ステップ 2 により，フィルタリングされた各波形の基本周波数候補と基本波らしさが計算される．ステップ 3 では，すべての候補から，各時刻における最終的な基本周波数を選定する．ただし，低域通過フィルタの条件より，以下の条件のいずれかを満たす候補は除外される．

- フィルタリングに用いられたカットオフ周波数の下限を下回る候補，上限を上回る候補
- 低域通過フィルタの通過域以外の周波数に存在する候補
- 低域通過フィルタのカットオフ周波数の半分以下の周波数に存在する候補

この選定処理で残った候補から，最も基本波らしさの大きい候補を，最終的な候補とする．

本方法は，低域に雑音が存在する音声に対する推定は困難であるが，低域の雑音が存在しない音声の場合，SWIPE'や NDF と実質的に同等の性能を達成しつつ，計算時間を SWIPE' の 1/42，NDF の 1/80 にまで低減可能である．

2-2-3 今後の展望

基本周波数は，音声の主要な要素であり，その高精度な抽出は，学術的だけではなく，歌

唱合成ツールなど産業的な価値も高い研究テーマである．特に，商用化を目指す場合，計算コストに関する問題があるため，限られた計算時間でより高精度な推定性能を達成するための工夫 [15] が求められる．また，CGMが一般化した現在，一般ユーザ向けの技術提供も重要な課題である§．

■参考文献

1) 中野倫靖, 後藤真孝, 平賀　譲, “楽譜情報を用いない歌唱力自動評価手法, ”情報処理学会論文誌, vol.48, no.1, pp.227-236, 2007.
2) 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, ” 情報処理学会研究報告，2008-MUS-74, pp.25-32, 2008.
3) W. Hess, “Pitch determination of speech signals,” Springer-Verlag, Berlin, 1983.
4) M.J. Ross, H.L. Shaffer, A. Cohen, R. Freudberg, H.J. Manley, “Average magnitude difference function pitch extractor,” IEEE Transactions on acoustic, speech, and signal processing, vol.ASSP-22, no.5, 1974.
5) A. Cheveigné and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” J. Acoust. Soc. Am., vol.111, no.4, pp.1917-1930, 2002.
6) A.M. Noll, “Short-time spectrum and “cepstrum” techniques for vocal pitch detection,” J. Acoust. Soc. Am., vol.36, no.2, pp.269-302, 1964.
7) A.M. Noll, “Cepstrum pitch determination,” J. Acoust. Soc. Am., vol.41, no.2, pp.293-309, 1967.
8) A. Camacho and J.G. Harris, “A sawtooth waveform inspired pitch estimator for speech and music,” J. Acoust. Soc. Am., vol.124, no.3, pp.1638-1652, 2008.
9) 大村浩, 田中和世, “基本波フィルタリング法による精細ピッチパターンの抽出, ”日本音響学会誌, vol.51, no.7, pp.509-518, 1995.
10) 阿竹義徳，入野俊夫，河原英紀，陸金林，中村哲，鹿野清宏，“調波成分の瞬時周波数を用いた基本周波数推定方法, ” 電子情報通信学会論文誌 D, vol.J83-DII, no.11, pp.2077-2086, 2000.
11) 佐宗晃，中村尚五，“ウェーブレット変換を用いたピッチ抽出の一方法, ” 電子情報通信学会論文誌 A, vol.J80-A, no.11, pp.1848-1856, 1997.
12) K.S.R. Murty and B. Yegnanarayana, “Epoch extraction from speech signals,” IEEE Transactions on audio, speech and language processing, vol.16, no.8, 2008.
13) H. Kawahara, A. Cheveigné, H. Banno, T. Takahashi and T. Irino, “Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT,” Proc. Interspeech2005, pp.537-540, 2005.
14) 森勢将雅，河原英紀，西浦敬信，“基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法, ” 電子情報通信学会論文誌 D, vol.J93-D, no.2, pp.109-117, 2010.
15) 森勢将雅, 中野皓太, 西浦敬信, “歌唱合成システムの実現を目的とした高品質音声分析合成法の提案, ” 電子情報通信学会応用音響研究会, vol.110, no.71, pp.89-94, 2010.
16) “歌声合成ソフトウェア UTAU, ” http://utau2008.web.fc2.com/index.html

§ 筆者により提案された音声分析変換合成システム WORLD[15] は，UTAU[16] 用に実装され，様々な楽曲制作者により利用されている．

■2 群－9 編－2 章

2-3 音の群化・自動採譜

(執筆者：亀岡弘和)［2011 年 6 月 受領］

我々人間は，多数の音が混じり合った音響信号から，個々の音を難なく聴き分けることができる．足し算が不可逆であるのと同じように，いったん重畳されてしまった波形から個々の波形を復元することは一般には困難である．にもかかわらず，混じり合っている個々の音を正確に聞き取れるのは，人間の聴覚の「アルゴリズム」がいかに優秀であるかを示している．人間は，物体を見たときに，どこまでをひとまとまりなのかをとらえ，物体と物体の「境界」を把握することができるのと同じように，音を聞いたときにも，どこまでをひとまとまりなのかをとらえ，音どうしの「境界」を把握することができる．このように，ひとまとまりの音を把握することを「音の群化」といい，このように形成されたひとまとまりの音の「塊」を「音脈」という．

音の群化といういわば逆問題を，人間がどのようなアプローチにより解いているのかについては未解明な点が多い．両耳に入ってくる二つの波形の微妙な違いに基づいて知覚される，波源位置の情報は音を聴き分けるための手がかりの一つに違いないが，我々はモノラル録音された音響信号からですら個々の音を聴き分けられる能力をもっている．このことは，人間には空間的手がかり以外の手がかりに基づく何らかの音の群化メカニズムが備わっていることを示唆し，この困難な逆問題を人間がどうにかして解いているという事実は，音を聴き分ける原理を追究することへの動機となっている．

更に，我々は音楽を聴くとき，その音響信号には様々な現象的な「揺らぎ」があるにもかかわらず，どのような楽器が，どの音高で，どのようなビート，リズムで奏でられているかを容易く理解できる．これは，低次の機能による信号処理モジュールだけでなく，知識に基づく高次の機能によるパターン処理モジュールを総動員した，人間の優れた音の認識メカニズムのなせる業である．計算機に，音楽の音響信号から自動的に楽譜化させることを自動採譜[1, 2]という．これは，人間の低次機能と高次機能を統合した圧倒的な情報処理メカニズムに迫ろうとする大いなる挑戦である．

本項目では，特に音の群化の問題に焦点を当て，問題を整理しながらその計算論的アプローチに関する近年の取組みについて紹介する．

2-3-1 基本周波数推定（周波数方向の群化）

音の群化の問題と多重音の基本周波数推定の問題との間には，極めて密接な関係がある．このことを明快にするため，分かりやすい例題として単一音のパワースペクトルから基本周波数を推定する問題について考えよう．もし信号が純音の場合，パワースペクトルのピーク周波数が基本周波数に対応する（**図 2･3**(a)）が，一般の周期信号には複数のピークがある（図 2･3(b)）．そして複数あるピークの内最大のピークの周波数が必ずしも基本周波数に対応するとは限らない（図 2･3(c)）．また，基本周波数成分はいつも大きいとは限らないため，複数あるピーク周波数のうち最も低い周波数を基本周波数と見なすのは頑健なやり方ではない（図 2･3(d)）．以上より，基本周波数を推定するためには，スペクトルピークのような限られた情

報だけで済ませようとするのではなく，対象とする音の信号波形やスペクトル構造の全体を手がかりにしたロバストな方法が必要になる（単一音の基本周波数推定すら容易でないことは長い研究の歴史が物語っている [3]）．しかしながら，複数の信号が混合されて観測される音響信号には，どの成分がどの音に帰属するのかという情報が欠落しているため，基本周波数を推定するための重要な手がかりが得られないのである．したがって，音の群化の問題が解かれない限り，個々の基本周波数を推定することは容易ではないわけである．一方で，もし，個々の音の基本周波数が既知であれば（極めて特異な状況であるが），各音に由来する成分の検討がつくため，音の群化の問題は大幅に解きやすくなる．すなわち，音の群化の問題を解く手がかりになる基本周波数の情報が，音の群化が解かれない限り安定的に求められない，といういわゆる「鶏と卵」の状況に陥るのである．

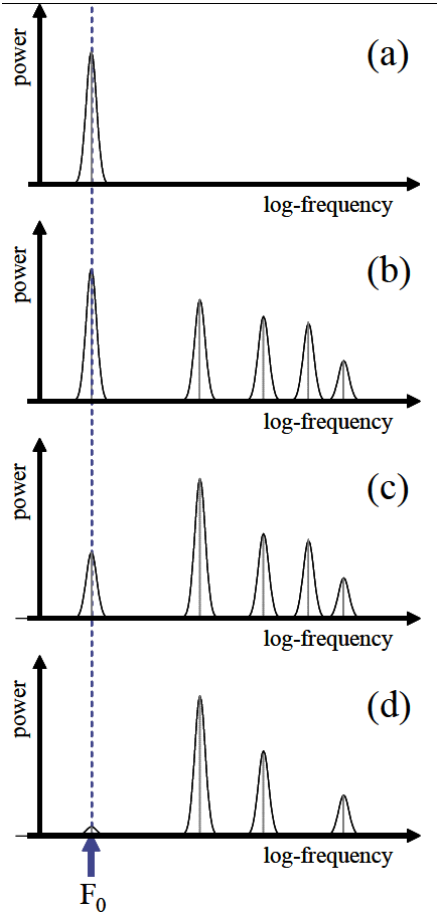

図 2・3　基本周波数推定の問題

この問題に対しては，音に関する先験的知識（調波性やスペクトル概形の仮定）を利用するのが主流な常套手段となる．例えば，観測音響信号をあらゆる基本周波数の音の重みつき混合としてモデル化してその重みを推定するアプローチ [4]，スペクトルクラスタモデルを用いて音の群化と基本周波数推定を反復的に行うスタイルのアプローチ [5, 6] や対数周波数領域で調波構造のシフト不変性を仮定して対数周波数スペクトルを調波構造パターンで逆畳み込みするアプローチ [7] などが試みられている．このほかにも，多重音から基本周波数を推定する手法は膨大にあるので，より詳しい動向については，［2-2 ピッチ抽出］節やほかの著書 [8] を参照されたい．

2-3-2 計算論的聴覚情景分析（時間方向の群化）

前節では，暗黙のうちに非常に短い時間区間における波形から個々の音に群化する問題について考えていた．我々人間でも，数十ミリ秒程度の混合信号から個々の音を聴き分けるのは必ずしも容易ではなく，容易に聴き分けるためにはある程度の信号の長さが必要になる．前節で考えていた問題は，周波数方向の群化と呼ぶものに相当し，人間はそれだけでなく音の時間的な連なりを形成する時間方向の群化も同時に行っているとされる．

近年，聴覚情景分析 [9] と呼ぶ心理学的アプローチの枠組みによって徐々に明らかになってきた人間の音の群化メカニズムに関する知見を積極的に利用して，音の群化問題の解決を図ろうとする試みが進められており，その枠組を総称して計算論的聴覚情景分析（Computational Auditory Scene Analysis: CASA）と呼ぶ．具体的には，知識を利用しない聴覚の低次の音の分離能力に関して，音響信号はスペクトログラムに似た要素に「分解」されること，同じ音源に由来する要素は「群化」されて音脈を形成すること，群化のされやすさ（分凝要件）は，（1）調波性，（2）調波成分の立上りの共通性，（3）調波成分の周波数及び振幅変化の共通性，（4）成分の連続性，（5）時間周波数の近接性，（6）音源位置の共通性などに関係する，ことなどが心理実験を通して示されている．瞬時瞬時において調波的関係にある周波数成分を一つの音としてグルーピングすることを周波数方向の群化といい，それらを分凝要件（2）～（5）に基づいて継時的にグルーピングすることを時間方向の群化という．これによって，例えば，二つの音声の基本周波数軌跡がある時点で交差していたとしても，本来は分離不能なはずの交差の瞬間における個々の音声信号の各周波数成分がどのように重なっているかを前後の時刻から推論できるようになるわけである．CASA の目的は，このような人間の低次機能による音の群化メカニズムを模倣することであり，上記の分解と群化のプロセスを，分凝要件に関係する物理量を用いてアルゴリズムとして実現し，音脈の認識に有用な特徴量（基本周波数など）を抽出したり，目的音に相当する音脈の再構成を行うことである．

その具体的なアプローチとしては，周波数方向の群化に相当する処理により各離散時刻において個々の構成音の瞬時特徴成分（例えばスペクトルや基本周波数）を抽出したのちに，マルチエージェントシステム [4, 12] やベイジアンネットワーク [10] や隠れマルコフモデル [11] や Kalman フィルタ [13～16] などの手段を通して，時間的にどの成分が同じ一連の音に対応しているかを瞬時特徴成分の時間的滑らかさなどを評価尺度にして推定する方法が主流である．また一方で，分凝要件（1）～（5）から逸脱しない範囲の自由度をもった時変スペクトルを直接的にモデル化し，これを混合したもので観測時間周波数スペクトルにフィッティングする，周波数方向及び時間方向の群化を同時最適化問題として定式化されたアプローチ [6, 17～20, 23, 24]

も考案されている.

2-3-3 スパース成分分析（記憶に基づく群化）

ところで我々は，ユニゾン（同一音高またはオクターブ違い）で弾かれたピアノとヴァイオリンの音を聴き分けることができる場合がある．一定の時間連続して一方の音の調波成分が完全に他方の音の調波成分と重なってしまうこの状況では，前後の時刻から調波成分の重なり具合を推論することが難しいため，これまで述べてきた群化メカニズムとは別の何らかのメカニズムが存在している可能性が示唆される．極めてわずかな基本周波数の違いによって二つの信号の間に干渉が生じており，それを手がかりにしている可能性もあるが，それよりもピアノやヴァイオリンがどのような音色であるかを漠然と記憶していて，それに基づいて個々の音脈を推論するような働きが関与しているとも考えられる.

ピアノの音とヴァイオリンの音を過去にもっと容易に聴き分けやすい状況で聴き分けた経験があったとして，その経験から，それぞれの音響的特徴に関する「辞書」が作られているとすると，この「辞書」はユニゾンのような群化が困難な状況においても高い精度で音を群化するための有用な手がかりになる．そして，こうして音の群化が高い精度でなされるたびに，信頼性の高い学習データを得たことになり，「辞書」の再学習が可能になる.

スパース成分分析の考え方を基礎として，以上のような観点で音の群化の問題をとらえたアプローチが近年脚光を浴びている．具体的には，各時刻で観測される混合信号ないし混合スペクトルを，時刻によらず共通な基底セット（辞書）の重み付き和によってモデル化し，できるだけ重みをスパース[**]に，かつ復元誤差を小さくするように基底と重みを学習すると，一つひとつの基底が最大限の情報量をもった効率的な分解表現へと誘導されるかたちとなり，結果，各基底が観測中に頻発する信号あるいはスペクトルのパターンとなるはずだとする考え方である [21]．通常，基底と重み[††]は交互に更新されるため，ちょうど上述の例えと同様な反復学習が行われることになる．非負値行列分解（Non-negative Matrix Factorization; NMF）は，基底と重みをいずれも非負制約のもとで学習する方法で，効率的な学習アルゴリズムが存在する点，非負制約以外の制約がなくとも副次的に重みがスパースになる基底の解が得られる点が特徴的である [22]．また，NMFを応用し，混合信号中の個々の音の間で共通しているスペクトル包絡や微細構造を自己組織化的に発見する方法も試みられている [25].

2-3-4 自動採譜への展望

以上で見てきたような音の群化アルゴリズムの実現は，音楽を人間と同等以上に計算機に理解させる自動採譜の実現への第一歩である．自動採譜の実現に向けて，以上で述べた低次機能に相当するアルゴリズムと同等かそれ以上に，高次機能のモデル化についての検討も必要である．いずれは，低次機能による個々の音への群化処理と，それを音楽的制約に基づいて構造化して記号化する認識処理とを統合的に行える [26]，で論じられているような大規模な情報統合モデルの構築が必要になるであろう.

[**] ほんの一部の基底の係数だけが大きな値をもち，それ以外の係数は 0 であることを「重みがスパースである」という.

[††] 重みの更新は，前段で更新された辞書信号あるいは辞書スペクトルを観測にフィッティングさせる操作に相当し，すなわち音の群化処理に他ならない.

