

聴覚芸術への情報学的アプローチと 音楽情報処理ツールの開発事例

茂出木 敏雄*

Informatics Approach Towards Auditory Arts and Development Examples of Musical Informatics Tools

MODEGI Toshio

Abstract

We human beings have auditory organs called as a cochlear, which make frequency analysis on the musical scale. As we recognize any kinds of acoustic signals including human speech sounds with the same auditory organs, it seems not impossible to transcript them in a staff notation. Therefore, we have begun developing a MIDI encoding tool for general acoustic signals, which can transcript any kinds of given acoustic signals including human speech sounds in a staff notation.

After that, we have tried to develop a lossless compression technique for audio signals; an automatic endless background music composition technique by synthesizing selected rhythm, chord and melody phrase parts; an audio fingerprint technique for identifying music works; and an inaudible sign information embedding technique for music works.

In this paper, we will review these development examples of music informatics techniques done by us for past 15 years, and discuss future research topics.

Key Word

musical informatics, MIDI encoding

“Auto-F”, lossless encoding “G-encoder”, BGM synthesizer

“Matrix Music”, audio fingerprint, audio watermark

[要約]

ヒト聴覚には蝸牛という螺旋形をした音階理論的に周波数分析を行っている器官がある。ヒトの会話をはじめ、同じ聴覚器官を用いて音階理論的な認識を行っている以上、任意の音を五線譜で表現することは不可能ではない。そこで筆者らは、ヒトの音声を含む、あらゆる音を五線譜に載せることを可能にする、一般音響信号のMIDI符号化ツールの開発に着手した。

その後、音響信号に対して劣化を加えない情報圧縮技術、リズム・コード・メロディの音

* 情報表現学科非常勤講師。大日本印刷株式会社 情報コミュニケーション研究開発センター勤務。

楽部品の組み合わせで無限長のBGMを合成できる自動作曲技術、音楽作品を特定する音の指紋照合技術、音楽作品に耳に聞こえないサインを埋め込む技術の研究開発を進めてきた。

本稿では、これら過去15年間に行ってきた音楽情報分野の開発事例を振り返り、今後の研究課題を考察したい。

キーワード

音楽情報処理、MIDI符号化「オート符」、
ロスレス符号化「ゲンコード」、
BGM合成技術「マトリックス音楽」、音楽電子指紋、音楽電子透かし

はじめに

西洋音階はピタゴラスが黄金比を基に数学的アプローチで確立したものを基礎としているが、この音階理論の妥当性はヒト聴覚の解剖学的構造からも説明することができる。鼓膜の奥の内耳に蝸牛という螺旋形をした周波数分析器官があり、その形状特徴から、認識可能な周波数が離散的で半音未満の近傍の音成分には抑制がかかること（臨界周波数帯域幅といわれる）、高音ほど認識可能な隣接音との周波数差が大きくなること、オクターブ周期で音の類似性を認識すること、といった音階理論の特徴を説明できる。また、音波を含む波動には共振という特性があり、整数倍の周波数の倍音（一般には高調波という）を生み出しやすく、これにより和声学の協和音や和声進行理論をある程度説明できる。このような音階理論や和声理論の規則を遵守して制作された自由度の少ない音楽作品ほど、ヒトの聴覚器官が共鳴し、複数の演奏者どうしでコラボレーションしやすく、聴取者に感銘を与えやすく、長期に渡って経済効果を生みやすいという特性をもつのが聴覚芸術である。従って、音楽は芸術分野というより、数学者や情報学者にとってアプローチしやすい応用科学分野ともいえる。

1. あらゆる音を音符に変換できるMIDI符号化技術「オート符[®]」

前述した通りヒトの聴覚器官では音階理論的な認識を行っているが、その対象は楽音信号に限定されない。世の中はヒトの会話をはじめ、五線譜では表現できない音響信号に満ち溢れているが、それらも同じ聴覚器官を用いて音階理論的な認識を行っている以上、それらを五線譜で表現することは不可能ではない。そこで筆者らは、ヒトの音声を含む、あらゆる音を五線譜に載せることを可能にする、一般音響信号のMIDI符号化ツールの開発に着手した。

1.1.医療分野向け音響信号のMIDI符号化ツールの開発

本研究のきっかけは、筆者が旧郵政省・通信総合研究所（現NICT）の「情報通信基盤整備プログラム」に特別研究員（95年10月～98年3月）として出向した際に実施した遠隔医療分野のテーマに端を発する¹⁾。遠隔医療ではCTなどの画像伝送技術が主眼に研究されていたが、筆者らは波形伝送に注目した。当時、心電図波形の伝送技術は存在したが、心音波形については伝送はおろか、収集や蓄積も行われていなかった。心音・呼吸音というのは、現在も状況は変わらないが、聴診器で担当医師が聞くだけで記録に残らず、患者は無論のこと第三者の医師すらも聞くことができない。そこで、マイクロフォンを付けてデジタルのWAVデータとして収集できるようにすることから研究が始まった。続いて、収集した心音データをどのように遠隔伝送するかであるが、当時規格化されたMP3（MPEG-1 Layer3）圧縮をかけると診断のポイントとなる信号成分が劣化してしまうという問題があり、独自に符号化方式を開発した。そこで、着目したのが音楽音階で、心音や呼吸音波形を音符情報（MIDI: Musical Instrument Digital Interface）に自動変換するツールを試作した²⁾。図1は各種心音に対してMIDI符号化を行った事例で、正常な心音は 音と 音という2つの拍で構成されてい