■参考文献

1) J.A. Moorer, “On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer,” Ph.D. Thesis, Stanford University, 1975.
2) H. Katayose, S. Inokuchi, “The Kansei music system,” Computer Music Journal, vol.13, no.4, pp.72−77, Winter 1989.
3) W. Hess., “Pitch determination of speech signals.,” Springer-Verlag, Berlin, 1983.
4) M. Goto, “A Real-Time Music-Scene-Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-World Audio Signals,” Speech Communication (ISCA Journal), vol.43, no.4, pp.311−329, 2004.
5) H. Kameoka, T. Nishimoto, S. Sagayama, “Separation of Harmonic Structures Based on Tied Gaussian Mixture Model and Information Criterion for Concurrent Sounds,” In Proc. 2004 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2004), vol.4, pp.297-300, 2004.
6) H. Kameoka, “Statistical Approach to Multipitch Analysis,” Ph.D. Thesis, The University of Tokyo, 2007.
7) S. Saito, H. Kameoka, K. Takahashi, T. Nishimoto, S. Sagayama, “Specmurt Analysis of Polyphonic Music Signals,” IEEE Transactions on Audio, Speech and Language Processing, vol.16, no.3, pp.639−650, 2008.
8) A. de Cheveigné, “Multiple F0 Estimation,” in Computational Auditory Scene Analysis: Principles, Algorithms and Applications, D.-L. Wang, G.J. Brown Eds., IEEE Press / Wisely, 2006.
9) A.S. Bregman, “Auditory Scene Analysis,” MIT Press, Cambridge, 1990.
10) K. Kashino, K. Nakadai, T. Kinoshita, and H Tanaka, “Application of the Bayesian Probability Network to Music Scene Analysis,” In D.F. Rosenthal and H.G. Okuno, editors, Computational Auditory Scene Analysis, pp.115−137, Lawrence Erlbaum As- sociates, 1998.
11) M. Wu, D.L. Wang and G.J. Brown, “A Multipitch Tracking Algorithm for Noisy Speech,” IEEE Transactions on Speech and Audio Processing, vol.11, pp.229−241, 2003.
12) T. Nakatani, M. Goto and H.G. Okuno, “Localization by Harmonic Structure and Its Application to Harmonic Sound Segregation,” In Proc. 1996 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP’96), pp.653−656, 1996.
13) 西, 安部, 安藤, “聴覚情景分析のための多重ピッチ追跡と調波分離アルゴリズム,” 計測自動制御学会, vol.34, no.6, pp.483−490, 1998.
14) 鵜木, 赤木, “聴覚の情景分析に基づいた雑音下の調波複合音の一抽出法,” 電子情報通信学会論文誌, vol.J82-A, no.10, pp.1497−1507, 1999.
15) 安部, 安藤, “共有 FM-AM の時間周波数統合に基づく聴覚情景分析(I)−Lagrange 微分特徴量とその周波数統合−,” 電子情報通信学会論文誌, vol.J83-D-II, no.2, pp.458−467, 2000.
16) 安部, 安藤, “共有 FM-AM の時間周波数統合に基づく聴覚情景分析(II)−最適な時間軸統合とストリーム音の再合成−,” 電子情報通信学会論文誌, vol.J83-D-II, no.2, pp.468−477, 2000.
17) H. Kameoka, T. Nishimoto, S. Sagayama, “A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering,” IEEE Transactions on Audio, Speech and Language Processing, vol.15, no.3, pp.982−994, 2007.
18) 亀岡弘和, ルルージョナトン, 小野順貴, 嵯峨山茂樹, “調波時間構造化クラスタリングによる CASA へのアプローチ,” 日本音響学会聴覚研究会, vol.36, no.7, H-2006-103, pp.575-580, 2006.
19) H. Kameoka, T. Nishimoto, S. Sagayama, “Audio Stream Segregation of Multi-Pitch Music Signal Based on Time-Space Clustering Using Gaussian Kernel 2-Dimensional Model,” In Proc. 2005 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2005), vol.3, pp.5-8, 2005.
20) J.Le Roux, H. Kameoka, N. Ono, A. de Cheveigne, S. Sagayama, “Single and Multiple Pitch Contour Estimation through Parametric Spectrogram Modeling of Speech in Noisy Environments,” IEEE Transactions on Audio, Speech and Language Processing, vol.15, no.4, pp.1135-1145, 2007.
21) S.A. Abdallah and M.D. Plumbley, “Unsupervised Analysis of Polyphonic Music Using Sparse Cod-ing,” IEEE Transactions on Neural Networks, vol.17, no.1, pp.179-196, 2006.
22) P. Smaragdis, J.C. Brown, “Non-Negative Matrix Factorization for Music Transcription,” In Proc. 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA2003), pp. 177-180, 2003.
23) K. Miyamoto, H. Kameoka, T. Nishimoto, N. Ono, S. Sagayama, “Harmonic-Temporal-Timbral Clustering (HTTC) for the Analysis of Multi-instrument Polyphonic Music Signals,” In Proc. 2008 IEEE International

Conference on Acoustics, Speech and Signal Processing (ICASSP2008), pp.113-116, 2008.
24) 糸山克寿, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博, “楽譜情報を援用した多重奏音楽音響信号の音源分離と調波・非調波統合モデルの制約付パラメータ推定の同時実現,” 情報処理学会論文誌, vol.49, no.3, pp. 1465-1479, March 2008.
25) 亀岡弘和, 柏野邦夫, “複合ソースフィルタモデルによる音響信号の三要素テンソル分解,” 電子情報通信学会 2008 年総合大会講演論文集, vol.AS-5-5, pp.S-56--S-57, 2008.
26) 柏野邦夫, “音楽音響信号を対象とする聴覚的情景分析に関する研究,” 東京大学大学院工学系研究科博士論文, 1994.

■2群－ 9編－2章

2-4 オーディオアライメント・ビートトラッキング・リズム認識

(執筆者：武田晴登) [2011年6月 受領]

2-4-1 はじめに

タイトルに挙げられている項目は，音楽演奏からリズムやテンポやビートなどの音楽の時間構造に関する情報を得るための技術をまとめたものである．これらの技術の中には既に実用化されているものもあり，例えばテンポの自動解析機能は，PC上の音楽再生ソフトや携帯音楽プレイヤに実装されており，楽曲推薦やプレイリスト自動作成に用いられている．音楽のビートやテンポについては，人間がビートを認知する機構の解明として心理学の分野での研究があるが，計算機パワーが豊富になり信号処理やパターン認識技術が普及するにつれ工学の研究テーマとして取り上げられるようになった．以下，後者の観点から解説する．

2-4-2 オーディオアライメント

オーディオアライメント（audio alignment）は，演奏曲の楽譜が与えられているときに，演奏の音響信号と楽譜と時間的対応づけを自動で求める技術である．楽譜に紐付いている情報を音響信号に紐付けることができ，例えばクラシック音楽のように楽譜をもとにしたアナリーゼの情報が手に入る場合に，「78小節目からここが第2主題」というような楽譜に付与されている情報を，音楽を聴きながら参照することに利用できる．また，名演奏家のテンポなどを算出して演奏解析や解析結果を利用した自動演奏などにも用いることができる．実際の人間の歌や器楽演奏では，音色や音の大きさやテンポやリズムなど様々な不確定で変動する要素があるため，演奏中のすべての音を楽譜と対応づけさせることは大変難しく，現在でも質の高いメタデータを楽曲中に付与するには人手に頼らなければならない．

音響信号と楽譜の拍との対応づけを行うアライメント処理の一般的な手順は，与えられた音楽音響信号からは音響的特徴の時間変化を特徴量の時系列として抽出し，一方で楽譜からはこれらの特徴量と対応づけられる拍やピッチの情報を抽出し，この両者の時系列の間で適切なアライメントを求める手順で行われる．一つのアプローチとして，楽譜をシンセサイザを用いて音響信号に変換し，楽譜と音響信号のアライメントを二つの音響信号のアライメントに問題として扱う手法が提案されている．すなわち，楽譜から生成した音響信号は楽譜との時間の対応づけは分かるので，一つの音響信号の対応関係から演奏と楽譜との対応づけが求められる．音響信号からはパワースペクトルを平均律の12音階にパワーを分解したベクトル（クロマベクトル，chroma vectorと呼ばれる）や，パワースペクトルの時間変化を使用し，DTW（dynamic time warping）によるマッチングが用いられている[1,2]．また，楽譜を音響信号に変換せず，確率モデルを用いてモデル化を行い，楽譜のモデルと入力された音楽音響信号の特徴量時系列をマッチングするアプローチも提案されている．このアプローチでは，マッチング対象をモデルにすることで統計学習により事前知識をモデルの与えることでより頑健なマッチングが可能になることが期待されている[3]．

2-4-3 ビートトラッキング

オーディオアライメントは演奏曲の楽譜が与えられていることを仮定しているが，ビート

トラッキング（beat tracking）とリズム認識（rhythm recognition）は，演奏曲の楽譜が未知である場合にビートあるいはリズムパターンを求める技術である．ビートトラッキングは，一般にはビート（拍を打つ時刻）とビート間隔（テンポ）を求める問題から出発し，それから小節などの複数の拍にまたがる構造や 1 拍より短い長さの構造を推定する処理系の全体を指し，通常は複数の処理系をボトムアップに積み上げて構成される [5, 12～14]．典型的な構成においては，まず音響信号から楽器音の発音時刻を検出し，検出した発音時刻の中からビートに対応するものをビートの追跡（トラッキング）として求め，後段で拍節構造についての付加的な情報の推定が行われる．

1. 発音時刻の検出

最初に音響信号から演奏された音の発音時刻（onset time）を求める．楽器音のうち，音の開始時の音量変化が急激で，その後すぐに音量が減衰する音は，音量の変化を手がかりに発音時刻を求められる．例えば，**図 2･4** に見られるように音量 $y(t)$ の時間変化のピークが明確に見られる場合は，これらのピークを抽出すればよい [6]．音量の変化は，発音時刻を検出するために使用できる特徴の一例であり，これ以外にも多くの特徴が検討されている．例えば，音量の変化を周波数帯域を限って計算したエネルギーを発音検出の指標に用いたり [7, 13]，スペクトルのパワーではなく位相の変化に注目してピッチの変化の安定しない箇所を発音時刻として抽出する手法 [9] が報告されている．また，発音の開始が緩やかに起きる場合でも，ピッチに変化がある場合はピッチに注目して発音時刻を求めることができる [8]．特徴量だけでなく，発音の識別器の学習方法についても半教師つき学習の利用が検討されている [10, 11]．

2. テンポの推定

次に，発音時刻の検出結果を用いて，あるいはそれとは別に，テンポすなわち，一拍の長さ（ビート間隔の時間長）を求める処理が行われる．これは，入力信号の周期を求める問題であるととらえられるので，通常の周期を求める手法と同様に自己相関に着目する手法が使われてきた．図 2･4 に示すように，音量の変化に周期性がある場合は音量の時間変化 $x(t)$ の自己相関のピークを与える時間差をビートの間隔として求めることができる．また，発音時刻の結果を使用できる場合は，IOI（発音時刻の間隔，inter-onset interval）の統計からビートを求めることもできる．単純に IOI の出現頻度の最も高いところを周期とするのではなく，例えば IOI をクラスタリングするなど [12] 様々な解析方法が検討されている．いずれの手法もビートの間隔がほぼ一定であること，つまりテンポが一定ある区間を対象とした解析であり，その範囲で有効に動作することが報告されている．また，ピッチ抽出のように周期を求める問題に見られるように，倍や半分の周期を求めてしまう誤りはここでも原理的に起こり得る．

3. ビートの追跡と拍節構造の解析

以上の発音時刻とビートの時間長の情報をもとに，発音時刻を追跡し，ビートや拍節構造を求める．事前に定めた規則に基づいて複数のエージェントがビートの仮説を追跡するシステムが作られている [12, 13]．また，小節の開始を特徴としてコードの変化に着目し，周波

数解析を行い狭帯域ごとの時間変化から時刻を求めたり [14]，ドラムの音のように周波数軸全体に広がる打楽器音の音色に着目してスペクトル形状のマッチングからドラムのビートパターンを推定する手法 [13]も提案されている．

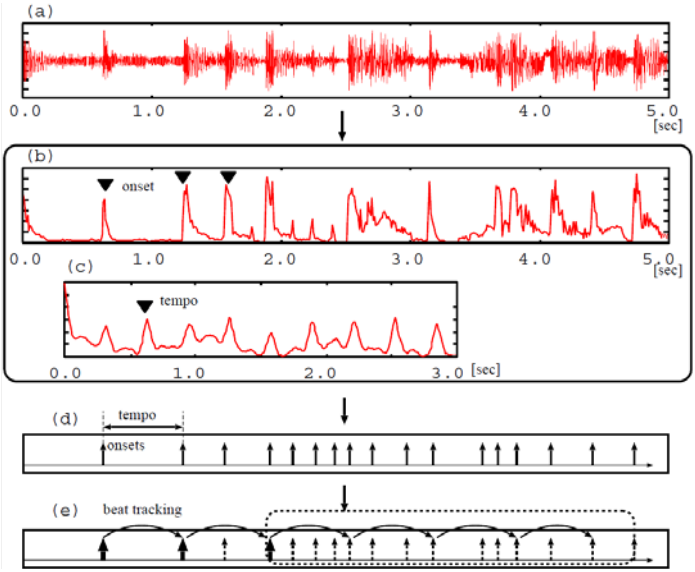

図 2・4　ボトムアップに構成される単純なビートトラッキングの例

(a) 入力された音響信号から(b) 音量の変化に着目してピーク抽出により発音時刻を求め，また(c) その自己相関からビートを求め，(d) これらの情報を用いて(e) ビートトラッキングと拍節情報を求める

2-4-4 リズム認識

ビートトラッキングでは，一般にテンポが一定であることを仮定して，ビートの推定を基準としたボトムアップに処理系を構成したが，一方で，テンポの変化を含むリズムパターンをモデルで表現して推定を行うモデルベースのアプローチの研究もある．リズム認識を目的としたモデル化でテンポの変化を扱うために，テンポを内部変数としたモデル化されている．テンポはマルコフ過程 [15] や，あるいは演奏解析の研究で用いられる [16]ように時間について連続的に変化する曲線としてモデル化 [17] が学習や推定のアルゴリズムと共に提案されている．

2-4-5 おわりに

ここで紹介したテンポ，ビート，リズムの推定は，現在でも更なる高精度化は技術課題である．一方で，例えばリズムに注目した楽曲分類 [18] のように音楽検索，ビートを認識してテンポを合わせてダンスするロボットの開発 [19] など，テンポやビートの情報を応用した音楽情報処理の研究の進展している．これらの研究により，ユーザが体験できる音楽の幅は今後ますます広がるであろう．

■参考文献

1) R.B. Dannenberg and Ning Hu, “Polyphonic audio matching for score following and intelligent audio editor,” in Proc. Int. Comp. Mus. Conf., pp. 27-34, 2003.
2) S. Dixon, G. Widmer, “MATCH: A music alignment tool CHEST,” in Proc. Int. Conf. Music Inf. Retrieval, 2005.
3) C. Raphael, “A hybrid graphical model for aligning polyphonic audio with musical scores,” In Proc. Int.Conf. Music Inf. Retrieval, 2004.
4) S. Hainsworth, “Beat tracking and musical metre analysis,” in Signal Processing Methods For Music Transcription, A. Klapuri and M. Davy, Eds. New York: Spriger, 2006.
5) F. Gouyon and S. Dixon, “A Review of automatic rhythm description systems,” Comp. Mus. J., vol.29, no.1, pp.34-54, 2005.
6) J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, “A tutorial on onset detection in music signals,” IEEE Trans. Speech Audio Process., vol.13, pp.1035-1047, 2005.
7) A. Klapri, “Sound onset detection by applying psychoacoustic knowledge,” in Proc. Int. Conf. Aouct. Speech Signal Process., pp. 3089-3092, 1999.
8) N. Collins, “Using a pitch detector for onset detection,” in Proc. Int. Conf. Music Info. Retrieval, pp.100−106, 2005.
9) J.P. Bello and M. Sandler, “Phase-based note onset detection for music signals,” in Proc. IEEE Int.Conf. Acoust. Speech. Signal Process., pp.49-52, 2003.
10) Ning Hu and Roger B. Dannenberg, “Bootstrap learning for accurate onset detection,” Mach Learn, vol.65, pp.457-471, 2006.
11) W. You and R.D. Danennberg, “Polyphonic music onset detection using semi-supervised learning,” In Proc. Int. Conf. Music Inf. Retrieval, pp.279−282, 2007.
12) S. Dixon, “Automatic extraction of tempo and beat from expressive performances,” J. New Music Res., vol.30, no.1, pp.39-58, 2001.
13) M. Goto, “An audio-based real-time beat tracking systme for music with or without drum-sounds,” J.New Music Res., vol.30, no.2, pp.159-171, 2001.
14) A. Klapuri, A. Elonen, and J. Astora, “Analysis of the Meter of acousitic music signal,” IEEE Trans. Audio Speech Lang. Process., vol.14, no.1, pp.342-355, 2006.
15) A. Cemgil, B. Kappen, P. Desain, H. Honing, “On tempo tracking: tempogram representation and Kalman filtering,” J. New Music Res., vol.28, no.4, pp.259-273, 2001.
16) B.H. Repp, “Diversity and commonality in music performance: An analysis of timing microstructure in Schumann’s “Träumerei”, ” J. Acoust., Soc. Amer., vol.92, no.5, 1992.
17) H. Takeda, T. Nishimoto, and S. Sagayama, “Rhythm and tempo analysis toward automatic music transcription,” in Proc. Int. Conf. Acoust., Speech, Signal Process., 2007.
18) S. Dixon, F. Gouyon, and G. Widmer, “Towards characterisation of music via rhythmic patterns,” in Proc. Int. Conf. Music Inf. Retrieval, pp.509-516, 2004.
19) 村田和真, 中臺一博, 武田　龍, 奥乃　博, 長谷川雄二, 辻野広司, “ロボットを対象としたビートトラッキングロボットの提案とその音楽ロボットへの応用,” 日本ロボット学会誌, vol.27, no.7/8, pp.793-801, 2009.