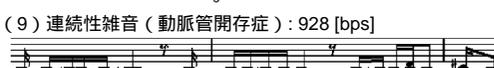
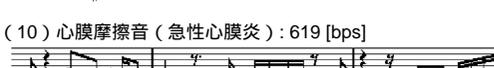
<p>[正常者心音] 181 [bps]</p> 	<p>[心雑音例]</p> <p>(6) 収縮期駆出性雑音（大動脈弁狭窄）: 289 [bps]</p>  <p>(7) 収縮期逆流性雑音（大動脈弁狭窄）: 536 [bps]</p>  <p>(8) 拡張期逆流性雑音（大動脈弁閉鎖不全）: 278 [bps]</p>  <p>(9) 連続性雑音（動脈管開存症）: 928 [bps]</p>  <p>(10) 心膜摩擦音（急性心膜炎）: 619 [bps]</p> 
<p>[異常心音例]</p> <p>(1) 音亢進（僧帽弁狭窄）: 232 [bps]</p>  <p>(2) 音分裂（心房中隔欠損症）: 622 [bps]</p>  <p>(3) 重合奔馬調律（拡張型心筋症）: 495 [bps]</p>  <p>(4) 四部調律（急性心筋梗塞）: 492 [bps]</p>  <p>(5) 収縮中期クリック（僧帽弁逸脱）: 318 [bps]</p> 	

図1 各種心音のMIDI符号化事例サンプル

1) 旧郵政省・通信総合研究所（現、（独）情報通信研究機構）からの「オート符」に関するニュースリリース <http://www2.nict.go.jp/pub/whatsnew/press/980317/980317.html>
 2) 茂出木敏雄・飯作俊一、「コンピュータミュージックによる医療診断の支援」、『コンピュータサイエンス誌 bit a, Vol.29, No.11、共立出版、pp.14-23、1997-11.

るが、同図(1)～(5)の異常心音には音や音などが付加される。また、同図(6)～(10)のように、音と音の収縮期あるいは音と音の拡張期に雑音が付加される病態がある。本変換を行うと情報量は約1/1000に圧縮されるが、診断のポイントとなる信号成分は劣化せず、むしろ強調され、97年の日本医療情報学会で発表し好評であった³⁾。

しかし、この段階では医学教育用途以外には実用性が乏しく、経済産業省より支援を受けながら⁴⁾、本技術により診断支援の機能をもたせるように改良を行った。具体的には内科学の教科書をもとにMIDI符号に変換された異常心音に関する知識ベースを構築し、変換されたMIDI符号に属性を与え構造化する。当時、HTML言語に対して拡張型マークアップ言語XML(eXtensible Mark-up Language)の導入が叫ばれた時期で、医療分野も電子カルテにXMLの導入が検討され始めたため、構造化されたデータをXML形式で符号化することにした。図1(3)の異常心音の原波形は図2(1)となり、これをMIDI符号に変換した結果が図2(2)である。これに対して属性付けを行った結果が、図2(3)である。音はM1(僧帽弁音)、T1(三尖弁音)、A1(大動脈弁音)の3つの弁開閉音成分から構成され、音はA2(大動脈弁音)、P2(肺動脈弁)の2つの弁開閉音成分から構成されていることがわかる。図2(3)をXML形式に符号化した結果が、図2(4)である。

また、MIDI符号化とは別のアプローチで、心電波形・呼吸音といった生体信号を用いてMIDI形式の音楽に変調を加える生体情報モニターへの応用研究にも着手した。例えば、心電図波形のP波を基にテンポを変化させ、呼吸音の振幅を基に再生音量を変化させることにより、生体信号の変動を音楽の変動を通じて遠隔モニターすることが可能になる⁵⁾。

1.2. 音楽の自動採譜ツールの開発:「音響情報の構造化記述手法に関する研究・開発」

次のステップとして、前述のMIDI符号化の対象音源を生体信号以外の音響信号一般に拡充することに取り組んだ⁶⁾。心音の次に取り組んだ対象音はヒトの声(話声、歌声)で、これが適切に符号化されるようになれば、あらゆる複雑な音源にも対応可能になると考えたからである。ヒトの声は子音と母音で構成され、母音の発音機構は管楽器と同じで、声帯という振動源(リード)が声道という管で共鳴し、声帯振動数の整数倍の倍音を出す機構までは同じである。しかし、楽器の倍音分布は固定的で高音になるにつれエネルギーが小さくなる傾向があるのに対し、母音の倍音分布にはところどころに特有のピーク成分があり、これらのピーク周波数(フォルマントとよばれる)は母音(ア・イ・ウ・エ・オ)ごとに変化する。そこで、フォルマント周波数に対応する音を複数の管楽器に割り当てて和音演奏すれば、理論