■2群－9編－2章

2-5 音楽推薦・プレイリスト生成

(執筆者：吉井和佳) [2011年6月 受領]

現在，iTunes Music Storeなどのインターネットを介した音楽配信サービスは，数百万曲規模のデータベースを運用し，大きな成功を収めている．ユーザ個人レベルでは，ほとんどの楽曲が聴くに値せず，度を超えた楽曲の豊富さは意味がないように思える．しかし，全ユーザを合わせると，大量に存在する知名度の低い楽曲（テール）の再生回数は少数のヒット曲（ヘッド）のそれに匹敵するというロングテール効果[1]が観測できる．つまり，商品数を豊富にそろえることがe-commerce成功要件の一つであるといえる．ユーザはこのようなサービスを用いて好みに合う楽曲を入手し，iPodなどの携帯型音楽プレイヤに数千曲を登録している．今やユーザの音楽的嗜好の多様性には際限がないことが分かり，少数のヒット曲だけで満足できる時代ではなくなった．ユーザは本当に自分の音楽的嗜好（明確に言葉で表現しにくい場合が多い）に合う未知の楽曲との出会いを求めている．したがって，もう一つのe-commerce成功要件は，このようなユーザの希望を実現することにある．

2-5-1 音楽推薦技術の重要性

大規模な楽曲データベースからユーザ自らが好みに合う楽曲を見つけ出すのは容易ではないので，音楽推薦技術の重要性は高まっている．携帯型音楽プレイヤに登録した手持ちの大量の楽曲からユーザの好みに合ったプレイリストを自動生成する技術も音楽推薦技術の一種とみなすことができる．通常の音楽検索システムでは，アーティスト名や楽曲名をクエリとして利用するため，未知のアーティストや楽曲を探すことは困難である．この問題を解決するため，音楽情報処理分野では音楽内容に基づく検索が盛んに研究されている．例えば，ある楽曲がクエリとして指定されると，それと音楽内容の似た楽曲を提示する類似楽曲検索という形態が挙げられる．しかし，ユーザが好みに合う楽曲を検索しようとする場合には，ユーザ自身が入力したクエリが最適である保証はない．また，散歩中や運転中に携帯型音楽プレイヤやカーナビを用いて音楽を聴く場合など，検索を行うことが煩わしかったり禁止されたりしている場面（「ながら」聴き）も存在する．このような問題に対処するには，ユーザの音楽的嗜好を自動的に推定し，ユーザに代わって適切な楽曲を検索して提示するシステム，すなわち推薦システムが必要である．

2-5-2 従来の推薦技術

ユーザ群を U，楽曲群を M とすると，あるユーザ $u \in U$ に対する音楽推薦とは，u 自身更にはほかのユーザの行動履歴（評価・鑑賞・購入）を参考にして，u に知られていない（未評価・未鑑賞・未購入）が好まれそうな楽曲 $m \in M$ をデータベース中から選び出すことである．このとき，楽曲の音楽内容（人手による記述あるいは計算機による自動解析）を参考にすることもある．ユーザの行動履歴には，明示的なもの（五段階評価など）と暗黙的なもの（再生回数など）との2種類がある．本節では特に前者を意識して手法の解説を行うが，手法の適用可能性を限定するものではないことに注意されたい．

従来の推薦手法は，推薦に利用するデータの違いから「協調フィルタリング」と「内容に基づくフィルタリング」に大別することができ，相反する性質をもっている．

(1)　協調フィルタリング

協調フィルタリングでは，あるユーザに楽曲を推薦する際にほかのユーザの楽曲評価を参考にする [2]．基本的には，楽曲 A, B を好むユーザに推薦を行う場合，楽曲 A, B, C を好むユーザがほかに多数いれば，楽曲 C を推薦する．しかし，知名度が低かったりリリースされたばかりで評価が与えられていない楽曲は推薦できないという欠点がある．更に，推薦される楽曲におけるアーティストのバリエーションが乏しくなりがちである．

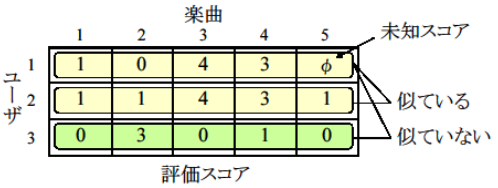

図 2・5　ユーザ間類似度に基づく協調フィルタリング

典型的な方法は，対象ユーザが評価を与えていない楽曲に対して，そのユーザが与えるであろうスコアを予測する（**図 2･5**）．いま，$\tilde{r}_{u,m}$ をユーザ u の楽曲 m に対する評価スコアの予測値であるとすると，$\tilde{r}_{u,m}$ はほかのユーザの評価スコアを参考にすることで求まる．

$$\tilde{r}_{u,m} = \bar{r}_u + k \sum_{\{u' | u' \neq u, u' \in U\}} w_{u,u'} \left(r_{u',m} - \bar{r}_{u'} \right) \qquad (5・1)$$

ここで，$\bar{r}_u$ 及び $\bar{r}_{u'}$ は，それぞれユーザ u, u' が付けた評価スコアの平均値である．$w_{u,u'}$ とは，ユーザ u, u' 間の嗜好の類似度を表し，k は $\sum_{u'} |w_{u,u'}| = 1$ とするための正規化定数である．式（5･1）に従ってすべての未評価の楽曲に対する評価スコアの予測値を求め，$\tilde{r}_{u,m}$ の大きいものから並べたランキングをユーザに提示する．このとき，嗜好の類似度を計算する尺度として，以下のピアソンの相関係数法がよく利用される．

$$w_{u,u'} = \frac{\sum_m \left(r_{u,m} - \bar{r}_u \right) \sum_m \left(r_{u',m} - \bar{r}_{u'} \right)}{\sqrt{\sum_m \left(r_{u,m} - \bar{r}_u \right)^2 \sum_m \left(r_{u',m} - \bar{r}_{u'} \right)^2}} \qquad (5・2)$$

ここで，$\sum_m$ は，ユーザ u, u' がともに評価を行った楽曲についてのみ和をとることに注意する．しかし，多くの場合にそのような楽曲数がほとんどゼロとなり，推薦精度が大幅に低下することが多いため，様々な改良が提案されている．

別の方法として，評価スコアを予測することなしに，対象ユーザが評価を与えていない楽曲を順位づけをすることもできる．この立場をとる手法として，Aspect Model に基づく推薦手法が挙げられる（**図 2･6**）．Aspect Model とは，内部に潜在変数を内包する確率モデル（ベ

イジアンネットワーク）であり，観測データ（ユーザによって与えられた評価スコア）の発生過程を表現している．ここで，潜在変数は音楽のジャンルに対応していると解釈すると理解しやすいが，この解釈は便宜的・概念的なものであることに注意されたい．まず，あるユーザ u がいて，そのユーザが確率 $p(z|u)$でジャンル z を選択する．次に，そのジャンル z が選ばれた場合に，確率 $p(m|z)$で楽曲 m が観測される．これらの確率を全ユーザの評価スコアを利用して求めることができれば，ユーザ u に対し，$p(m|u) \propto \sum_z p(m|z)p(z|u)$ の大きい順に楽曲 m に順位付けすればよい．しかし，観測できるのはユーザ u の楽曲 m に対する評価スコアのみであり，各枝の確率値は直接求めることはできないが，このような不完全データからのパラメータ推定問題は，EM アルゴリズムを適用して解くことができる．

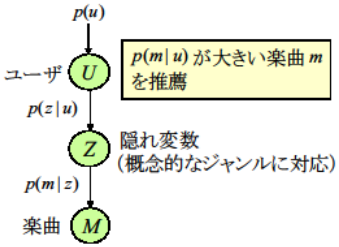

図 2・6　アスペクトモデルに基づく協調フィルタリング

(2)　内容に基づくフィルタリング

内容に基づくフィルタリングでは，音色やリズムなどの音楽内容の点で，対象ユーザに好まれそうだと予想される楽曲を推薦する．これは類似楽曲検索と似ているが，ユーザが任意の複数の楽曲に対して付与した「好き」か「嫌い」かの評価スコアを総合的に考慮して嗜好を推定する処理が必要である．

Hoashi ら [3] や Logan ら [4] は音楽内容として MFCC（Mel Frequency Cepstral Coefficients）と呼ばれる音声認識でよく利用される低次の特徴量を用いている．これらの方法では評価のなされていない楽曲も推薦可能で，アーティストのバラエティも豊かであるが，推薦精度に改善の余地が残されている．音楽内容には音響信号から得られるものだけでなく，テキスト（レビューコメントや人手によるタグ）から得られるものもある．

本節では，Logan らの提案した，音楽内容に基づく検索法を推薦目的で使えるよう自然に拡張した手法について説明する（正確には若干の改良を加えてある）．いま，楽曲 m の音楽内容がベクトル c_m で与えられると仮定し，ユーザ u が好きあるいは嫌いと評価した楽曲の集合をそれぞれ M^+_u，M^-_u と定義する．このとき，ユーザ u に楽曲推薦を行うアルゴリズムは以下のとおりである．

1. ユーザの音楽的嗜好の解析

内容ベクトルの集合$\{cm|m \in M^+_u\}$，$\{cm|m \in M^-_u\}$ はユーザ u の音楽的嗜好を表しており，それぞれの各要素を P ベクトル，N ベクトルと呼ぶことにする．

2. 楽曲の内容ベクトルとの類似度計算

ユーザ u によって評価を与えられていない楽曲 m' の内容ベクトル $c_{m'}$ に対して，各 P ベクトル及び各 N ベクトルとの類似度を計算する（**図 2･7**）．このとき，最大となる類似度を $s^{+}_{u,m'}$，$s^{-}_{u,m'}$ とすると，それらの差 $d_{u,m'} = s^{+}_{u,m'} - s^{-}_{u,m'}$ はユーザ u が楽曲 m' をどれくらい好きになりそうかを示す．この値は未評価楽曲すべてについて計算する．

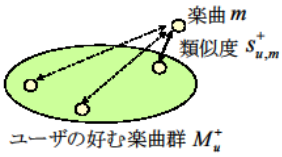

図 2･7　内容に基づくフィルタリング

3. 楽曲の順位づけ

未評価楽曲 m' に対して，$d_{u,m'}$ が大きいものから順位づけを行う．

類似度尺度としては，コサイン距離尺度が利用されることが多い．Hoashi らの研究でもコサイン距離尺度が利用されている．

2-5-3 ハイブリッド型音楽推薦

最近では，楽曲に対するユーザの評価だけでなく，楽曲自体の内容も合わせて考慮するハイブリッド型音楽推薦システムの研究が活発になってきている [5]．その最新の成果として，本節では筆者らの研究を紹介する [6]．従来，推薦システムの研究はテキストベースで行うことが通例であり（本や文書を推薦する場合はその文章，音楽や映画を推薦する場合でもレビュー文に着目），楽曲それ自体の内容を考慮するものはほとんど存在しなかった．しかし，別個に研究されてきた推薦技術と音楽解析技術を効果的に組み合わせることで，優れた精度を保ちつつバラエティ豊かな推薦ができる音楽推薦システムを設計することができる．

我々は，Popescul らの Three-way Aspect Model に基づくテキストベースの推薦手法 [7]を音楽推薦に適用することを試みた．このときの主な課題は，テキスト内容と同一の枠組みで扱えるように楽曲内容を表現することである．各テキストの内容は Bag-Of-Words モデルを用いて語彙中の単語の出現頻度ベクトルで表現するのが一般的である．我々は各楽曲の内容を音色の重みベクトルとして表現するため，Bag-Of-Timbres モデルを提案した．具体的には，音響信号から抽出した MFCC 群に混合ガウス分布をフィッティングさせ，各要素分布の重みをベクトルの各次元に対応させた．ここで，混合される要素分布はすべての楽曲で共通であり，重みのみが楽曲ごとに異なることに注意する．このようにすることで，各要素分布がある代表的な音色に対応すると解釈できる．

Three-way Aspect Model は前節で解説した Aspect Model を拡張したグラフィカルモデルである（**図 2･8**）．ユーザの楽曲評価だけでなく，楽曲の内容を考慮するためのノードが追加された構造をもつ．まず，あるユーザ u がいて，そのユーザは確率 $p(z|u)$ でジャンル z を選択

するとする．次に，その状態のもとで，確率 $p(m|z)$で楽曲 m が，確率 $p(t|z)$で音色 t が観測されるとする．これらの確率を観測データ (ユーザの楽曲評価と楽曲の内容) から求めるには，通常の Aspect Model と同様に EM アルゴリズムを適用すればよい．楽曲推薦時には，ユーザ u に対し， $p(m|u) \propto \sum_{z,t} p(t|z)p(m|z)p(z|u)$ の大きい順に楽曲 m に順位づけを行う．

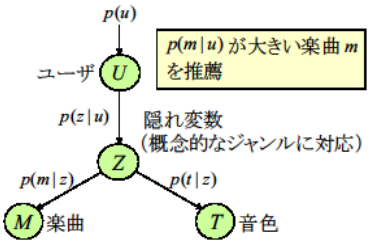

図 2・8　ハイブリッド型フィルタリング

2-5-4 音楽推薦技術の今後

今後の音楽推薦研究の方向性としてはいくつか考えられる．文献 8～11)は推薦研究全般に関する網羅的なサーベイであり，研究を進めるうえでの指針となる．これらのほか，特に音楽推薦に着目した視点では，以下のような課題が考えられる．

(1)　複数の高次音楽内容の利用

ユーザの音楽的嗜好を推定する際の手がかりとして，高次の音楽内容であるメロディ，リズム，コードなどを有効利用するための研究が必要である．音楽情報処理分野では，これまで音楽音響信号の内容解析技術が数多く提案されてきた．最近では市販 CD レベルの複雑な音響信号に対しても，かなりの精度で自動抽出できるまでになっている．しかし，解析された音楽内容の扱い方についての研究がほとんどなく，推薦や検索といったアプリケーションレベルでは，低次特徴量である MFCC などを利用することが多い．また，複数の音楽内容を総合的に考慮することでユーザの嗜好を推定する研究も重要である．

(2)　音楽情報処理技術と推薦技術との融合

音楽推薦システムを設計するうえで，音楽情報処理分野と推薦分野それぞれで培われた最新の技術を組み合わせることは有力な方法である．音楽情報処理分野において，推薦は検索の亜流であるとの見方が強いように思われる．そのため，本分野の中心的な研究テーマである音楽情報検索（MIR）に関する技術から出発し，それらを改良することで推薦を行おうとする研究が多い．しかし，推薦分野では，テキストを対象として古くから推薦技術の研究が行われてきており，最近ではハイブリッド型推薦技術に関して活発な議論がなされている．このような異分野の技術を音楽情報処理分野に積極的に取り組むことは重要である．

(3) 実用化のための効率性向上

音楽推薦システムを実用化するためには，巨大な楽曲データベースを扱うことができ，データ変化（新曲のリリースやユーザの入れ替わり）に対しても迅速に適応できるような技術の開発が必須である．多くの場合，推薦技術の開発や評価は実験室環境で行われており，システム評価に用いるデータは静的で，推薦を行うのに要する時間に制約はない．そのため，実用化において必須であるスケールアップを困難にしている．この問題を解決するには，推薦システムを設計する段から最終的なサービス運用時に必要な要件を列挙し，それを満たすように技術開発していくことが重要である．

■参考文献

1) Anderson, C., “The Long Tail: Why the Future of Business Is Selling Less of More,” Hyperion, 2006.
2) Breese, J. et al., “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” Proc. of UAI, pp.43-52, 1998.
3) Hoashi, K. et al., “Personalization of User Profiles for Content-based Music Retrieval based on Relevance Feedback,” Proc. of ACM Multimedia, pp.110-119, 2003.
4) Logan, B., “Music Recommendation from Song Sets,” Proc. of ISMIR, pp.425-428, 2004.
5) Celma, O., Ramir, P., “Foafing the Music: A music Recommendation System based on RSS Feeds and User Preferences,” Proc. of ISMIR, pp.464-467, 2005.
6) Yoshii, K. et al., “An Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model,” IEEE Transactions on Audio, Speech and Language Processing, vol.16, no.2, pp.435-447, 2008.
7) Popescul, A. et al., “Probabilistic Models for Unified Collaborative and Content-based Recommendation in Sparse-data Environments,” Proc. of UAI, pp.437-444, 2001.
8) 神嶌敏弘, “推薦システムのアルゴリズム(1),” 人工知能学会誌, vol.22, no.6, pp.826-837, 2007.
9) 神嶌敏弘, “推薦システムのアルゴリズム(2),” 人工知能学会誌, vol.23, no.1, pp.89-10, 2008.
10) 神嶌敏弘, “推薦システムのアルゴリズム(3),” 人工知能学会誌, vol.23, no.2, pp.248-263, 2008.
11) Adomavicius, G. and Tuzhilin, A., “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,” IEEE Trans. on Knowledge and Data Engineering, vol.17, no.6, pp.734-749, 2005.

■ 2 群－9 編－2 章

2-6 楽音分析・楽音合成

（執筆者：小坂直敏）［2011 年 6 月 受領］

楽音の分析とは，波形あるいはスペクトルに関する特徴量を用いて楽音の特徴を論ずることであり，また合成は波形あるいはスペクトルで楽音を表現することである．楽音合成は，コンピュータ音楽制作を目的として 1960 年代から発展してきた．初期は，加算合成，及び波形テーブル合成が主流であった．やがて 70 年代に変調方式が登場し，中でも FM 合成方式はそれまでの合成音品質を飛躍的に向上させた．更に，物理モデルが登場し，音質は更に一段向上した．一方音声通信技術で発展した分析／合成方式は様々なエフェクトを作るため，また楽音分析のために用いられてきた．説明の都合上，まず，様々な合成方式を分類し，概略を記したうえで楽音分析について述べる．

2-6-1 楽音合成方式とその機能

合成する楽音の機能を四つに分類する．機能とは音楽的な意図，あるいはコンテンツにおける役割という意味合いである．

1) 既存楽音の表現
2) 楽音の基本的加工
3) 楽音の応用的加工（エフェクト）
4) 創造的合成

1)～3)までは，対象とする楽音が存在し，1)は対象音そのものを合成するもの，2)，3)は対象楽音をもとにこれを加工するもの，4)は新たな音楽音を作るもので，いわゆる電子音やディジタルサウンドである．2)はピッチ，音色，テンポ（音声の場合は話速）などの特徴量をどれか一つを独立に制御する技術をいう．これらは音声技術と重複する．各種音合成方式をアルゴリズムにより分類し，それぞれがどのような機能で用いられているかを整理して**表 2･1** にまとめた．○印が主に用いられている使途である．以下は方式の概略である．

2-6-2 波形テーブル参照型

(1) 波形テーブル合成

これは最も基礎的な合成方式で，正弦波をはじめとする関数発生波を合成する，これらは波形をメモリに加工し，ここから周波数を変換を行って波形合成を行う．必要な演算はテーブル上のアドレシング計算，線形補間など演算コストが低い．

(2) サンプリング合成

楽音のシミュレータとして，実際の楽音の収録音をメモリに蓄えて合成させる手法である．任意長の発音を得るため，持続区間を繰り返して発音させるルーピング技術が中心課題である．

表 2・1　各種音合成方式とその主要な用途

区分	方式名		擬似楽音	基本加工	応用加工（エフェクト）	創造的合成
波形テーブル参照型	波形テーブル					○
	多重波形テーブル		○		○	○
	サンプリング合成		○			
	二連音素合成		○			○
	Karplus Strong 方式		○			○
ユニットジェネレータ	加算合成		○			○
	減算合成					
非線形処理	変調合成	リング変調			○	○
		振幅変調			○	○
		周波数変調	○		○	○
	ウェーブシェーピング				○	○
物理モデル	ウェーブガイド		○			○
分析／合成系	ボコーダ				○	○
	細粒合成				○	○
	LPC			○	○	
	フェーズボコーダ			○	○	
	正弦波モデル			○	○	
走査合成	地表面軌道合成					○
	Scanned synthesis		○			

(3)　二連音素合成

二連音素（diphone）を単位としてこれらの素辺を接続して合成する手法で，音声合成の手法と類似している．二連音素内では，補間可能部分と非補間部分があり，補間部分で伸縮を考慮しながら接続する．

(4)　Karplus Strong 方式

p 個の波形テーブルを用意し，新たな波形値を p 個前とその次の波形値の平均として順次定義していくと，動的な音色が得られる．この手法を拡張して，撥弦及び打楽器の音を合成する手法である．非常に自然な音質が得られる．

2-6-3 ユニットジェネレータ

(1)　加算合成

1960 年に Mathews が Bell 研究所にて開発した音楽用合成音作成ツール Music III 上で初めて実現された．正弦波発信器をユニットジェネレータとし，周波数と振幅の異なるいくつも重ね合わせることにより楽音を表現する手法である [1)].

(2) 減算合成

雑音，のこぎり波，矩形波など，倍音が豊富な音をユニットとし，これに様々なフィルタをかけることにより，周波数情報を削ぎ落としていくことにより合成する手法である．

2-6-4 非線形処理方式

(1) 変調合成

変調合成はリング変調（RM），振幅変調（AM），周波数変調（FM）が代表的な合成方式である．音楽応用での変調方式と通信技術との相違点は，通信技術では，搬送波には高域の周波数を，また変調波には音声などの信号をあてがうが，音楽応用では両者にまったく任意の信号を用いうることである．音楽的な機能として，対象音エフェクトをかけるほか，FM 合成 [2]では，楽音のシミュレーションがある．70 年代に考案されて以来，音質がそれまでの加算合成方式に比して飛躍的に向上し，物理モデルが登場するまで，楽音合成の主導的な地位を築いた．

(2) ウェーブシェーピング

チェビシェフ関数などを用いることにより，入力信号に非線形歪みを加えることにより高次倍音を豊かに付与する合成方式である [3]．

2-6-5 物理モデル

(1) ウェーブガイド方式

損失のない撥弦の振動は 2 階の線形微分方程式で表され，解は前進波と後進波の和として表すことができる．Smith は出力波形を，波形テーブルと，これに遅延を加えて加算を施す方法により実時間合成を達成し [4]，現在の物理モデルの基礎的な考え方となっている．

2-6-6 分析／合成

分析／合成方式とは，モデル化されたパラメータを原音から推定する方式が存在するものをいう．すなわち，任意の音源に対しモデル表現できることを意味し，音声通信技術としては必須条件である．楽音合成では，対象音を分析するためのツールとして用いたり，様々な加工（エフェクト）を施すための基本方式として用いられた．LPC，フェーズボコーダ，正弦波モデルなどが代表的な方式である．

(1) ボコーダ

音声通信方式としてのボコーダは,音声を音源と声道フィルタとに分解して表現している．音楽情報処理分野での応用は，音源を別の信号に置き換えたり，音声以外の楽音にも適用して新たなエフェクトとしての効果を確立した．後述の LPC は，音源を分析された残差を用いずに別のモデル音に入れ替えるとボコーダとなる．

(2) 細粒合成

音は音の小さな単位（粒子）の集合体として表現できる，という考えのもとで，粒子に分解し，これを再合成する際に時間軸を伸縮したりランダムに配置するなど，時間軸上での任

意に制御するエフェクトである．Gabor の定式化 [5] により，分析／合成系として表現できるが，実際の音楽応用では，細粒として，原音にガウス窓や三角窓などをかけ，自由に再配置されている [6]．

(3) LPC

LPC（Liniear Predictive Coding; 線形予測符号化）は代表的な音声表現技術である．音声を声帯振動を表す音源と声道情報を表すフィルタとに分解して表現して表すもので，チャネルボコーダの流れを汲む．音楽独自の応用としては，クロス合成がある．音声を本手法を用いて音源と声道情報に分解し，音源情報をほかの楽音に置き換えると，声道フィルタは音韻性を表すため，喋る楽音の合成が可能となる．

(4) フェーズボコーダ

対象音を帯域フィルタバンクに通過させ，個々の出力を位相情報も含めて忠実に分析／再合成する枠組みをいう．これをもとにエフェクトとして楽音の伸張を行うのが一般的である．Portnoff は STFT（短時間フーリエ変換）を用いて定式化を行った [7]．ボコーダとの違いは，原音の位相を保持している点であり，これが音質の良さにつながっている．

(5) 正弦波モデル

本方式は，時刻 t での楽音 $y(t)$を

$$y(t)=\sum_{i=1}^{L} A_i \cos(\varphi(t)) \qquad (6\cdot1)$$

の形式で表現するものである．ここに$\varphi(t)$は位相，A_iは L 個の調波のうち i 番目調波の振幅を表す．信号の分析には STFT を用いる．一般にこの方式は楽音の調波表現などには適しているが，雑音部の表現には適さない．Serra による SMS（Spectral Model Synthesis）[8]と MQ アルゴリズム [9] が代表的な方式として知られる．

2-6-7 その他の合成方式

その他には走査合成が挙げられる．これは何らかの形状をもとにして，オーディオレート（ピッチ周波数）で走査して音響信号とする方式である．地表面軌道合成は 3 次元上の平面の走査ルートを時変で制御し信号波形を得る手法，また Scanned synthesis は 1～2 Hz 程度のゆるやかな弦振動を空間方向にオーディオレートで走査するもので，前者は元データが固定されて走査が時変，後者は逆に元データが時変で走査が固定されている．この方式はまだあまり探求されていない．

2-6-8 楽音分析

楽音の分析には，一般に楽音をモデル表現できる上記の分析／合成方式を用いる．Risset と Mathews は，60 年代に周波数領域で金管楽器の調波ごとの時間－振幅特性を調べ，それを調波ごとに直線近似し [10]，これらを加算合成のパラメータとして与えられる合成プログラム Music IV を作成した．これを用いて，原音とほとんど違いのない音色を作るのに成功した．しかし，70 年代以降，次節で述べる新たなディジタル周波数分析／合成法が登場して以来，

フレームワイズの精密な分析が可能となった．以下ではこれらの分析について音声分析と対比して述べる．

2-6-9 音声分析と対比した楽音分析の特徴

音声は楽音や環境音と異なり，呼気流か声帯振動が音源となり，これが声道を通過して音声となる生成メカニズムの大枠は固定されている．そのため，古くからソースフィルタモデルとして，有声／無声，あるいは韻律，ひいては情動を表現する音源と，音韻性を表現するスペクトル包絡とに分離し，古くは LPC，また近年では STRAIGHT など，物理的な実態とグローバルに対応する数理モデルが一般的に用いられてきた．

一方，楽音は環境音まで含めて考えるとは上記の音生成モデルは必ずしも一般的でないため，信号モデルとしては，必ずしも音源とフィルタに分離されてこなかった．むしろ，80 年代後半以降，個別楽器構造に対応するかたちで物理モデルが発展した．

また，モデルに要求される仕様も音声と楽音では異なっている．音声は音韻性が第一義的であり，韻律や感情，という情動的な部分は優先度が低く発展してきた．一方，音楽は音韻性という呪縛はないかわりに，高音質への要求が厳しく，帯域幅も 20 kHz 前後までと音声よりも格段に高い．

楽音分析は，主に楽音合成の性能を向上させることを目的として行われる．そのほか，セントロイドや MFCC などの音色知覚と関連の深い各種特徴量の抽出は 2-1 節，ピッチ抽出，あるいはピッチに立脚したメロディ，和声，あるいはリズム，テンポなどの音楽の高次情報を抽出するための分析は 2-2 節あるいは 2-4 節で，また，多音源のピッチ抽出，音源分離などは 2-3 節で論じている．ここでは，一つの楽音（音脈）の合成あるいはその加工音の合成を前提にするときの周波数分析法について述べる．

音声では耳は位相に疎いという考えで，伝統的に位相よりもパワースペクトルやフィルタの振幅特性に興味がもたれてきた．一方，楽音分析ではパワースペクトルに特化するより，あるがままの信号をどう扱うか，ということに興味がもたれ，位相情報は最後まで捨てずに保持するという姿勢を保ってきた．すなわち，周波数情報に分解しながら，最終的に正確な波形表現を行う，という考えで発展してきた．この考え方で，モデルはフェーズボコーダと正弦波モデルとに大別できる．

フェーズボコーダは，音声信号を帯域フィルタバンクに分解する周波数領域での表現として，Flanagan が提案した [11]．これは瞬時周波数の定義も与えるなど，明確に数理表現されていたが，Portnoff がディジタル表現し，特に FFT を用いて高速に狭帯域フィルタ表現が可能となった [7] ことから，容易に使えるツールとして進歩した．**図 2･9** は Portnoff の表現で k 番目の帯域フィルタの入出力を示したものである．

しかし，分析／再合成音として性能の良いフェーズボコーダも，信号の伸縮など，時間領域での加工に残響感を与える歪みが登場することも知られた．これを改善するため，新たな周波数領域表現として，追跡型のフェーズボコーダが登場した．これが正弦波モデルとして発展していく．一方，フェーズボコーダの改良の方向性として周波数分析における，狭帯域信号の分析精度の向上が行われてきた．

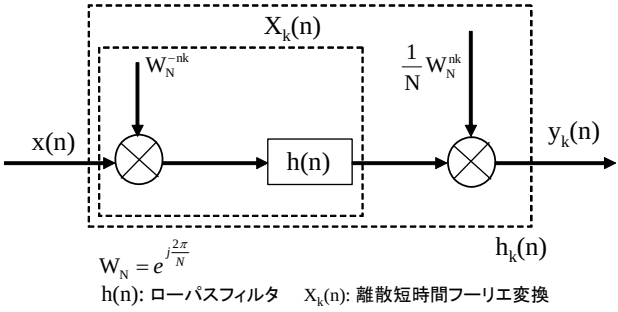

図 2・9　k 番目の帯域フィルタの入出力

一つには解析解が徹底的に求められた．図 2･9 で，帯域フィルタの出力信号は実信号であるが, その信号をフィルタ表現で行うと $1/N \cdot W_N X_k(n)$ と複素信号の乗算のかたちで表される．任意の周波数の正弦波入力を考えると，時間領域の定数と三角関数の和として表現される窓信号が，周波数領域での離散表現はディリクレ核が複合したものとなり，この信号の周波数を正確に求めるのはさほど単純ではない．文献 12)で，Puckette らは正弦波入力時の周波数の振る舞いを解析的に詳細に論じている．このような検討は雑音下での正弦波信号の推定精度向上が期待できる．また，これらの詳細の検討の結果はフェーズボコーダの変形時の音質向上と結びついており, 時間伸縮，ピッチ変換などの変形が文献 13),14)などで論じられている．

しかし，フェーズボコーダでは，単一の正弦波でさえ，その周波数や分析条件によっては複数の帯域にエネルギーが分散する．例えば周波数 bin の整数倍とその半分の周波数などである．分析／再合成ではこの現象は問題ないものの，加工時には歪みの原因にもなる．正弦波モデルはこの欠点を解消し，一つの正弦波には一つの三角関数波を対応させるモデルである．しかし，その分雑音の表現力はより低下し，音声では破裂音，また楽音では打楽器の立ち上がり部などの再合成時の音質が劣化している．

阿部らは，正弦波モデルにおいて，Flanagan の定義するフィルタ出力の偏角を時間微分した瞬時周波数という考えを推し進め，中心周波数のまわりでの不動点をみつけ，これを周波数アトラクタとして表現することにより，より精緻化した瞬時周波数を提案した [15]．狭帯域信号の瞬時周波数は，聴覚との対応もよく，理解しやすいためである．

このほか, 過渡的な信号の効率的表現では文献 16)などもある. 文献 15)で過渡的な信号や，帯域をよぎって周波数が変わる信号に対応できるよう，時間軸を一次式により変換する方法も提案している．これにより，調波構造をもつ信号の調波抽出精度が向上している．これにより，ソナグラムなどの時間-周波数表現が向上している．

2-6-10 今後の楽音分析と楽音合成

信号表現については，音声研究の動向を更に取り込む方向に発展するであろう．これは一つには歌声合成が盛んになってきたこともあるが，研究者の発表のコミュニティの中に音楽と音声の双方を扱う研究者が増えつつあることも要因である．分析は環境音のほか，より広

範な音や音楽表現が対象となろう. また, それに見合った表現が探されていくと考えられる.

■参考文献

1) J.A. Moorer, “Signal Processing Aspects of Computer Music: A Survey,” Proceedings of the IEEE, vol.65 No.8, pp.1108-1137, Aug. 1977.
2) J.M. Chowning, “The synthesis of complex audio spectra by means of frequency modulation,” Journal of the Audio Engineering Society, vol.21, no.7, Sep. 1973.
3) M.Le Brun, “Digital Waveshaping Synthesis,” Journal of Audio Engineering Society, vol.27, No.4, pp.250-266, Apr. 1979.
4) J.O. Smith, “Physical modeling Using Digital Waveguides,” Computer Music journal, vol.16, no.4, Winter 1992.
5) D. Gabor, “Acoustical Quanta and the Theory of Hearing,” Nature, vol.159, no.4044, pp.591-594, may, 1947.
6) C. Roads and J. Strawn, “Granulay Synthesis of Sound, Foundations of Computer Music, Cambridge, Massachusetts, 1987.
7) M.R. Portnoff, “Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform,” Acoust, Speech, Signal Processing, vol. ASSP-24, no.3, pp.243-248, June 1976.
8) Serra, X. and J. Smith, “Spectral modeling Synthesis,” Computer Music Journal vol.14, no.4, pp.12-24, 1990.
9) R.J. McAulay and T.F. Quatieri, “Speech Analysis/Synthesis Based on a Sinusoidal Representation,” IEEE Trans. On Acoust., Speech, and Signal Processing, vol. ASSP-34, no.4, Aug. 1986.
10) J.C. Risset and M.V. Mathews, “Analysis of musical instrument tones,” Physics Today, vol.22, no.2, pp.23-30, 1969.
11) J.L. Flanagan, “Speech Analysis Synthesis and Perception,” Springer-Verlag, pp.378-386, 1972.
12) M.S. Puckette and J.C. Brown, “Accuracy of Frequency Estimates Using the Phase Vocoder,” IEEE Trans, on Speech and Audio Processing, vol.6, no.2, pp.166-176, March 1998.
13) J. Laroche, “Improved Phase Vocoder Time-Scale modification of Audio,” IEEE Trans. on Speech and Audio Processing, vol.7, no.3, May, 1999.
14) J. Laroche and M. Dolson, “New Phase-Vocoder Techniques for Real-time Pitch-shifting, Chorusing, Harmonizing and other exotic Audio Modifications,” Journal of the Audio Engineering Society, vol.47, no.11, Dec. 1999.
15) 阿部敏彦, 誉田雅彰, “瞬時周波数アトラクタに基づく正弦波分析合成法, ” 信学技法, vol.SP 2002-169, pp.1-4, Jan. 2003.
16) Axel Roebel, “A New Approach to Transient Processing in the Phase Vocoder,” Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03), London, Sep. 2003.

■2群－9編－2章

2-7 歌声分析・歌声合成

(執筆者：中野倫靖)[2011年6月 受領]

歌唱は，音楽という人間の知的活動の中において，多くの人にとって最も身近な表現手段の一つであろう．音楽とのかかわり方として「作曲」「演奏」「聴取」の三つを考えると，歌唱は「演奏（歌う)」「聴取」という行為に関して，多くの人がかかわりやすい．歌唱の音源である「声」は，ほとんどの人が日常的に接しているため，同一言語であれば「聞き取る」「歌う」といった基本技能に特別な訓練を必要としない．したがって，音楽的知識や経験が乏しい人であっても，好きな歌手の曲をカラオケで歌ったり，楽曲のメロディを「ラララー」などのようにハミングしたりでき，楽曲中からボーカルを聴き分けることも容易に行える．その結果，多くの人は「感動する歌が聴きたい」「うまく歌いたい」といった，歌唱のより深い楽しみ方を望んでいる．歌声の分析と合成に関する知見や技術は，そのような楽しみを提供するために重要であり，研究成果の適用範囲が音楽家だけに留まらない点で，応用可能性が大きい．

2-7-1 歌声の分析／合成研究の意義とその相互関係

歌声の分析及び歌声の合成に関する研究はそれぞれ，学術的にも実用的にも意義がある．歌声分析に関する研究は，歌唱という側面から人間を知るという知的探求の観点から学術的に意義があり，その結果を歌唱指導などへ応用できる点で実用的である．また，歌声合成に関する研究は，歌声生成機構の計算機上への実現方法を知るという観点から学術的に意義があり，合成技術は歌唱付き楽曲の制作を計算機上で行えることを支援する．

ここで，歌声の分析と合成には密接な関係があり，お互いの分野の知見や技術が，それぞれの研究分野の発展に繋がることも多い．まず，歌声合成においては，歌声分析の知見が必要不可欠である．なぜなら，歌唱音声を特徴づける音響的特徴を知らなければ，「人間らしい」歌唱音声を合成することはできないからである．また逆に，歌声合成のための技術や知見も，歌声分析の研究分野の発展につながる．例えば，高品質な歌声合成手法を確立することで，合成のための各特徴と歌声知覚の関係を調査する心理実験が実施できるようになる．このような心理実験の実施は，人間の歌声知覚機構の解明につながるといえる．

2-7-2 歌声分析／合成のための基本知識及び要素技術

「歌声」を扱うための基本知識や要素技術は，「話声」を扱う音声情報処理 [1] とほとんど同様である．したがって，歌声を扱う場合には，既知の知見や要素技術を用いて，「歌声に特有の」特性分析や，合成のためのモデル化を行うことが多い．すなわち，歌声はメロディや音楽的なリズムをもつこと，歌唱様式や表情付けによって，歌い方や発声のされ方に様々なバリエーションがあることを考慮して扱う．

歌声の分析や合成に必要となる要素技術として特に重要なのは，基本周波数（*F*0: Fundamental Frequency）の推定と，スペクトル包絡やフォルマント周波数の推定である．基本周波数は歌声の音高に相当し，スペクトル包絡やフォルマント周波数は音韻や声質を決定

づける．以降，歌声の分析と合成に関する従来の知見や技術を述べる．ただし，本節では紙面の都合上，すべての情報を載せることができないため，文献 2〜5) なども参考にしていただきたい．

2-7-3 歌声分析における知見

これまで，歌声特有の特性分析としては，主に基本周波数変化の特性とスペクトル特性に着目して研究が行われてきた．まず，基本周波数変化の概形は，多くの場合メロディに対応しているが，それとは別に歌声特有の動的な変動成分（F_0：動的変動成分）の存在が報告されている[4,6]．その結果，F_0動的変動成分としては，歌唱スタイルや歌唱者に依存せずに，次の 4 種類が存在することが明らかになっている（**図 2･10**）[4]．

オーバーシュート（Overshoot）：
滑らかな音高変化，及びその直後に目的音高を超える時間的な変動成分

ビブラート（Vibrato）：
同一音高区間で観測される 4〜8 Hz の準周期的な変動成分

プレパレーション（Preparation）：
音高変化直前に変化とは逆方向に触れる瞬時的な変動成分

微細変動（Fine fluctuation）：
発声区間全体に観測される不規則で細かい変動成分

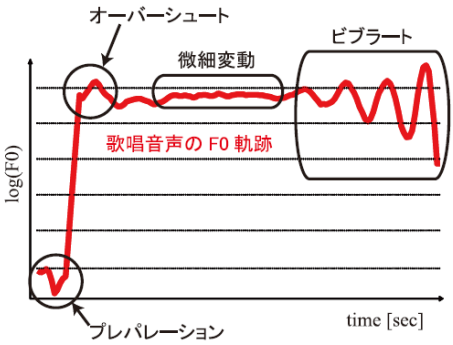

図 2・10　4 種類の F_0 動的変動成分

このような歌声の基本周波数の軌跡は，モデル化を行ってそれにパラメータフィッティングすることで分析を行う[4,6]．基本的には二次系のモデルが考えられているが，F_0軌跡の相平面（F_0－ΔF_0 平面）を利用して F_0軌跡を表現する手法の提案もある[7]．