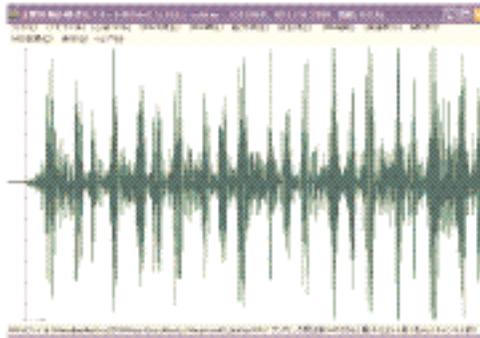
3) 茂出木敏雄・飯作俊一、「聴診音の楽譜による提示手法の提案」、『第17回医療情報学連合大会・論文集』、1-H-1-2、1997-11。http://www.jami.jp/17taikai/paper/084-009.htm

4) (財)デジタルコンテンツ協会 dCONSupport 「オート符」ソフトウェア・サンプル音源の無償配布サイト http://www.dcaj.org/d-con/frame09.html

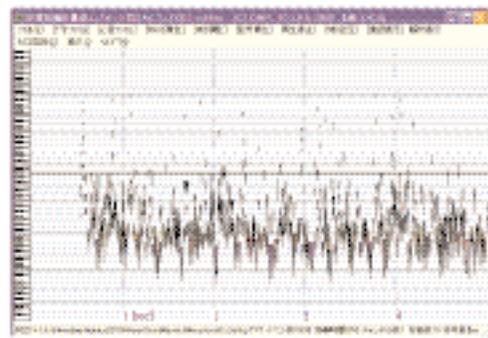
5) 茂出木敏雄、「監視業務向けオーディオ再生制御システムの提案」、『電気学会 電子情報システム部門誌』、Vol.120-C, No.2、オーム社、pp.285-291、2000-2。

6) Modegi, Toshio, 'Application of MIDI Technology for General Audio Signal Coding', *Information Systems and Technologies for Network Society*, World Scientific Publishing Co. Pte. Ltd., pp.163-168, 1997-10.

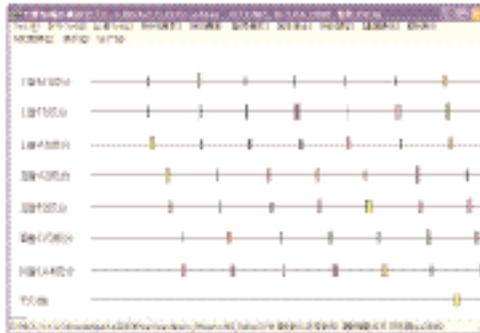
(1) 心音原波形（重合表示）



(2) 同・MIDI符号化変換データ



(3) 同・心音成分構造化データ



(4) 同・心音構造化XML形式変換データ

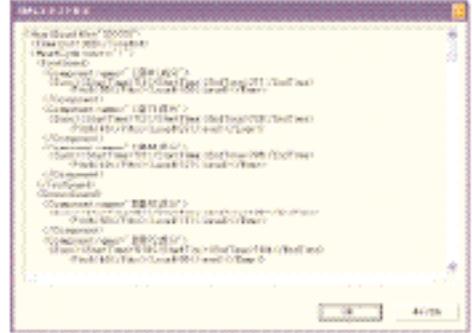


図2 心音のMIDI符号化および構造化処理例

上は管楽器で母音を再現できる。子音については倍音成分だけでなく、口の中で摩擦音などの雑音を発生させているので、それらの周波数成分を考慮して、母音より多くの楽器を割り当てて和音演奏する必要がある。

従って、対象音をヒトの声を中心とした一般音響信号に拡充するためには、周波数解析における和音解析精度がポイントになる。例えば、ソース音響信号を楽器演奏音に限定すると、自動採譜という研究分野になり、この分野ではライバルがそこそこ存在する。例えば、当時代表的なDAW（Digital Audio Workstation）ツールであるYAMAHA/XGworks等に添付されている鼻歌音符入力ツールでは、現在も状況は変わらないが、和音解析性能が貧弱で、CDに収録されているような自然な演奏音では歯が立たず、鼻歌・ハミングまたはそれに相当する不自然なテンポで演奏された単音の楽器音でないと適切に変換してくれないため、実用から程遠い状態であった。

従来の和音解析性能が悪い理由は、周波数解析手法としてオーソドックスなフーリエ変換法を用いているためである。フーリエ変換法では周波数次元が音楽音階と異なり線形であるため、音高の分解能が不均一で、和音の近傍の多くの擬似成分を誤って拾いやすくなる。これに対し、筆者は一般化調和解析法を採用した。このアルゴリズムは文献⁷⁾に譲るが、フー

7) Modegi, Toshio, 'Multi-track MIDI encoding algorithm based on GHA for synthesizing vocal sounds', *Journal of Acoustic Society of Japan*, Vol.20, No.4, pp.319-324, 1999-7.