歌声特有のスペクトル形状については，スンドベリ（Sundberg）が，Singer's formant と呼

ばれる 3 kHz 付近に現れるスペクトルピークの存在を明らかにしている [2]．**図 2･11** に，男性テナー歌手が日本語 5 母音を歌った場合と話した場合の長時間平均スペクトルを示す．音声データは，文献 8) のものを利用した．Singer's formant は，第 3～第 5 フォルマントが互いに近づいて一つの山を形成したものであり，その周波数は母音によらず一定であるとされている [2]．これは「響く声」「通る声」「張り・艶のある声」を特徴づけ，当初は男性歌手（コンサート歌手，オペラ歌手）の母音（有声音）にのみ存在すると報告されていたが，その後，女声や邦楽歌唱にも存在することが明らかになった [10]．

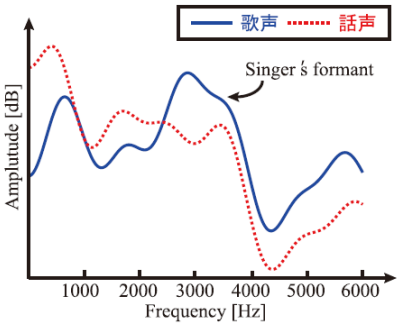

図 2・11　Singer's formant の例
日本語 5 母音における，テナー歌手の歌声と話声の長時間平均スペクトル包絡．3kHz 付近にピークがあることが確認できる．

また最近では，歌声中のブレス（息継ぎ，吸気）音に着目した研究がある [9]．中野らは，ポピュラー音楽における歌声のブレス音のスペクトル形状を分析し，歌唱者・曲・言語・歌唱力が異なっても，1.6 kHz（男声）及び 1.7 kHz（女性）付近にスペクトルピークが存在することを報告している [9]．**図 2･12** に，歌唱音声中のブレス音における長時間平均スペクトルを示す．

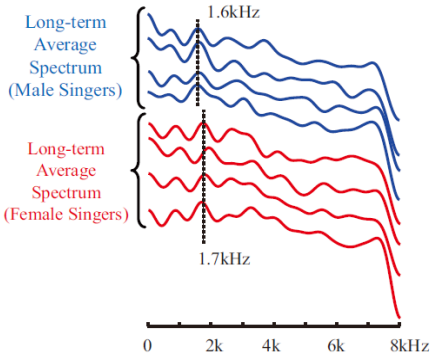

図 2・12　ブレス音の長時間平均スペクトル包絡の例．1.6 kHz，1.7 kHz 付近のピークが確認できる．

2-7-4 歌声合成技術

歌声合成は，基本的に音声合成と同様の技術が用いられており，現在は，波形素片接続方式[11]，フォルマント合成方式[12]，分析合成（ボコーダ）方式[3, 13]，HMM合成方式[14]，による歌声合成が主に研究されている．そのほかには，周波数変調（FM）合成，フォルマント波形関数（FOF）合成による歌声合成の研究がなされてきた[13]．これらの合成方式の中でも，サンプリングした歌声波形の素片を連結する波形素片接続方式は，人間の歌声データを用いるため自然性が高く(肉声に近く)，歌声合成の市販ソフトウェアのほとんどはこの方式である．

これらの技術をユーザが利用するためには，歌詞と楽譜情報（何を歌わせるか）と，歌唱の表情（どう歌わせるか）を入力するインタフェースが必要となる．市販ソフトウェアとしては主に，前者として歌詞と楽譜（音高・発音開始時刻・音長）を与え[14, 15]，後者としてユーザが表情に関するパラメータを調整する方法[15]が採用されている．歌声合成システムを実用的に扱うためには，このようなパラメータをユーザが細かく調整できる必要があるといえる．

このほか，歌い方や歌唱スタイルのモデル化[14, 16]，演奏記号（crescendo など）による合成[17]などがあり，これらの研究は歌い方を抽象化して扱うために必要である．また，歌詞のみを与えた合成手法[18]も研究されており，これは作曲ができないユーザでも利用できる．最近では，ユーザの話声から声質を保存したまま歌声を合成する方法[3, 16, 17]や，歌声間のモーフィング[19]，ユーザの歌声から歌い方を真似して合成する方法[20, 21]なども研究されている．

2-7-5 歌声に基づくアプリケーション

歌声の分析及び合成に関する研究成果は実用性･応用可能性が大きく，歌声を利用した様々なアプリケーションが考えられる．ここではいくつかの事例を紹介する．

(1) 音楽情報検索

歌唱による検索で代表的なものは，ハミング検索（QBH: Query-by-Humming）である[22, 23]．ハミングとは「ラララー」などの歌詞を伴わない歌唱を指し，楽曲タイトルが思い出せない場合でも，既存メロディを歌うことで楽曲を検索できる．最近では，QBSH（Query-by-Singing/Humming）と呼ばれ，歌詞情報を検索に利用する研究もある[24]．そのほかには，声質が類似している歌手の楽曲を検索するシステム[25]といった研究事例がある．

(2) 歌唱力向上支援

歌がうまくなりたいという欲求は，プロの歌手だけでなく一般ユーザにもあるため，歌唱力向上を助けるアプリケーションは有用である．これまで，歌唱トレーニングの支援を目的として，歌唱に関する音響特性をリアルタイムに可視化して，ユーザにフィードバックするシステムが提案されてきた[5]．ここで，特に重視されるのは音高で，教師音と同時に提示する，ビブラートを自動検出するといった機能が提案されている．

(3) 楽曲制作

楽曲制作場面においては，これまで述べてきた歌声合成技術の発展が最も望まれているだろう．現在でも既に市販ソフトウェアが DTM（DeskTop Music）場面で活用され始めている．

また，マイクを通したハミングや歌唱による，メロディラインの入力も既に商用化されている．ドラムパートを入力したい場合にも，Beatboxing やロ（くち）ドラムと呼ばれる，歌唱によるドラム音の表現が研究されている [5]．更に，歌唱中のブレスを自動検出し，ブレスを消したり強調したりする機能も実用化されている [9]．

2-7-6 今後の展望

これまで述べてきたように，最近では歌声分析や歌声合成に関する研究が盛んに行われており，歌声の特性についても徐々に解明されてきている．また，現時点で商用化されている歌声合成システムは，若干不自然さは感じるものの，非常に高い水準の自然性で歌声を合成できる．今後は，歌声の個人性や歌声合成の自然性を向上させるための特性解明，歌唱力の解明とその応用としての歌唱力向上支援，様々な歌い方や声質をユーザが自由自在に操作できる歌声合成，歌唱力や歌い方に着目した音楽情報検索，などが研究対象として興味深い．

■参考文献

1) 古井貞熙, “音声情報処理,” 森北出版株式会社, p.173, 1998.
2) J. Sundberg（著），榊原健一（監訳），伊藤みか，小西知子，林良子（訳），“歌声の科学,” 東京電機大学出版局，p.218，2007.
3) 河原英紀，片寄晴弘，“高品質音声分析変換合成システム STRAIGHT を用いたスキャット生成研究の提案,” 情報処理学会論文誌，vol.43，no.2，pp.208-218，2002.
4) 齋藤　毅，“歌声知覚・生成機構の解明に向けた歌声合成システム構築に関する研究,” 北陸先端科学技術大学院大学博士論文，p.139，2006.
5) 中野倫靖，“歌唱理解及び歌唱表現の解明とその応用システム構築に関する研究,” 筑波大学博士論文，p.216，2008.
6) H. Mori, W. Odagiri and H. Kasuya, “F0 Dynamics in Singing: Evidence from the Data of a Baritone Singer,” IEICE Trans. Inf. & Syst., vol.E87-D, no.5, pp.1086-1092, 2004.
7) 大石康智，後藤真孝，伊藤克亘，武田一哉，“歌声の旋律と動的変動を特徴付けるための確率的な表現手法に関する検討,” 情報処理学会研究報告音楽情報科学 MUS，vol. 2007-MUS-07-19, pp.111-118, 2007.
8) 中山一郎，“日本語を歌・唄・謡う,” 日本音響学会誌，vol.11, no.59, pp.688-693, 2003.
9) 中野倫靖，緒方　淳，後藤真孝，平賀　譲，“無伴奏歌唱におけるブレスの音響特性と自動検出,” 日本音響学会 2008 年春季研究発表会講演論文集, 1-11-12, 2008.
10) 安藤由典，“第 11 章歌い声,” 新版楽器の音響学，音楽之友社，pp.235-246, 1996.
11) J. Bonada and S. Xavier, “Synthesis of the Singing Voice by Performance Sampling and Spectral Models,” In IEEE Signal Processing Magazine, vol.24, Iss.2, pp.67-79, 2007.
12) J. Sundberg: “The KTH Synthesis of Singing,” Advances in Cognitive Psychology, Special issue on Music Performance, Vol. 2, Iss. 2-3 (2006)
13) P.R. Cook, “Singing Voice Synthesis: History, Current Work, and Future Directions,” Computer Music Journal, vol.20, no.3, pp.38-46, 1996.
14) 酒向慎司，才野慶二郎，南角吉彦，徳田恵一，北村　正，“声質と歌唱スタイルを自動学習可能な歌声合成システム,” 情報処理学会研究報告音楽情報科学 MUS，vol. 2008-MUS-74-7，pp.39-44，2008.
15) 剣持秀紀，大下隼人，“歌声合成システム VOCALOID－現状と課題,” 情報処理学会研究報告音楽情報科学 MUS，vol.2008-MUS-74-9，pp.51-58，2008.
16) 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,” 情報処理学会研究報告音楽情報科学 MUS, vol. 2008-MUS-74-5, pp.25-32, 2008.
17) 森山　剛，小沢慎治，“好みの歌唱様式による歌詞朗読音声からの歌唱合成,” 情報処理学会研究報告音楽情報科学 MUS，vol. 2008-MUS-74-6，pp.33-38，2008.

18) 米林裕一郎，中妻　啓，西本卓也，嵯峨山茂樹，“Orpheus: 歌詞の韻律を利用した Web ベース自動作曲システム，”情報処理学会インタラクション 2008 論文集，pp.27-28，2008.
19) 河原英紀，生駒太一，森勢将雅，高橋　徹，豊田健一，片寄晴弘，“モーフィングに基づく歌唱デザインインタフェースの提案と初期検討，”情報処理学会論文誌，vol.48，no.12，pp.3637-3648，2007.
20) J. Janer, J. Bonada and M. Blaauw, “Performance-Driven Control for Sample-Based Singing Voice Synthesis,” In Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06), pp.42-44, 2006.
21) 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，”情報処理学会研究報告音楽情報科学 MUS，vol. 2008-MUS-75，pp.49-56，2008.
22) 蔭山哲也，高島洋典，“ハミング歌唱を手掛りとするメロディ検索，”電子情報通信学会論文誌．vol.J77-D-II，no.8，pp.1543-1551，1994.
23) A. Ghias, J. Logan, D. Chamberlin and B.C. Smith, “Query by humming: Musical information retrieval in an audio database,” in Proc. ACM Multimedia, vol.95, pp.231-236, 1995.
24) M. Suzuki, T. Hosoya, A. Ito, and S. Makino, “Music Information Retrieval from a Singing Voice Using Lyrics and Melody Information,” EURASIP Journal on Advances in Signal Processing, vol.2007, Article ID 38727, p.8, 2007.
25) 藤原弘将，後藤真孝，“VocalFinder: 声質の類似度に基づく楽曲検索システム，”情報処理学会研究報告

■2群－9編－2章

2-8 自動作／編曲

(執筆者：平田圭二)[2011年6月 受領]

2-8-1 位置づけと分類項目

作曲や編曲は，人が行う知的で創造的な活動の一つに数えられる．計算機に音楽を作／編曲させるという目標への取組みは，人工知能（Artificial Intelligence）という言葉の誕生（1956年ダートマス会議）とほぼ同時期といえるほど古い（1957 年に作曲されたイリアック組曲（Illiac Suite）がその嚆矢）．当時は，音楽的な知性や創造性に対する科学的な興味，計算機を利用した新しいスタイルの音楽の開拓，作／編曲の技術や才能をもたない人に対する表現機会の提供などが目標であったと考えられる.その後の技術発展の経過や成果を振り返ると，現在これら目標の一部は実現されてはいるものの，まだ不十分なところも残っている．近年においても上述の目標はある程度有効だが，ゲーム操作中のBGM，公共空間での著作権フリーな楽曲，UGC/CGM向けの楽曲など，大量生産／大量消費される楽曲への需要が高まるにつれ，工学的な応用という側面も強く意識されるようになった．

自動作／編曲を研究対象として眺めてみよう．もし「自分が作るような（嗜好を反映したような）曲で，音楽的に一定のクオリティを満足するレベルをもったものを自動作曲させたい」という研究目標を設定したとすると，次にそれを技術的な課題へとブレークダウンしなければならない．例えば，自分が作るような曲とはどのようなもので，その情報をどうやって計算機に伝えるのか．計算機はその情報をどう処理するのか，どうやって音楽的に一定のクオリティを満足させるのかなどである．ここで重要なことは，目標や音楽的対象が異なればそれらを適切に実現する技術も評価手法も異なってくるので，まず自動作／編曲を様々な観点から分類し，ほかの研究者による再現や追試が可能な程度に目的，条件，方法などを明確化することである（**表 2･2**）． 表中，例えば最上段の「目標」の欄の意味は，自動生成された楽曲だけで出来上がるような作品を作るのか，それとも楽曲にビデオや絵や写真などほかのメディアを組み合わせて作品とするのかのいずれかを選択する項目である．続く，長期間繰り返し聴くのかそれとも 1 回だけ聴くのかという選択肢は，先の選択肢とはまた独立の選択肢であり，実際の目標はこれら選択肢の組合せの数だけ存在する．

2-8-2 本来的に備える困難な性質

このように目的，条件，方法などを限定し明確化しても，自動作／編曲という課題が本来的に備える性質により，研究課題の設定と実験結果の評価法（測定・比較の方法）にはどうしても曖昧で不完全な部分が残ってしまう．その性質とは，(1)技術そのものの評価と生成された表現の評価が対応していない点，(2)その楽曲が音楽規則を守っているという意味で正しいか否かということとその人の意図を表現しているかという意味で好ましいか否かということが対応していない点，(3)楽曲は楽曲のみで存在しているわけではないという点である．

まず(1)に関して，これまで様々な自動作／編曲手法が提案され，中には，他分野で高い有用性をもつことが実証された技術も含まれているが，そのような技術が高い音楽性や選好性を保証するわけではない．技術の改良が直接的間接的に音楽性や選好性の改良に貢献するか否かは曖昧である．技術的に改良すべき点を同定するような評価を実現することも難しい．

表 2・2　自動作／編曲に関する分類項目（の一部）

目標:	•楽曲自体が作品　•長期間繰り返し聴く　•自分だけ聴く　•専門家 •作品の一部(BGM)　•1回だけ聴く　•他者に聴かせる　•非専門家
音楽的対象:	•旋律，和声，リズム(音楽の3要素) •パート(ソプラノ，アルト，テノール，バス) •楽器(ピアノ，管楽器，弦楽器，打楽器) •曲長(4/8/16小節，あるいは数百小節) •バンドスコア/合奏譜，即興演奏 •スタイル/ジャンル
生成方式:	•決定的　－ルール －事例 －統計
	•非決定的　－確率(乱数，カオス) －インタラクション(即興演奏，進化計算)
制御方式:	•数値を与える　•完全自動 •形容詞を与える　•半自動/支援 •例示曲を示す

人は，その表現が自分の意図に照らし合わせて適切なのか／好ましいのかどうかよく分からない場合があるが，これは，自動作／編曲における汎用的な評価尺度の構築が難しいことを示唆している．

次に(2)に関して，一般に，人の意図を効果的に表現したり，人が求めているような質の高い音楽を生成するには，構造的な普遍性を規定する音楽的な規則に従って楽曲を生成するだけでは不十分で，そこから多少逸脱することで音楽性や選好性を高める必要があると考えられている[#]．しかしその逸脱に規則性を認めるのは難しく，また統計的に最尤な振る舞いが常に妥当とも限らない．音楽性を高める逸脱と選好性を高める逸脱の区別も曖昧である．更にユーザの意図を計算機に伝達するには，ユーザの意図を何らかの記号の形式で表現しなければならない．ユーザの意図を最も具体的に指示する方法は，楽譜エディタ上で生成する楽曲の 1 音 1 音をユーザが操作することであろう．逆に最も抽象的な指示方法は，ユーザがいくつかの大域的なパラメータや指示を与えて自動作曲するものであろう．前者は精密な操作が可能だがユーザに高度な音楽的スキルを求める．一方後者の指示方法は簡便であるがユーザの意図を正確に伝達するのが難しい．この意図指示に関する抽象度と操作性，あるいは簡便さと意図の伝達度は一般にトレードオフの関係にある．

最後に(3)に関して，システムが楽曲を自動生成して五線譜上の音符として記述したとしよう．しかし人がその楽曲を聴取するときは，五線譜上の音符の情報だけを聴きとっているわけではない．編曲，演奏，音色，その楽曲に関連した文章（例えばアルバムのライナーノーツ），楽曲がビデオ作品の BGM として利用されるようなときは映像に関するような付帯的な情報も同時に鑑賞している．鑑賞には，聴取環境や以前聴取した楽曲との関連性も影響を及ぼす．したがって，自動作／編曲の技術は，自動作／編曲された楽曲を鑑賞する環境まで

[#] そもそも音楽には正しい楽曲か否かを判別する明確な規則が存在しない．対して，例えば自然言語では，ある文法規則に関して正しい文と非文の区別は容易である．

考慮することが望まれる.

前述の(1)や(2)で述べた曖昧さが音楽自身に由来するとすれば, (3)は外部的な要因に由来する曖昧さに関連している.

自動作／編曲には上に挙げた本来的に備える困難な性質があるため，従来なら効果を発揮してきたような方法論が，うまく使えない場面が出てくる．例えば，音楽情報処理の分野では（ほかのメディア処理の分野でも），ほかの研究者による実験の再現や新手法の検証を可能とするために，標準的な正解集（コーパス）を構築することが一般的である．しかし，自動作／編曲の場合，もしシステムがそのコーパスに含まれないような楽曲を生成してもそれを単純に不正解とすることはできない．なぜなら，コーパスに含まれないような新しい表現を生成する自動作／編曲システムの実現というのを目標に掲げることもできるからである．

音楽情報処理において（ほかのメディア処理の分野でも），個人の選好を適切に扱える技術の一つにソーシャルフィルタリングがある．この技術は，既に大量に存在している作品やコンテンツに対する個人の選好（振る舞い）を処理の対象とする．一方，自動作／編曲は，基本的に，この世に存在しない未知の楽曲を生成する技術である．もし自動作／編曲で生成された楽曲の評価にソーシャルフィルタリング技術を応用する場合には，楽曲の類似度を定義する（楽曲のモデル化をする）必要がある．しかし一般的に，研究コミュニティが芸術性や選好性の要素を含む類似度に関して合意に到達するのは難しいだろう.

2-8-3 研究事例紹介

自動作／編曲分野を網羅的に紹介した文献には 1) や Wikipedia の Algorithmic Composition のページ [2] があるので参考にされたい. 以下代表的な自動作／編曲システム研究を紹介する.

音楽学者である David Cope は 1981 年より作曲システム Experiments in Musical Intelligence（EMI）の開発を開始した [3, 4]．EMI はある類似度をもって内部データベースから適切なメロディ断片を検索し，SPEAC という音楽文法に従って断片を接続して楽曲を生成する．EMI は数多い自動作曲システムの中でも質の高い楽曲を創作することで有名である.

コンピュータ音楽の作曲家である Robert Rowe は Cypher というインタラクティブな自動作曲システムを製作した [5]．Cypher に入力された音楽は，listener モジュールによって特徴量空間にマップされ，フレーズが検出されて音楽を理解する．player モジュールはその情報をもとに実時間でユーザに応答を返す．インタラクティブにすることで，ユーザの意図の曖昧さを低減させることに成功した.

Generate & Test手法を使うと，自動生成の問題の一部を認識の問題に帰着できる．Gerhard Widmerは, 対位法の楽曲事例から作曲ルールを帰納推論するシステムをProlog 言語を用いて構築した [6]．帰納推論する際，既存の音楽理論（Generative Theory of Tonal Music（GTTM）とImplication-Realization Model（IRM））を背景知識とすることで，音楽知識を表現するための基本概念が与えられ, 学習の効率が高まった．こうして学習されたルールをGenerate & TestのTestに用いることは，上述(1)，(2)の曖昧さ克服に効果的であろう.

Francois Pachet は，ユーザの演奏スタイルを実時間で学習する Continuator という即興演奏器を作成した [7]．入力されるフレーズのピッチを可変次数マルコフモデルによって学習し，その獲得されたモデルに基づいて応答の旋律を生成する．和声やリズムも学習，模擬できるように学習モデルに修正を加えた．入力された旋律の学習と応答の生成が同時に実時間オン

ラインで実行されるようアルゴリズムに工夫を加えた．この工夫により，例示による意図指示が可能となり，上述の意図指示に関する抽象度と操作性のトレードオフの課題に対処している．

浜中らは，既存の音楽理論（GTTM と Tonal Pitch Space（TPS））が定義する旋律のタイムスパン簡約構造に基づいて，ユーザが弾く可能性の高い音列を予測する予測ピアノを制作した [8]. GTTM と TPS では, タイムスパン簡約構造の安定度を算出する手順が与えられており，安定度の高い構造ほど音楽的に正しい解釈を与えていると考えられている．音楽的に正しい旋律を生成するという曖昧な課題を，音楽的に妥当な解釈をもつ旋律を生成するという課題に置き換え，音楽に内包されている曖昧さの問題を軽減させた．

深山らは，ユーザが与えた歌詞の韻律を反映したような旋律を自動作曲するシステム Orpheus を構築している [9]. 様々な旋律候補の中から，テンプレートとして与えられている和音パターンやリズムパターンを最もよく満足する旋律を出力する．自動作／編曲に，歌詞の韻律と旋律の対応という新しい視点を持ち込んだ点は興味深い．更に Web 上では，人工音声による歌声トラックや伴奏トラックを付加する編曲のサービスが提供されており，生成した旋律をどのように聴かせるかという環境まで考慮した例の一つである．

安藤の構築した作曲支援システムでは，意図指示に関する抽象度と操作性のトレードオフの課題に対処するため，クラシック音楽の作曲手法を模擬するような木構造型遺伝子と進化プロセスを用いたインタラクティブな遺伝アルゴリズム（Genetic Algorithm）を採用している [10]. 楽曲プールを世代更新する際，人が楽曲プール中の候補楽曲を直接評価し淘汰するか否かを決定する．評価作業を行うユーザの負担という課題はあるものの，human-based computation の観点からも興味深い方法論である．

■参考文献

1) Gerhard Nierhaus, “Algorithmic Composition: Paradigms of Automated Music Generation,” Springer, 2009.
2) http://en.wikipedia.org/wiki/Algorithmic composition
3) David Cope, “Experiments in Musical Intelligence,” A-R Editions, Inc., 1996.
4) David Cope, “A Musical Learning Algorithm,” Computer Music Journal, vol.28, no.3, pp.12-27, 2004.
5) Robert Rowe, “Interactive Music Systems－Machine Listening and Composing,” The MIT Press, 1993.
6) Gerhard Widmer, “Qualitative Perception Modeling and Intelligent Musical Learning,” Computer Music Journal, vol.16, no.2, pp.51-68, 1992.
7) Francois Pachet, “The Continuator: Musical Interaction with Style,” In Proceedings of ICMC 2002, pp.211-218.
8) Masatoshi Hamanaka, Keiji Hirata, Satoshi Tojo, “Melody Expectation Method Based on GTTM and TPS,” In Proceedings of the 9th International Conference on Music Information Retrieval, ISMIR2008, pp.107-112, 2008.
9) 深山　覚，中妻　啓，米林裕一郎，酒向慎司，西本卓也，小野順貴，嵯峨山茂樹，“Orpheus 歌詞の韻律に基づいた自動作曲システム，” 情報処理学会音楽情報処理科学研究会研究報告, 2008-MUS-76, no.30, pp.179-184，2008.
10) 安藤大地，“対話型進化論的計算による作曲支援に関する研究，” 博士論文東京大学大学院新領域創成科学研究科，2009. あるいは，人工知能学会誌特集「人工知能分野における博士論文」, vol.25, no.1, 2010.

■2 群－9 編－2 章

2-9 演奏の表情付け

（執筆者：橋田光代）［2011 年 6 月 受領］

演奏表現を題材とする演奏表情付け（performance rendering）は，自動作編曲と並び生成系音楽情報処理研究の中核に位置する[1]．演奏表情付けシステムのさきがけとしての取組みは，1980 年代のFrydén[2], Clynes[3] らの研究にさかのぼる．1990 年以降には，GTTM[4, 5] やIRM[6] などの認知的音楽理論の利用，学習システムや事例ベース推論によるアプローチも見られるようになった．2002 年からは，演奏表情付けシステムによる生成演奏の聴き比べコンテストRencon[§§7] が開催されている．この頃から，FinaleやBand-in-a-boxなど商用の音楽制作ツールにおいても表情付け機能が導入・強化されるに至っている．

研究としての演奏表情付けは，人工知能研究の応用対象，つまり，演奏生成処理の自動化処理の実現としての興味から取り組まれてきたものが多かった．2005 年以降，CGM の普及とともに，一般ユーザが演奏デザインを実施する機会も増えてきている．自動化に加えて，ユーザの表現意図をより反映させることを目的とした研究の重要性も増してきている．

2-9-1 表情付けシステムの主な処理構成

図 2･13 に表情付けシステムの典型的な処理形態を示す．表情付けシステムの目的は，表情付けシステム入力された楽譜[***]に対して音楽構造解析処理と演奏表情付与の処理を行い，演奏データ（標準MIDI ファイル（SMF）や音響信号）を出力することにある．

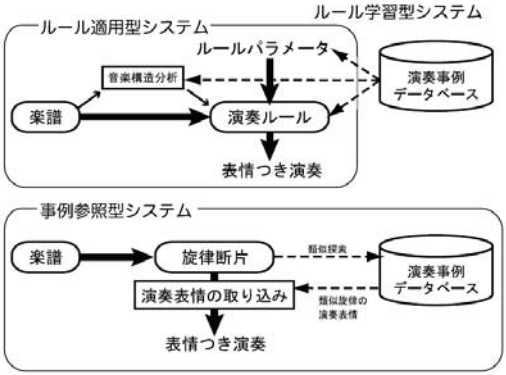

図 2・13　演奏表情付けシステムの処理の流れ

音楽構造解析処理では，演奏表情を付与する対象となる音楽構造（フレーズ，和声，拍節など）を取得する．その自動化は表情付け研究の最重要課題の一つであるが，音楽理論の曖昧性や，人間の音楽認知・聴覚心理なども考慮した総合的な解析能力が求められるため，シ

§§ http://www.renconmusic.org/
*** SMF や MusicXML 形式がよく採用されている．

ステムが人間と同じレベルで一意の音楽構造を同定する手法は確立されていない．現在研究されている演奏生成システムの多くは，手作業を介してあらかじめ入力楽譜に構造情報を付与（アノテート）したり，大量の演奏事例を用いて学習したりするなどで対処している．

演奏表情付与処理では，得られた楽曲構造をもとに，入力楽譜に対し，演奏表情として，各音符の音量・音長やテンポの揺らぎなどを付与する．処理手法については，システムによって，(1)ルール型，(2)事例型，(3)学習型に分類される．

2-9-2 ルール型

楽譜に対する演奏知識をルールとして組み込んでおき，入力楽譜に対して該当するルールを適用するシステムである．演奏知識とは，「スタッカートなら音符の時間長を短くする」「フレーズの最初の音符は音量を強くする」などのように，楽譜に記述される演奏記号やフレーズなどに対して慣習的・経験的に与えられる演奏表現を指す．生成された演奏の根拠が明確であり，人間にとって処理の中身が分かりやすいのが特徴である．

(1) MIS

演奏ルールの抽出に最も早く取り組んだのが片寄らの研究グループである．MIS では，動的計画法を用い，楽譜をガイドとして実演奏から表情データを抽出し，メロディの類型（繰り返しパターン）に対する共通の表現特徴をルールとして抽出した [8]．その後，重回帰分析のイタレーションによる演奏ルールの抽出 [9, 10]，コーパスに基づくフレーズ分析（GTTM の延長簡約の同定）などの研究 [11] に取り組み，学習型システム（2-9-4 節）へと発展させている．

(2) Director Musices

Sundberg らは，コードの変化，フレーズの変化点などに関連した演奏表現ルールを音楽家からの聴取（アナリシスバイシンセシス法）によって取得し，そのルールによって表情付けを行うシステムを示した [2]．この枠組みを利用して，Bresin らは 1990 年代初頭にニューラルネットワークを用いた Director Musices を提案した．その後，Director Musices の改良に取り組むとともに，ルールベースシステム，ニューラルネットによる演奏生成システムの比較 [12] を実施し，アーティキュレーション（レガートやスタカートなど個々の音符に対する奏法）にかかわる演奏ルール，特に，その感情表現に関するルールの検討を行った [13]．

(3) SuperConductor

Clynes は，拍，その上位の拍，小節などそれぞれのレベルにおける二つないしは三つのイベントのテンポや音量の比の組合せ（composer's pulse）をもとに演奏表現を行う機構を提案した [14]．その後，作曲家ごとのパルスを分析 [15] し，更にビブラートなどの表現を加え，様々なパターンを試聴できる SuperConductor と呼ばれるツールの開発を進め商品化を行っている．

(4) Pop-E / jPop-E / Mixtract

Pop-E[16] は，複数旋律音楽に対する自然な演奏の生成と，演奏デザインの効率的な支援に焦点をあてた演奏表現モデルである．声部別に演奏ルールを適用したうえで，楽曲構造に考

慮した声部間同期処理を行うことにより，声部間の微細なずれを保ちつつ演奏全体のテンポ統制を図っている．Pop-E の発展版である Mixtract[17] では，保科が提唱するフレーズ表現理論 [18] の定式化を行い，頂点音の導出と階層的フレーズ表現のデザインの支援を行っている．

(5) Finale

Finale[†††]に代表される記譜ソフトウェアは，印刷に耐え得る高品質の楽譜を作成するのが主目的である．開発当初から，作成した楽譜をMIDIで再生する機能はついているものの，表情付け機能としては楽譜上の記号に沿った単純なものに限られており，音符レベルでの細かな表情付けには不向きとされていた．しかし，2004 年以降に，ロマン派風やロック風，ジャズ風など，いくつかの演奏様式に沿ってMIDIの演奏パラメータを自動的に調整するHuman Playback機能が搭載され，独自のソフトウェア音源の提供も開始された．

2-9-3 事例参照型

対象楽曲のフレーズ構造に類似する演奏事例をデータベースから検索し，演奏表情の特徴量を転写するのが事例型の表情付けシステムである [19~22]．「自分が好きなあの曲の雰囲気に似た演奏表情を付けたい」といったように，演奏表現の中身（演奏ルール）が分からない，あるいは音楽に対する専門知識がなくても，事例演奏の演奏表情を直接転写することで表情付けができるというメリットがある．

(1) SaxEx

Arcos らによる SaxEx[19] は，演奏表情の自然さを表す形容詞（tender, aggressive, sad, joyful など）をユーザの制御子として，演奏事例の表情を対象楽曲に付与するシステムである．演奏表情の特徴を事例から適切に抽出するために事例ベース推論が用いられている．その際，演奏事例を構成する音符列を，GTTM や IRM 理論を利用して抽象化することにより，類似旋律を検索しやすくする工夫がなされている．

(2) Ha-Hi-Hun

平田らによる Ha-Hi-Hun[20] は，演繹オブジェクト指向データベース（DOOD）の枠組みを用いた音楽知識表現と，事例推論を採用したシステムである．表情付けにおけるユーザ意図を「自然言語で記述される指示」「音符列に直接的に付与される指示」に分けたうえで，音楽的に重要な音符の発音時刻，音長，音量の変化分に変換し，音楽的に重要な音の周囲の音にその変化分を伝搬させるという二段階の表情付けを行う．

(3) Kagurame Phase-II

清水らによる Kagurame Phase-II[21] は，鍵盤楽器による複旋律の楽曲を対象とした表情付けシステムである．対象曲や対象事例を様々な長さの旋律断片に分割し，旋律断片を対象として参考事例の検索を行う．旋律断片の分割により，事例の総数を増加させ，小数の演奏事例の効率的な利用を可能にしていることと，個々の検索の単位を短くし，類似旋律の検索に成

[†††] http://music.e-frontier.co.jp/product/finale/

功する可能性を向上させている.

2-9-4 学習型

学習型の表情付けシステムは，演奏ルールやルールに対する制御パラメータ，楽曲構造に対する演奏表情のパターンなどを，実演奏から学習することで生成処理の自動化を目指すものである．2000 年に入って以降，計算機が扱えるデータ容量も大幅に増大したことや音楽配信や音楽情報検索（Music information retrieval）に関する研究や事業が盛んになり，音響信号や演奏情報の特徴量を集積した大規模な音楽演奏データベースの整備も進んでいる [23, 24].

このような状況と重なって，近年の表情付けシステムは学習型・事例型システムが主流になりつつある [9, 10, 25～29].

Widmer らは，1990 年代から帰納学習法（IBL-SMART）と数値内挿法を組み合わせた手法によって演奏ルールの抽出を行い，そのルールを用いて演奏生成を行うシステムの開発に取り組んできた [25]. GTTM におけるタイムスパン木，IRM の一部を条件節として用意し，該当部分での音符の音量やテンポが平均値より大きいか小さいかなどの基準で正負事例を与えることで演奏ルールの抽出を行った後，制御量（平均値よりどれだけ大きくするか小さくするか）の数値的なフィッティングを行う．その後は，最近傍予測による類似フレーズの検索とその演奏表現適用の組みあわせ [30] やベイジアンネットワークなど機械学習によるフレージング生成モデル [28] に取り組んでいる.

2-9-5 インタラクティブ演奏支援システム

人間による演奏の支援や，人間との協奏（セッション）を目的としたリアルタイム系のシステムとしてインタラクティブ演奏支援システムがある．演奏生成の効率化をはかるためには，処理の自動化だけでなく，ユーザが演奏表現を練り込むプロセスや演奏を完成させるまでの作業時間を短縮する支援を行うことも重要である.

このような支援を目指す場合，そのシステムの実装形態は多くの場合リアルタイムに動作する演奏インタフェースになる．Radio-Baton[31] や orchestra in a box[32]，iFP[33] は，演奏の制御に必要なパラメータを音量とテンポの二つに集約しており，ユーザが自身の演奏表現を実施することを可能とする．Coloring-in Piano[34] は拍内の微妙な表現の制御を重視したシステムであり，ユーザは演奏時の音高ミスを気にせずに拍内の微妙な表現に集中できる.

2-9-6 技術的課題

ルール型システムの場合，本質的な演奏ルールの発見と制御パラメータの適切な集約が大きな課題である．演奏ルールを増やせば，より精緻な演奏表現を行うことができる一方で，ユーザが制御すべき制御パラメータの数が飛躍的に増大するという難点がある．しかし制御パラメータを抑制するために演奏ルールの数を減らすことは，ユーザにとって必要な意図を反映させられないというジレンマに陥る可能性もある．形容詞を用いて複数の制御パラメータをグルーピングすることは有効な手段の一つであるが，形容詞のニュアンスと演奏表現のニュアンスとの対応が適切かを新たに保証する必要が生ずる.