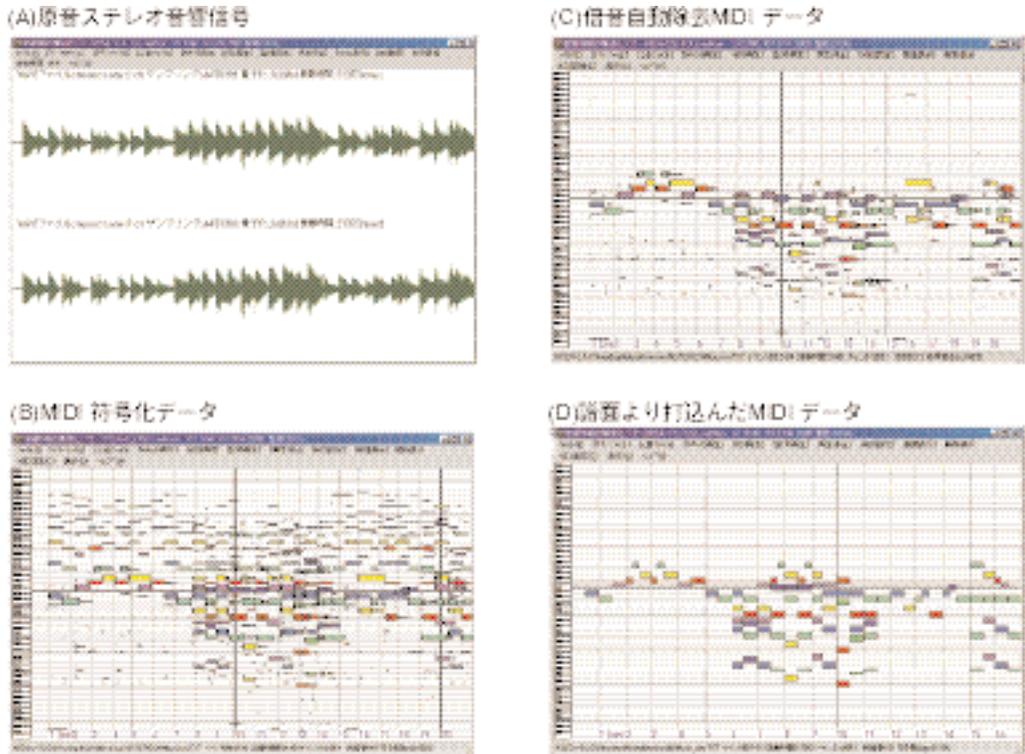


図3 ピアノソロ演奏のMIDI符号化事例（ムソルグスキー「展覧会の絵」）

リエ変換により得られた周波数スペクトルより単一の和音成分のみを抽出して、その成分を原音信号より削除して、再度フーリエ変換を実行して次の和音成分を抽出するという処理を抽出対象の和音数分だけ繰り返し実行する。

図3(A)はムソルグスキーの組曲「展覧会の絵」ピアノソロ演奏の音響信号で、これに対して本ツールでMIDI符号化を行い、自動倍音除去を行った結果が図3(B)(C)である。図中の着色された小さい四角形が音符を示し、横軸は時間で四角形の横幅は音価（デュレーション）を示し、縦軸は音高で四角形の縦幅は強度（ペロシティー）を示す。図3(D)は譜面をもとに同箇所を手入力したMIDIデータで、これと図3(C)を比較するとかなり譜面に近いデータが再現されていることが判る。

本ツール開発に関しては、平成12年度旧通商産業省・（財）マルチメディアコンテンツ振興協会の「コンテンツ制作基盤ツール等開発事業」⁴⁾に採択され、専門委員と討議を重ねな

8) 茂出木敏雄、「音響情報のMIDI符号化ツール「オート符®」の開発」、『芸術科学会誌DiVA』、No.2、夏目書房（株）pp.42-48、2001-12。

9) Modegi, Toshio, 'Very Low Bit-rate Audio Coding Technique Using MIDI Representation', *Proceedings of ACM 11-th NOSSDAV Workshop*, IEEE-press, pp.167-176, 2001-6.

10) Modegi, Toshio, 'Development of MIDI Encoder (Auto-F) for Creating MIDI Controllable General Audio Contents', *Entertainment Computing: Technologies and Applications*, Kluwer academic publishers, pp.229-236, 2003-1.

11) 茂出木敏雄、「音響信号の平均律音階に基づく汎用解析ツール（オート符®）の開発」、『電気学会 電子情報シス

がらツールの完成度を上げ、音楽制作者に配布できるフリーソフトに仕上げた（99年7月～00年3月）⁴⁾。また、本学・芸術情報学部・情報表現学科の渡辺秀先生と本ツールに関するプロモーション映像（およびサンプルコンテンツ）の制作を行い本成果発表・展示会等で公開した。ヒトの音声をMIDIで表現することに世界で初めて成功し、2チャンネル音源分離技術、XMLによる音響情報の構造化を併せて提唱した^{8)・11)}。配布サイト⁴⁾は2002年6月が最終版であるが、改良開発は現在も継続中である。

1.3. MIDI符号化ツール「オート符」を用いた楽器音による音声合成機能の実現

前述の通り、本ツールは、音声信号に適用すると解析されたフォルマント成分がMIDI形式に和音近似され、一般的なMIDI音源を用いてボーカルが再現できるという特徴をもつ。本ツールを無償配布してから、和音演奏可能な玩具楽器や、複数のアコースティック楽器等をアンサンブル演奏することにより、ボーカルが再現できないかという要望が多く寄せられるようになり、技術的な検討を進めてきた。まず、既開発のMIDI符号化ツールで生成されるMIDIデータは標準32和音で構成され、木目細かいデュレーションやベロシティなどの演奏指示情報が付与されているため、そのままでは判読性のある五線譜に変換することは困難であった。検討の結果、日本語音素については、デュレーションやベロシティを均一にした二連の単純な和音で表現する方法を検討し、カナテキスト入力によりMIDI音源や楽器演奏により音声を再現できるMIDIデータを合成するシステムの試作に着手した。また、MIDI音源による音声の再生音の明瞭性を向上させる方法を検討し、ソース音響信号に対して時間軸方向に2～4倍拡大して周波数解析を行うようにした。