事例参照型システムの場合は，データスパースネス問題への対処が最大の課題となる．大量の演奏事例データベースを用意するのはもちろんのこと，対象楽曲や旋律に類似する事例

の探索手法についても工夫が必要となる．データスパースネスを回避する代表的な手法として，参照事例の演奏表情を統計的に処理することが挙げられる．この手法は，ロック風，ショパン風など楽曲の演奏スタイルを対象曲に適用する場合には有効である．しかし，事例型システムの場合は，参照する事例ならではの演奏表情が正しく抽出されなければ意味をなさない．システムを利用するユーザの目的にかなった演奏表情転写手法を構築することが重要である．

■参考文献

1) A. Kirke and E.R. Miranda, “A survey of computer systems for expressive music performance,” ACM Computing Surveys (CSUR), vol.42, no.1, 2009.
2) L. Frydén and J. Sundberg, “Performance rules for melodies. origin, functions, purposes,” Proc. of International Computer Music Conference (ICMC), pp.221-225, 1984.
3) M. Clynes, “Secrets of life in music,” Proc. of International Computer Music Conference (ICMC), pp.225-232, 1984.
4) F. Lerdahl and R. Jackendoff, “A Generative Theory of Tonal Music,” MIT Press, 1983.
5) 浜中，平田，東条，“音楽理論 gttm に基づくグルーピング構造獲得システム，”情報処理学会論文誌，vol.48, no.1, pp.284-299, 2007.
6) E. Narmour, “Beyond Schenkerism: The Need for Alternatives in Music Analysis,” The University of Chicago Press, 1977.
7) 平賀，平田，片寄，“蓮根：めざせ世界一のピアニスト，”情報処理，vol.43，no.2，pp.136-141, 2002.
8) H. Katayose and S. Inokuchi, “Kansei music system,” MIT Press, Computer Music Journal, vol.13, no.4, pp.72-77, 1990.
9) 石川，片寄，井口，“重回帰分析のイタレーションによる演奏ルールの抽出と解析，”情報処理学会論文誌，vol.43, no.2, pp.268-276, 2002.
10) 青野，片寄，井口，“重回帰分析を用いた演奏表現法の抽出，”情報処理学会論文誌，vol.38，no.7, pp.1473-1481, 1997.
11) H. Katayose, Y. Uwabu and O. Ishikawa, “A music interpretation system -schema acquisition and performance rule extraction,” Proc. of ICAD-Rencon: Performance Rendering Systems: Today and Tomorrow, pp.7-12, 2002.
12) R. Bresin, “Artificial neural networks based models for automatic performance of musical scores,” Journal of New Music Research, vol.27, no.3, pp.239-270, 1998.
13) R. Bresin and A. Friberg, “Rule-based emotional coloring of music performance,” Proc. Intl. Computer Music Conf., pp.364-367, 2000.
14) M. Clynes, “Secrets of Life in Music,” pp.3-15, no.17, Royal Swedish Academy of Music, 1984.
15) A. Friberg, V. Colombo, L. Frydén and J. Sundberg, “Microstructural musical linguistics: Composer’s pulses are liked best by the best musicians,” COGNITION, International Journal of Cognitive Science, vol.55, pp.269-310, 1995.
16) 橋田，長田，河原，片寄，“複数旋律音楽に対する演奏表情付けモデルの構築，”情報処理学会論文誌，vol.48, no.1, pp.248-257, 2007.
17) M. Hashida and H. Katayose, “Mixtract: an environment for designing musical phrase expression,” Proc. of Sound and Music Computing (SMC), 2010.
18) 保科，“生きた音楽表現へのアプローチ：エネルギー思考に基づく演奏解釈法，”音楽之友社，1998.
19) J. Arcos, R. de Mantaras and X. Serra, “Saxex: A case-based reasoning system for generating expressive musical performances,” Journal of New Music Research, vol.27, no.3, 1998.
20) K. Hirata and R. Hiraga, “Ha-hi-hun: Performance rendering system of high controllability,” Proc. of Rencon Workshop in Intl. Conf. on Auditory Display (ICAD), pp.40-46, 2002.
21) 清水，鈴木，野池，金子，徳永，杉山，“事例に基づく演奏表情生成システムにおける旋律断片自動生成アルゴリズムの改良と評価，”情処研報 2007-MUS-72, pp.7-12, 2007.
22) 伊藤，橘田，片寄，“複数の生成プロセスが制御可能な演奏生成システム「itopul」，”情処研報

2007-MUS-12, pp.45-50, 2007.
23) 後藤，橋口，西村，岡，“Rwc 研究用音楽データベース：研究目的で利用可能な著作権処理済み楽曲・楽器音データベース，”情報処理学会論文誌，vol.45, no.3, pp.728-738, 2004.
24) 橋田，松井，北原，片寄，“ピアノ名演奏の演奏表現情報と音楽構造情報を対象とした音楽演奏表情データベース，”情報処理学会論文誌, vol.50, no.3, pp.1090-1099, 2009.
25) G. Widmer, “Learning expressive performance: The structure-level approach,” Journal of New Music Research, vol.25, no.2, pp.179-205, 1996.
26) K. Teramura, H. Okuma, Y. Taniguchi, S. Makimoto and S. Maeda: “Gaussian process regression for rendering music performance,” International Conference on Music Perception and Cognition (ICMPC), pp.167-172, 2008.
27) S. Flossmann, M. Grachten and G. Widmer, “Expressive performance rendering: Introducing performance context,” In Proceedings of the 6th Sound and Music Computing Conference, 2009.
28) G.Widmer, S. Flossmann and M. Grachten, “Yqx plays chopin,” AI Magazine, vol.30, no.3, pp.35-48, 2009.
29) T. Kim, S. Fukayama, T. Nishimoto and S. Sagayama, “Performance rendering for polyphonic piano music with a combination of probabilistic models for melody and harmony,” Proc. of Sound and Music Computing (SMC), 2010.
30) G. Widmer and A. Tobudic, “Playing mozart by analogy: Learning phrase-level timing and dynamics strategies,” Proc. of Rencon Workshop in Intl. Conf. on Auditory Display (ICAD), pp. 28-35, 2002.
31) R. Boulanger and M. Mathews, “The 1997 mathews radio-batton and improvisation modes,” Proc. of International Computer Music Conference (ICMC), pp.395-398, 1997.
32) C. Raphael, “Orchestra in a box: A system for real-time musical accompaniment,” Workshop on methods for automatic music performance and their applications in a public rendering contest, Intl. Joint Conf. of Artificial Intelligent (IJCAI), 2003.
33) H. Katayose and K. Okudaira, “sfp/punin: A performance rendering interface using expression model,” Workshop on methods for automatic music performance and their applications in a public rendering contest, Intl. Joint Conf. of Artificial Intelligent (IJCAI), 2003.
34) 大島，西本，宮川，白崎，“音楽表情を担う要素と音高の分割入力による容易な midi シーケンスデータ作成システム，”情報処理学会論文誌, vol.44, no.7, pp.1778-1790, 2003.

■2群－9編－ 2章

2-10 自動伴奏

（執筆者：武田晴登）［2011年6月 受領］

2-10-1 はじめに

自動伴奏は，人間の演奏者が一緒に集まって演奏する合奏やセッションを，人間とコンピュータとが一緒にあたかも人間の演奏者どうしの場合と同じように演奏を行うための技術である．カラオケでは歌い手が決められたテンポで再生される伴奏に合わせて歌わなければならず，歌い手は自分の望むノリで思いどおりに歌うことはできない．これに対して自動伴奏が目指すのは伴奏を演奏するプロの演奏家たちが行うように演奏者の演奏意図を理解し，演奏者の演奏に合わせて演奏するシステムを実現することである．

人間が合奏中に，人の演奏を聴きつつ自分もそれに合わせつつ自分の意図した音楽を演奏で表現する．この枠組を計算機に行わせるためには，少なくとも演奏者の演奏を「聴く」ことにより演奏を理解することと，その結果自分が「演奏する」ことの二つの機能が必要である．このうち，「聴く」ことに相当する処理は，楽譜が演奏曲の楽譜が与えられている場合には演奏が楽譜のどの部分を演奏しているかをリアルタイムに推定する技術として楽譜追跡（score following）と呼ばれる．楽譜追跡の結果に基づいて伴奏を「演奏する」するためには，演奏者の演奏に合わせて演奏するためのスケジューリングが必要になる．以下では，この二つについて解説する．

2-10-2 楽譜追跡

演奏者の演奏がMIDI（musical instruments digital interface）信号で与えられる場合，例えば演奏者が電子ピアノなどのMIDI楽器を用いている場合の楽譜追跡を紹介する．MIDI信号により音高と打鍵速度の情報が送信されるので，演奏が単旋律で演奏誤り（ミスタッチ）がなければ，入力される音を順に楽譜の書かれている音と対応づければ演奏している箇所を知ることができる．しかし，実際に演奏には演奏誤り（ミスタッチ）が含まれるため，演奏誤りの可能性を考慮して楽譜の音との対応づけを求めなければならない．

Dannenberg は演奏と楽譜との一致した音の数をコスト関数に設定し DP（動的計画法，dynamic programming）を用いて単旋律の MIDI 演奏の楽譜追跡を行えることを 1984 年に発表した [1]．DP マッチングは二つのストリング（文字列）のマッチングを求めるためなどに用いられるアルゴリズムである．単旋律の楽譜追跡では音高が「ド，ミ，ソ」のようにシンボル列で与えられているので，ストリングマッチングと同様に DP を用いて演奏と楽譜との対応を求めることができ，したがって楽譜追跡を行うことができる．なお，通常の DP マッチングはすべての時系列の最後までマッチングの仮説を計算した後にバックトレースにより最適マッチングを求めるオフラインで使用されるアルゴリズム手法であるが，楽譜追跡では各時刻での最適のマッチングの仮説を使用するオンライン処理として使用する．

実用的な場面に用いるには，演奏誤りこれ以外に和音のように同時に演奏される音，トリルのように演奏される音の個数が楽譜からは定まらない装飾音を扱えることが好ましい．このためにこれまでに DP の原理に修正を加える手法が検討された [2, 3, 4]．更に，繰り返しのあるなしや演奏し直しや演奏を一部分スキップする場合など，楽譜どおりでない演奏順序で演

奏することも，楽器練習ではしばしば行われるので，ここにも対応できることが望ましい．装飾音や弾き直しやスキップなどの様々な演奏の可能性を確率モデルでモデル化して，HMM によってモデル化されたものを Viterbi 探索することによって楽譜追跡する手法も提案されている [5]．

実楽器の演奏をマイクで入力した場合の楽譜追跡についても研究が行われている．初期の研究ではピッチ推定と音高列のマッチングによる楽譜追跡手法を組み合わせることも試みられた [6]．しかし，現在でも単旋律で伴奏音が小さいなどの条件がない限りはピッチ推定や発音時刻をリアルタイムに高精度で得るのは難しく，ピッチ推定を用いたシステムは主に単旋律を対象に用いられている．また，ピッチのみに頼らずテンポの情報を用いて行う手法が研究されている．マイクから入力される音の発音時刻とその音の継続時間をマルコフモデルでモデル化し，マイク入力の音に対して HMM の状態推定を行い，楽譜のどの音符の発音に対応するかを求める手法が検討されている [7,8]．なお，これらの手法はオンラインで用いることを前提としているが，同じ楽譜との対応を求める技術で，オフラインで行うものはオーディオアライメントとして別項で紹介されている．

2-10-3 自動伴奏システムにおける伴奏の再生

演奏者の演奏に合わせて伴奏を演奏させる方法の一つに，演奏者が演奏した音が入力された瞬間に楽譜追跡を行い，そこで求めた楽譜位置に対応する伴奏音を再生させる方法がある．この方法では演奏者の音が入力されてから伴奏音を再生させるので原理的に伴奏は演奏者の演奏に遅れるが，その遅れが聴覚の時間分解能 [10] より小さいならば，伴奏音はテンポに演奏者に合わせて同時に演奏しているように聞こえる．例えば伴奏音のアタックが明確である楽音をシンセサイザで再生により実現できる．

通常の人間の伴奏者は，演奏者のテンポから次に演奏するフレーズの発音時刻を予測して演奏することから，同様の予測を計算機で行わせようとする研究がある．これは，オンラインで行うビートトラッキングで次のビートを予測する処理を行い，その結果を用いて伴奏を再生するととらえることができる．演奏されるべき次の音の発音時刻は，例えば直前の演奏テンポを用いて線形モデルで予測 [8] する手法が提案されている．また，演奏者に合わせるだけでなく伴奏側自体に人間のように揺らぎをもった演奏としての特徴をもたせるために，演奏者のテンポと事前に用意した伴奏側の演奏の逸脱情報の両方を用いて伴奏のタイミングを計算する手法 [12] も検討されている．ここで使用する演奏の逸脱情報については，演奏者のノリを表す発音時刻の変動を自動学習する手法 [15] として研究されたり，人手でアライメントをつけたデータベース [14] をつくる試みがなされている．

なお，この予測に基づいて伴奏を再生させるには，伴奏の再生テンポをリアルタイムに制御を技術が必要である．MIDI を用いる場合は MIDI 信号のタイミング制御のみで行えるが，音響信号を用いる場合は，音響信号では任意のタイミングでテンポを変化させるには音のピッチを変えずに音長を伸縮させるアルゴリズムが必要である．音響信号のタイムストレッチを phase vocoder[11] を用いて行う自動伴奏が報告されている [8]．

2-10-4 おわりに

ここで紹介した楽譜追跡や演奏再生の技術は，それを用いた自動伴奏システムでこれまで

報告されていて，対象とするものもオーボエのソロ演奏であったり，ピアノ初心者の学習支援[13]など様々である．また伴奏ではなく演奏そのものを自分の意図した演奏表情で演奏させるためのシステムとして，タッピングや手ぶりを使用する手法[12]や自動車の運転と同じインタフェースでハンドルなどを用いたテンポを制御手法[16]など，楽器以外のインタフェースを用いて演奏を体験できるシステムもつくられている．このほか，人間の演奏フレーズを学習してそれに呼応して機会が演奏を行うシステムContinuator[17]をはじめ，インタラクティブに音楽を生成するシステムは多く発表されている．今後，自動作曲や自動演奏の進展やソフトウェアやライブラリの拡充とともにこれからも，便利で楽しいシステムが創られると期待されている．

■参考文献

1) R.B. Dannenberg, “An On-line Algorithm for Real-Time Accompaniment,” in Proceedings of International Cumputer Music Conference (ICMC), pp.193-198, 1984.
2) J.J. Bloch and R.B. Dannenberg, “Real-Time Computer Accompaniment of Keyboard Performances,” In Proc. Int. Comp. Mus. Conf., pp.279-280, 1985.
3) R.B. Dannenberg and H. Mukaino, “New Techniques for Enhanced Quality of Computer Accompaniment,” In Proc. Int. Comp. Mus. Conf., pp.243-249, 1988.
4) Miller Puckette, “EXPLODE: A User Interface for Sequencing and Score Following,” In Proc. Int. Comp. Mus. Conf., pp.259-261, 1990.
5) 武田晴登，西本卓也，嵯峨山茂樹，“HMM を用いた MIDI 演奏の楽譜追跡と自動伴奏，”情報処理学会研究報告，2006-MUS-66，pp.109-116, 2006.
6) M. Puckette, “Score following using the sung voice,” in Proceedings of the International Computer Music Conference (ICMC), pp.175-178, 1995.
7) N. Orio, F. Dchelle, “Score Following Using Spectral Analysis and Hidden Markov Models,” in Proc. International Computer Music Conference (ICMC), pp.125-129, 2001.
8) C. Raphael, “Orchestral Musical Accompaniment from Synthesized Audio,” in Proceedings of International Computer Music Conference, 2003.
9) R.B. Dannenberg, N, Hu, “Polyphonic Audio Matching for Score Following and Intelligent Audio Editors,” in Proceedings of International Computer Music Conference(ICMC), pp. 27-33, 2003.
10) R.A. Rasch, “The perception of simultaneous notes such as in polyphonicmusic,” Acustica, vol.40, pp.21-33, 1978.
11) M. Dolson, “The Phase Vocoder: A Tutorial,” Computer Music Journal, vol.10, no.4, pp.14-27, 1986.
12) 片寄晴弘，奥平啓太，橋田光代，“演奏表情テンプレートを利用したピアノ演奏システム：sfp，”情報処理学会論文誌，vol. 44，no. 11，pp. 2728-2736，2003.
13) 大島千佳，西本一志，鈴木雅実，“家庭における子どもの練習意欲を高めるピアノ連弾支援システムの提案，”情報処理学会論文誌，vol.46, no.1, pp.157-171, 2005.
14) M. Hashida, T. Matsui, and H. Katayose, “A New Music Database Describing Deviation Information of Performance Expressions,” in Proc. Int. Conf. Music Info. Retrieval, pp.489-494, 2008.
15) M. Hamanaka, M. Goto, H. Asoh, N. Otsu, “A Learning-Based Jam Session System that Imitates a Player’s Personality Model,” in Proceedings of the 2003 International Joint Conference on Artificial Intelligence, pp.51-58, 2003.
16) E. Chew, A. Franc, ois, L. Jie and Y. Aaron, “ESP: A Driving Interface for Musical Expression Synthesis,” in Proc. Conf. on New Interfaces for Musical Expression, 2005.
17) F. Pachet, “The Continuator: Musical Interaction with Style,” In Int. Comp. Mus. Conf., pp.211-218, 2002.

■2 群－9 編－2 章

2-11 インタラクティブパフォーマンス・新世代楽器

（執筆者：竹川佳成）［2011 年 6 月 受領］

見たもの，聞いたもの，感じたもの，考えたことを誰かに伝えたいという思いは，人類の普遍的な要求であり，古来より，絵画，音楽，舞踊，演劇，彫刻，文学など様々な芸術領域において表現手段が模索され続けている．また，20 世紀は，シンセサイザを代表とする音楽音響処理技術をはじめ，映像処理技術，LSI 技術，通信技術，センシング技術などマルチメディアやインタラクションにかかわる技術が発展した．演ずることを基本とするパフォーミングアートにおいても多くの研究者や芸術家によりこれらのテクノロジを活用した作品が提案され，インタラクティブパフォーマンスという新たな芸術分野を形成するようになった．

本項では，特に音楽に焦点を当て，インタラクティブパフォーマンスの概要や，その関連技術について論じる．

2-11-1 インタラクティブパフォーマンス

インタラクティブパフォーマンスの代表事例を紹介し，次いで，その特徴や可能性について述べる．

(1) 代表事例

一般に，インタラクティブパフォーマンスは，パフォーマの演奏・演技・動作といったジェスチャをセンサで計測し，そのセンサ情報を計算機で処理し，音響・映像・照明など各種メディアをリアルタイムに制御することで，作品として成立させている．例えば，SensorBand[1] というグループは，筋電位センサ，赤外線センサ，接触センサなど各種センサを用いてジェスチャをセンシングし，そのセンサ情報に基づき音楽表現を行っている．T. Machover の作品「Brain Opera」[2] では，センサやデバイスを搭載した演奏ツールによるパフォーマンスや，インターネット中継でライブを鑑賞している多数の視聴者が音パターンを送り，ステージ上にいるパフォーマの演奏と共演している．三輪眞弘の作品「東の唄」[3] では，2 台のピアノを使用し，ピアニストとピアニストの演奏を解析しリアルタイムに自動作曲を行うコンピュータとのピアノ二重奏を披露した．また，これらの演奏を演奏中にサンプリングしプレイバックする試みも行われている．志村哲らによる作品「竹管の宇宙」[4]（**図 2･14**）では，尺八や演奏者にとりつけた各種センサによって，尺八のもつ繊細な音楽表現や演奏技法に関連する身体動作を取得し，音響や映像をリアルタイムに操作したり，センサ情報に基づいて演奏するコンピュータと尺八演奏者とが共演している．

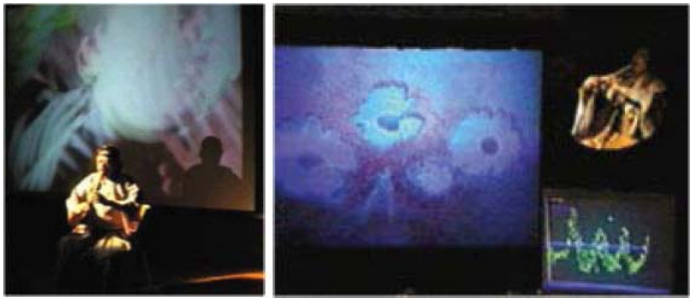

図 2・14　竹管の宇宙

(2)　特　徴

インタラクティブパフォーマンスは，従来のパフォーマによる楽器演奏だけでなく，センサを用いることで身体動作や生体信号など人が生成できるあらゆる情報伝達手段（ジェスチャ）を使ってメディアを制御できるという特徴をもつ．また，このジェスチャとメディア制御のマッピングは自由に行え，更に，ダンサーが踊りによって効果音を生成し照明をコントロールするといった複数メディアの同時制御も可能となる．

また，パフォーマの操作からシステムの応答までがリアルタイムに行われる点も特徴的である．このリアルタイムフィードバックゆえにパフォーマの微妙な運動制御，ひいては微妙な演奏を可能にしている．

(3)　芸術表現としての可能性

インタラクティブパフォーマンスにおいて，取得可能なジェスチャは無数に存在し，制御対象となるメディアも多種多様である．したがって，例えば，筋電位による演奏といったように，これまでにない制御手段を取り入れた新しい芸術表現が可能となった．また，複数メディアの同時制御によるシンクロニズム，歌で映像を制御するといったメディア変換も，従来の芸術表現ではなし得なかったものである．

また，インタラクティブパフォーマンスは，作品に対する鑑賞スタイルの拡充ももたらした．これまでの作曲家と演奏家によってつくり込まれた音楽を鑑賞する受動的な鑑賞スタイルと異なり，「Brain Opera」のように鑑賞者がパフォーマンスに参加する能動的な鑑賞を実現できるようになった．芸術表現としては，その場の雰囲気や状況によって演奏生成結果が大きく異なる一回性を作品に強く反映させられる．

2-11-2 インタラクティブパフォーマンスの関連技術

インタラクティブパフォーマンスの作品には，何らかのかたちで情報処理技術が使用されている．作品の中には，特殊なセンサの使用，限られた計算リソースの中でのリアルタイム処理，人間の演奏意図を理解する人工知能の開発など，高度な技術が求められる場合も少なくない．これら難題に対し，新世代楽器，自動伴奏システム（詳細は，2-10 自動伴奏を参照），

セッションシステム[‡‡‡]といった研究領域を立ち上げ，多くの研究者が挑んでいる．ここでは，特にインタラクティブパフォーマンスに最も関連のある新世代楽器について説明する．

また，インタラクティブパフォーマンスの作品制作においては，ハードウェア及びソフトウェア両方の広範な知識が求められる．このため，簡易に作品を制作でき表現の幅を広げるツールに対するニーズがあり，様々な制作ツールが提供されている．そこで，作品制作に有用なツールについてもいくつか紹介する．

(1) 新世代楽器

新しい電子楽器を作り出す試みは，製品としても研究としても多数行われている．奏法という観点で分類すると，「新しいセンサやデザインによるこれまでになかった楽器」と「従来の楽器を電子デバイスを用いて拡張した楽器」の二つに分類される．以下それぞれについて論じる．

(a) 新しいデザインの電子楽器

センサ技術の向上やシンセサイザの登場により，これまでの物理的な発音機構の制約に縛られない新しいデザインの楽器を開発できるようになった．これらは，モノにセンサを埋め込むタイプと人体にセンサを装着するタイプの 2 種類に大きく分けられる．

前者の事例は数多く，ここでは歴史的に重要な作品である RADIO DRUM, VIDEO HARP, LEMUR について紹介する．

M. Mathews らによって開発された RADIO DRUM[7] は，電磁波センサによって二つのドラムスティックの 3 次元位置座標を検出する楽器である．取得した位置情報は，発音タイミングなどを制御するトリガー信号やモジュレーションなどを制御するバリュー信号に変換され，打楽器のコントローラや指揮コントローラとして利用されている．RADIO DRUM の前身となるシーケンシャルドラムは，MIDI の制定以前に発表されており先駆性が注目される．

D. Rubine らによって開発された VIDEO HARP[8] は，光学スキャンセンサによって指の動作を検出し，設定モードのアルゴリズムに基づきセンサ情報を MIDI データに変換する機能をもつ．RADIO DRUM 同様先駆的な楽器で，その後の光学スキャン楽器に大きな影響を与えた．

Jazzmutant 社が開発した LEMUR[9] は，最大 10 点の個別認識が可能なマルチタッチスクリーンをもち，スクリーン内にスライダやボタン，マルチボールといったモジュールを自由に配置することができたり，各モジュールの色や大きさなどを自由に設定できる．これらのモジュールを組み合わせることで，シーケンサーやソフトウェアシンセサイザ，ソフトウェア楽器，VJ ソフトウェアなど様々なソフトウェアに適したコントローラを実現できる．

一方，後者は人が装着している各種センサから身体情報や生理情報を取得して音楽コントロールを行っている．

身体情報は，加速度センサ・ジャイロセンサ・ベンドセンサ・超音波センサ・磁気センサ・静電センサ・イメージセンサなどによって取得できる．一例として，肩・肘・手首に専用のベンドセンサを装着し身振りで演奏できる YAMAHA 社の MIBURI[10]，J. Pradiso らが開発し

[‡‡‡] セッションシステムとは，人間の演奏に協調して即興で演奏するシステムである．システムは，人間の演奏意図を認識し，それに基づいて音高列や各音の音量・調性など演奏データを生成する [6].

た靴底・つま先・足首などセンサを取り付けた 16 自由度のダンシングシューズ [11]，R. Aylward[12] らの画像処理を用いた動き検出による音楽コントロールなどが挙げられる．

また，生理情報は，筋電位センサ・脳波センサ・皮膚抵抗センサ・視線センサ・体温センサ・心拍センサ・呼気センサなどによって取得できる．生理情報を利用した一例として，脳波や筋電位を計測できる BioControl Systems 社の BioMuse[13] を使って演奏している事例 [14] がある．

更に，楽器の楽音生成及び発音にこだわった楽器もある．D. Trueman らによって開発された BoSSA[15] は，バイオリン風の楽器であるが，フィンガーボードと弓センサによって物理モデルに基づき音を生成し，楽器本体に搭載する 12 面体上のスピーカアレイから発音する．電子楽器の多くは，電子的な合成音をモニタスピーカやヘッドフォンから発音するが，発音体（弦やリードなど）の振動をより大きな物体（筐体）に伝えることで音量を増幅し空間上に音を拡散させるアコースティック楽器の発音モデルを取り入れている．

(b)　既存楽器の拡張

MIDI キーボード，MIDI ギター，ウィンドウ MIDI，MIDI ドラムなど楽器を電子化する試みは各楽器メーカから積極的に行われているが，いずれもアコースティック楽器と同等の演奏表現を行うことに注力している．既存楽器の拡張においては，電子化以外に新しい演奏表現を追求している事例や，既存楽器の問題点を情報処理技術により解決している事例もある．

新しい演奏表現を狙った事例としては，例えば，長嶋洋一の超琵琶 [16] は，琵琶に加速度センサなどを搭載することで，弦を弾いて胴体を左右に揺すれば楽音生成の高調波成分を制御するといった新たな奏法を付加している．また，宮下芳明らの Thermoscore[17] は，長時間持続してほしくない鍵や打鍵してほしくない鍵盤に対し，鍵盤上に並べられたペルチェ素子を加熱することで演奏に制約をかけられる．宮下芳明は，これを即興演奏における一種の制約として利用している．

また，演奏の難しさやセッティングの煩わしさの軽減をめざした事例もある．TransPerformance 社は，サーボ機構と制御アルゴリズムによって弦の張りを調節する自動チューニング機能をもつギター[18] を製品化している．また，YAMAHA 社が開発した MIDI ギターである EZ-EG[19] は各フレットに光スイッチを採用した電子ギターで，コードのナビゲートや半自動演奏を可能にしている．トランペットの発音の難しさを解消するために開発された EZ-TP[20] は，声の音程や音量によって楽音を生成できる電子トランペットである．

更に，持ち運びの軽減に取り組む事例もある．例えば，山野楽器のハンドロールピアノ [21] は，ゴム性の素材からなり，くるくる巻けるピアノを開発した．また，竹川佳成らの MobileClavierII（**図 2･15**）[22] は，すべての白鍵間に黒鍵を敷き詰めることでスムーズなキートランスポーズ操作を実現している．更に，寺田努らの DoublePad/Bass[23] は，普段持ち運ぶ情報機器を楽器に転用し，2 個の PDA を用いたベースギターを実現している．

加えて，鍵盤楽器を 1 オクターブの鍵盤に分割し LEGO ブロックのように組み合わせることで，楽器の形状を柔軟に変更できるユニット鍵盤 [24] も提案されている．

図 2・15　MobileClavierII

(2)　制作ツール

作品制作に有用なツールをハードウェア技術及びソフトウェア技術に分類し紹介する．ハードウェア技術に関しては，特にセンサの取り扱いが重要となる．これに対し，容易にセンサ情報を取得できる開発ツールが提供されている．InfusionSystem 社の I-CubeX[25] は，加速度センサ，超音波センサ，ジャイロセンサなど各種センサを提供し，それらの出力を RS 規格や MIDI 規格に変換するデバイスを販売している．また，Phidgets[26] は専用のセンサやアクチュエータを Java, C 言語, C++言語などプログラミング言語から制御できる．更に，Gainer[27] や Arduino[28] は，オリジナルのセンサやアクチュエータを接続できる専用デバイス及び Max/MSP，Flash，Processing など各種プログラミング言語から専用デバイスの動作を記述できる制御ライブラリを提供している．I-CubeX や Phidgets はいずれも各社が提供しているセンサやアクチュエータしか利用できない一方，Gainer や Arduino はセンサやアクチュエータを駆動させるための周辺回路を構築する必要があるものの，好みのセンサやアクチュエータを自由に使用できる．

Crossvow 社の Mote[29] は，センサネットワークやユビキタスコンピューティングの分野でよく使われているセンサノードで，ほかのノードを中継してセンサ情報を基地局（パソコン）まで伝送するマルチホップ機能，ノードの通信状態を考慮した最適な経路構築機能，ノードの参加・脱退に柔軟に対応できるアドホック機能など強力な無線通信機能をもつ．

任天堂社の Wii リモコン[30] は，加速度センサや振動モータなどを内蔵しており，Max/MSP などから制御できる．加速度センサはモーション検出に有用なデバイスの一つで，Wii リモコンは安価で手軽に使えるモーションセンサとして注目されている．また，Wii リモコンのもう一つの魅力として無線通信機能がある．無線化は，通信範囲，消費電力，信頼性，通信速度など課題も多いが，自由なパフォーマンスを獲得するための有用な技術である．上で紹介した InfusionSystem 社も無線通信モジュールを販売したり，ATR（国際電気通信基礎技術研究所）から小型無線加速度センサが発売されるなど，今後も注目すべき技術といえる．

一方，ソフトウェア処理としては，取得したセンサ情報を，音や映像など各種メディアの制御パラメータにマッピングし，制御する必要がある．多くのインタラクティブパフォーマンスのアーティストから支持されているソフトウェアとして Max/MSP/Jitter[31] がある．Max は，用意されたオブジェクトを結線することでアプリケーションを構築するプログラミングスタイルを採用しており，視覚的な要素を取り入れたプログラミングスタイル，MIDI に関連したサポートが人気の理由の一つである．MSP は音響処理，Jitter は映像処理に関する Max

のプログラミング環境を拡張するオブジェクト集である．また，MAX の設計者であった M. Puckette によってフリーで使える Pd（PureData）[32] と呼ばれる Max と似た開発環境も提案されている．

2-11-3 まとめ

本項では，国内外のインタラクティブパフォーマンス・新世代楽器の最新事例について紹介した．本テーマは，様々な分野の芸術家・研究者が精力的に取り組んでおり，本項で取り上げた事例以外に高い評価を受けている作品も多数存在する．より詳細な情報は，情報処理学会の音楽情報科学研究会の研究報告，国際会議 NIME（International Conference on New Interfaces for Musical Expression）[33]，ICMC（International Computer Music Cofference）[34] のプロシーディングを参照されたい．

世界初のシンセサイザが発明されてから約半世紀経過しようとしている．その間，メディア情報処理技術の進歩があり，様々なコンセプトの作品や楽器が考えられてきた．時代の評価を受けている作品は現段階では数少ないが，センサを搭載したコントローラの販売，いつでもどこでもマルチメディアコンテンツを楽しめる携帯機器の普及，ウェアラブル・ユビキタスコンピューティング環境の高まりなど，我々の生活や娯楽などあらゆる場面でメディア情報処理技術が使われ，もはやインタラクティビティの流れは抑えようがない．

今後，人々のインタラクションに対する興味は深まり，インタラクティブな製品やコンテンツに注目が集まるであろう．一方で，その質に対する要求も高まってくると考えられる．他の分野にさきがけてインタラクションを追求してきたインタラクティブパフォーマンスにおいては，なおさら期待と責任が大きいはずだ．黎明期から成長期に移るこの分野において，ぜひ，多くの芸術家や技術者に果敢に挑戦していただきたい．

■参考文献

1) “Sensorband のホームページ,” http://www.sensorband.com/
2) “Brain Opera のホームページ,” http://web.media.mit.edu/ joep/TTT.BO/index.html
3) 長嶋洋一，橋本周二，平賀　譲，平田圭二，“コンピュータと音楽の世界,” pp.437-442，共立出版，1998.
4) 安西祐一郎，片寄晴弘，中津良平，草原真知子，笹田剛史，黒川隆夫，“岩波講座マルチメディア情報学〈10〉自己の表現,” pp.99-102，岩波書店，2000.
5) 長嶋洋一，橋本周二，平賀　譲，平田圭二，“コンピュータと音楽の世界,” pp.429-432，共立出版，1998.
6) 長嶋洋一，橋本周二，平賀　譲，平田圭二，“コンピュータと音楽の世界,” pp.283-305，共立出版，1998.
7) B. Boie, M. Mathews and A. Schloss, “The Radio Drum as a Synthesizer Controller,” Proceeding of International Computer Music Conference (ICMC1989), pp.42-45, 1989.
8) D. Rubine and P. McAvinney, “The VideoHarp,” Proceeding of International Computer Music Conference (ICMC1988), pp.49-55, 1988.
9) “LEMUR のホームページ,” http://www.jazzmutant.com/lemur overview.php
10) “MIBURI のホームページ,” http://www.yamaha.co.jp/news/1996/96041001.html
11) J. Paradiso, K. Hsiao and E. Hu, “Interactive Music for Instrumented Dancing Shoes,” Proceeding of International Computer Music Conference (ICMC1999), pp.453-456, 1999.
12) R. Aylward, J. Paradiso, “Sensemble: A Wireless, Compact, Multi-User Sensor System for Interactive Dance,” Proceeding of International Conference on New Interfaces for Musical Expression (NIME06), pp.134-139, 2006.
13) “BioMuse のホームページ,” http://www.biocontrol.com/
14) Atau Tanaka, “Musical Issues in Using Interactive Instrument Technology with Application to the BioMuse,” Proceeding of International Computer Music Conference (ICMC1993).

15) D. Trueman and P. Cook, “BoSSA : The Deconstructed Violin Reconstructed,” Proceeding of International Computer Music Conference (ICMC1999), pp.232-239, 1999.
16) 長嶋洋一,“インタラクティブ・メディアアートのためのヒューマンインターフェース技術造形,”静岡文化芸術大学研究紀要，vol.1, pp.107-121, 2001.
17) 宮下芳明，西本一志,“演奏者の触発インタフェースとしての楽譜その拡張と可能性,”ヒューマンインタフェース学会論文誌，vol.7，no.2，pp.37-42，2005.
18) “TransPerformance 社のホームページ,” http://rbi.ims.ca/5702-598
19) “EZ-EG のホームページ,” http://www.yamaha.co.jp/ez/product/ez-eg/index.php
20) “EZ-TP のホームページ,” http://www.yamaha.co.jp/ez/product/ez-tp/detail/index.php
21) “ハンドロールピアノのホームページ,” http://www.yamano-music.co.jp/docs/hard/handroll piano61k2.html
22) 竹川佳成，寺田努，塚本昌彦，西尾章治郎,“追加黒鍵をもつ小型鍵盤楽器モバイルクラヴィーア II の設計と実装,”情報処理学会論文誌，vol.46，no.12，pp.3163-3174，2005.
23) 寺田努，塚本昌彦，西尾章治郎,“二つの PDA を用いた携帯型エレキベースの設計と実装,”情報処理学会論文誌，vol. 44，no.2，pp.266-275，2003.
24) 竹川佳成, 寺田努, 西尾章治郎,“UnitKeyboard: さまざまな演奏スタイルに適応可能な電子鍵盤楽器,”インタラクティブシステムとソフトウェア XIV: 日本ソフトウェア科学会 WISS2006，pp.89-94，2006.
25) “I-CubeX のホームページ,” http://infusionsystems.com/catalog/index.php
26) “Phidgets のホームページ,” http://www.phidgets.com/
27) “GAINER のホームページ,” http://gainer.cc/
28) “Arduino のホームページ,” http://www.arduino.cc/
29) “MOTE のホームページ,” http://www.xbow.com/
30) “Wii リモコンのホームページ,” http://www.nintendo.co.jp/wii/controllers/index.html
31) “Max/MSP/Jitter のホームページ,” http://www.cameo.co.jp/products/cycling74/maxmsp/max5.html
32) “Pure Data のホームページ,” http://puredata.info/
33) “NIME のホームページ,” http://www.nime.org/
34) “ICMC2008 のホームページ,” http://www.icmc2008.net/