図4は試作した楽器音を用いた日本語音声合成システムの構成を示す¹²⁾。図上段の水平方向のフローはテキスト音声合成のメインフローを示す。与えられたテキストを音節に分解し更に各音節を1～3個の音素に分解し、時系列に配置されている各音素をあらかじめ定義されている8和音などで構成されるMIDIデータに順次展開し、時間軸方向にMIDIデータを合成してゆく。その中で参照している日本語音素MIDIコード・データベースを構築するフローが同図左端の縦方向のフローである。あらかじめ日本語71音の音節別に録音を行い、前述したMIDI符号化ツールを用いて各音節WAVファイルを音節MIDIデータに変換する。次に、複数の音節MIDIデータどうしを掛け合わせて、各音節に共通に含まれるMIDI音素データを分離する処理を行い、20種の音素MIDIコードを作成する。例えば、「か」「き」「く」「け」「こ」という5つの音節データを掛け合わせて共通に含まれる「K」なる音素データを分離する。具体的な掛け合わせ方法は文献¹³⁾に譲る。

テム部門誌』、Vol.123-C, No.10、オーム社、pp.1768-1775、2003-10。

12) 情報処理学会・2010年インタラクティブ予稿集、「オート符」音声合成機能の動画紹介 [http://www.interaction-ip-sj.org/archives/paper2010/paper2010.html\(PA18, wmv\)](http://www.interaction-ip-sj.org/archives/paper2010/paper2010.html(PA18, wmv))

13) 茂出木敏雄、「MIDI符号化ツール（オート符）を用いた音素MIDIコードの設計と楽器音による音声合成機能の実現」、『電気学会 電子情報システム部門誌』、Vol.130-C, No.7、オーム社、pp.1159-1167、2010-7。

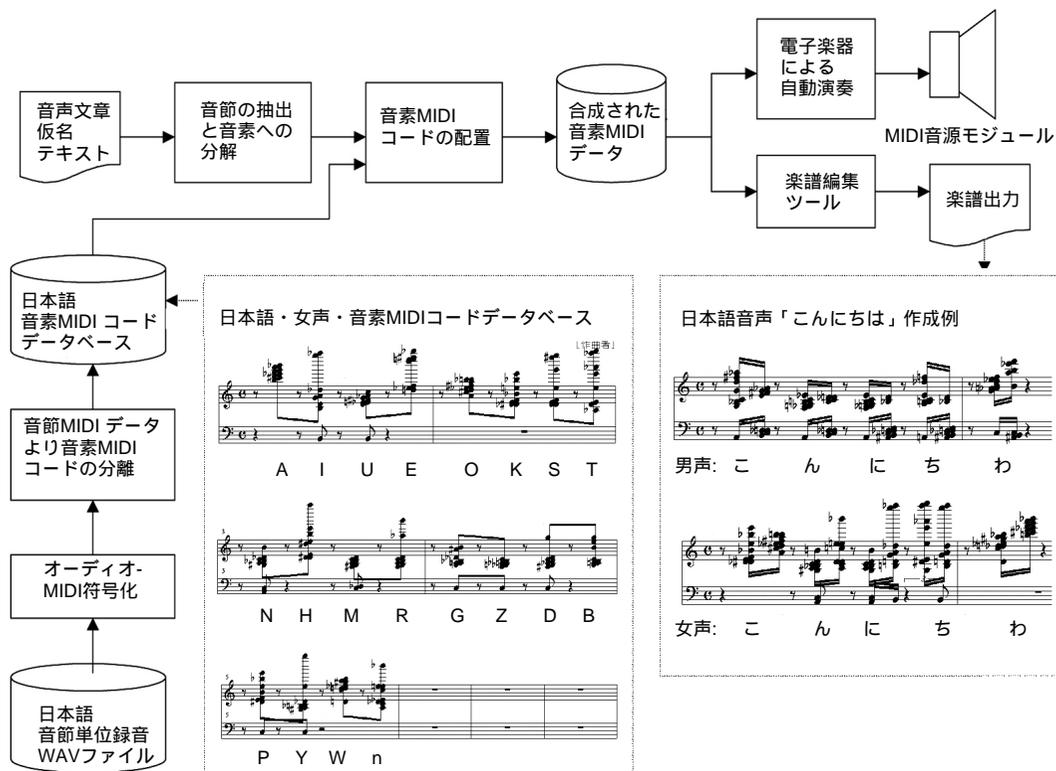


図4 試作した楽器音を用いたテキスト音声合成システムの構成

2. 音響信号に対して劣化を加えないロスレス情報圧縮技術「ゲンコーダ[®]」

コンピュータの記憶媒体が小型化・大容量化・低コスト化し、通信回線がブロードバンド化する一方で、従来より高精細で大容量の情報を容易に取得できるようになり、いつの時代においても、データ圧縮技術に対する需要は褪せない。品質上に殆ど差異がなければ、情報量が小さい方が迅速に処理ができ都合が良い。データ圧縮技術の基本は可逆圧縮（ロスレス符号化）であり、この場合は解凍処理を行うことによりオリジナルが完全に復元される。データ圧縮は特定の数値や文字の使用頻度が高いといったデータの偏りを利用しているため、完全にランダムに配列された数値データに対しては殆ど圧縮が働かない。

その典型例がオーディオデータや画像データで、マイクロフォンやカメラで入力された信号をデジタル化する過程において量子化雑音が発生し、サンプリングされたデータ配列がランダムになりがちなためである。そこで、オーディオデータや画像データに対しては、MPEGやJPEGといった非可逆圧縮（ロッキー符号化）方式が提案された。代表的なMP3オーディオ圧縮方式では、マスキングという近傍の周波数成分がヒトの聴覚では知覚されない特性を活用し、それらの成分を削除することにより、データに人為的な偏りを生じさせ1桁以下に情報量を削減することを可能にしている。しかし、復元されるオーディオデータはオリ

ジナルとは若干異なるため、業務用のオーディオ制作の過程では、このような圧縮技術を適用することは難しい。そのため、高精細オーディオの制作においては、大容量のワークデータが発生し、品質を維持するために多少の冗長性をともなっても、基本的に非圧縮で保存や伝送が行われている。作業用デジタル環境が大容量化してきているが、それらが追従できない勢いで制作現場への負荷が増大しつつある。そこで、音楽制作で発生するワークデータに照準を当てた新規なロスレス符号化ツールの開発につき、平成14年度経済産業省・(財)デジタルコンテンツ協会の「次世代デジタルコンテンツ制作支援事業」¹⁴⁾に提案し採択され、経済産業省の助成を受けながらロスレス圧縮ツール「ゲンコーダ[®]」の開発を実施した(02年7月～03年3月)。

前述した通りオーディオデータはそのままでは可逆圧縮の適用にならないが、隣接するサンプル間では相関があり、過去数サンプルで算出した線形予測値との誤差値の可逆な形式に変換すれば、データに偏りが生じ可逆圧縮がかかりやすくなることは既に知られていた。し

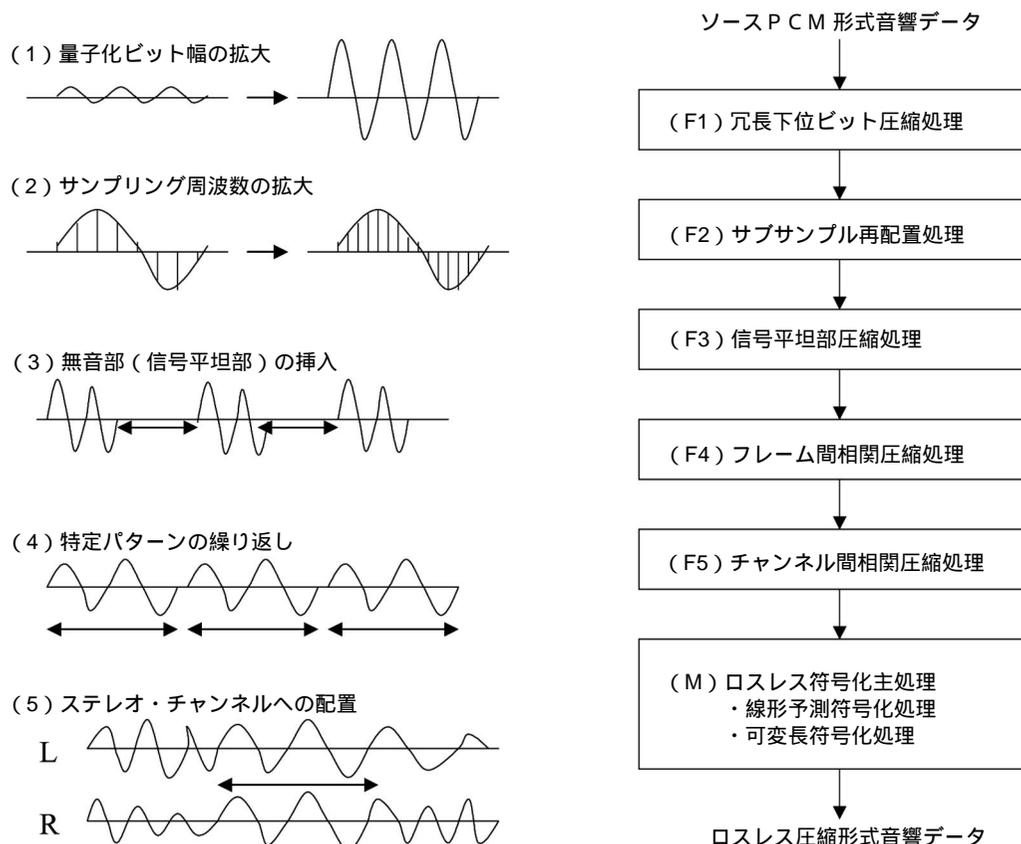


図5 音響制作途上で発生する各種冗長性と提案するロスレス符号化ツールの構成

14) (財) デジタルコンテンツ協会「平成14年度次世代デジタルコンテンツ制作基盤ツール等開発事業」成果報告集
<http://www.dcaj.org/bigbang/mmca/index.html>

かし、それだけでは現場が求める圧縮率を得ることは難しく、普及が難しかった。そこで、本研究は各種デジタル音響編集に伴う冗長度の増大を分析し、それらに適合した最適な可逆圧縮機能を多数提供するようにした。具体的には、本研究で対象とするPCM形式でサンプリングされた音響信号に対してデジタル編集されたデータを分析すると、図5に示されるような5通りの操作で冗長度が増大することがわかる。(1)量子化ビット幅の拡大、(2)サンプリング周波数の拡大、(3)無音部の挿入、(4)特定パターンの繰り返し、(5)ステレオチャンネルへの配置。(1)と(2)は、ミックスダウンを行う際に、ソースデータの品質を他の素材と合わせるために行われ、データ容量も増大する。(3)は他の素材とタイミングを合わせるため、(4)はリズムパターンを発生させるときに用いられる。(5)はステレオに変換する際に用いられる。これら(3)から(5)はデータ容量の増大は必ずしも伴わないが、無駄な領域が相対的に増大する。図5の(1)~(5)に対応した冗長性を除去するための専用圧縮ツールを図5の(F1)~(F5)に示すようにフロントエンドに配置し、既存の線形予測符号化と可変長符号化からなる(M)ロスレス符号化機能と組み合わせることにより、既存のオーディオ用可逆圧縮ツールより顕著な圧縮効果が得られることが確認できた。¹⁵⁾

表1に代表的な音源に対してロスレス圧縮を行った結果を示す。クラシック曲は50%前後まで圧縮できるが、ポップス・ボーカル曲は70%前後にとどまる。本研究を実施している時に、小澤征爾の「第九」CDがクラシック業界で異例のミリオンセラーに輝いたが、既存の代表的なカラヤンの「第九」CDに比べると優位に圧縮率が高いことを発見した。その後、ポップス曲でも調査してみると、宇多田ヒカルの"First Love"を筆頭にヒット曲はなぜか圧縮率が顕著に高いことに気づき、ヒット曲には線形予測が働きやすい“判りやすい音楽”という特徴をもっているのではという仮説が成り立ちそうなことを発見した。本ツールについては、その後、静止画(リモートセンシング画像、医療画像)、動画、CG、医療情報¹⁶⁾、ゲノム情

表1 代表的な音源におけるロスレス圧縮実験結果

素材内容	原信号サイズ [kbytes]	圧縮後サイズ [kbytes]	演奏時間 [min:sec]
(1)ベートーベン第九(カラヤン指揮)	695,126	350,783 (50.4%)	66:57
(2)ベートーベン第九(小澤征爾)	706,201	278,959 (39.5%)	68:10
(3)ピバルディ協奏曲「四季」	457,845	215,420 (47%)	42:51
(4)ムソルグスキー「展覧会の絵」pf版	354,401	125,611 (35%)	34:04
(5)「荒城の月」(滝廉太郎)	45,270	17,202 (37%)	4:19
(6)“Yesterday Once More” Carpenters	41,004	24,331 (59%)	3:50
(7)“First Love” 宇多田ヒカル	44,425	28,667 (64%)	4:17
(8)“Boys & Girls” 浜崎あゆみ	40,451	31,722 (78%)	3:54

15) 茂出木敏雄:「ロスレス符号化ツールの開発と評価用高精細音楽作品の制作」、『第50回音楽情報科学会・研究報告』, MUS-50, 情報処理学会, pp.7-12, 2003-3.

16) 茂出木敏雄、「医療情報の統合ロスレス符号化ツール(ゲンコーダ®)の開発」、『医療情報学』, Vol.24, No.1, 日本医療情報学会, pp.193-201, 2004-4.

報に展開し、ユニバーサル版も開発し、高精細音楽制作者などに配布を行った。また、本学・芸術情報学部・情報表現学科の渡辺秀講師と本ツールに関するプロモーション映像の制作を行い本成果発表・展示会等で公開した。

3. 楽曲部品の組み合わせによりBGMを合成できる自動作曲技術「マトリックス音楽」

発案者はフュージョンバンド「カシオペア」のキーボーディスト・向谷実氏（(株)音楽館代表取締役）で、同氏と共同で平成15年度経済産業省・(財)デジタルコンテンツ協会の「ブロードバンドコンテンツのブレイクスルー技術等開発支援事業」¹⁷⁾に提案したところ採択され、試作版を完成させた(03年10月～04年3月)。本ツールの基本アイデアは、音楽は少なくともメロディー、コード、リズムの3要素から構成されるが、複数の楽曲間で互いに要素を交換しても音楽的に整合性がとれるように素材を制作できると、それらの組み合わせで膨大な種類の音楽を生成可能になる。例えば、図6に示す試作ツールでは各楽曲が5パートで構成される5種類の楽曲をストックし、各パートごとに独立して5種類の楽曲より選択可能なように作曲しているため、 $5 \times 5 \times 5 \times 5 \times 5 = 3125$ 種の音楽を生成可能になる。一般の楽曲は各要素が更に複数のパートに分割されているため、パートの種類と楽曲数を増大させると、(パート数) × (楽曲数)のマトリックス状の基本素材を準備するだけで、天文学的な種類の音楽を生成可能になる。図6に示す今回試作したツールでは各楽曲が4分間の尺をもつため、毎回サイコロを振って組み合わせを変化させると、わずか25点の素材から連続2週間異なるBGMを再生可能になる。

また、図6の試作ツールの5×5の素材は、ジャンルの異なる5系統のアジアの民族楽器を主体にボーカルを加えて、全て96kHz/24bits/2-chでスタジオ演奏収録したもので、試作ツールで合成される音楽の再生品質は、BGMでしばしば使用されるMIDI打ち込みやループ素材で制作される音楽とは比肩できないものになる。各音楽素材を音楽CD品質に落とした状態ではCD2枚程度必要になるが、前節のロスレス圧縮ツールを適用すると平均25%に縮小でき、CD半分に全素材を劣化なしで収納できた。そして、ソフトウェアで選択された5トラック分の各圧縮音楽素材をリアルタイムに復号化しながら途切れなく合成再生することができた。したがって、通常の音楽CDでマトリックス音楽素材一式を流通させることが可能である。一方、ストックされる5×5の素材の殆どは、ミックスダウンされる前の単一楽器音であるため、違法流用されやすいという問題がある。そのため、各々の素材には可聴な電子透かしを埋め込み、著作権情報を語った音声メッセージが音楽と同程度のレベルで重畳するようにした。この電子透かしは、トラックどうして位相を反転させて、5本のトラックを合成再生させると消えるように設定している。例えば、5本のトラックに重畳する信号レベルを1:1:1:0.5:0.5とし、位相を+、-、+、-、- (+を位相角0度、-を180度)に設定すれば、5本のトラックを合成再生させると重畳される音声メッセージが打ち消される。しかし、各素

17)(財)デジタルコンテンツ協会「平成16年度ブロードバンドコンテンツのブレイクスルー技術等開発支援事業」
成果報告集 <http://www.dcaj.org/bbcontents/16/index.html>

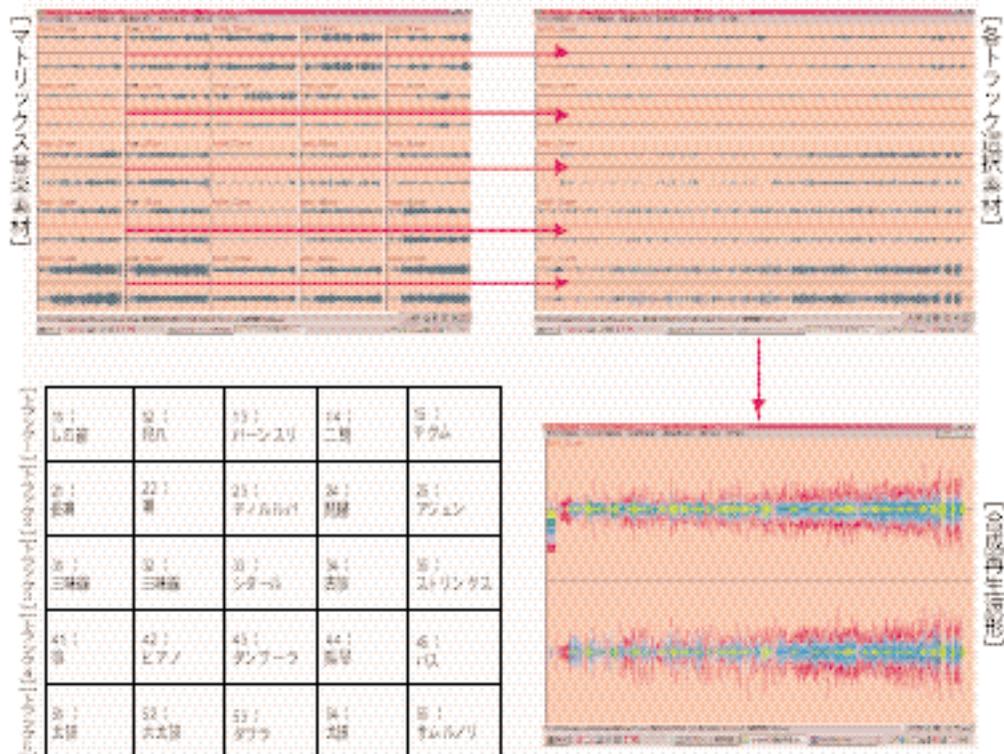


図6 試作した5×5のマトリックス音楽プレーヤの構成

材を個別に再生すれば、重畳された音声メッセージが聞こえ、これを信号処理で削除することは不可能であるため、違法に流用される心配はなくなる。

図6のマトリックス選別インターフェースとして、ユーザが直接ボタン操作で選別する方法では、好みの楽曲組合せパターンを探索することが難しい。そこで、3125通りの全組合せの合成音楽データに対して文献¹⁸⁾¹⁹⁾で示す方法を基本にした音響解析を行い、図7の音響マトリックスを構成する音量、ステレオ、音高、音符数、和声、倍音、テンポ、リズムからなる8項目の特徴量に関して平均値を記録した感性データベースを構築した。そして、ユーザが感性マトリックスの各音響特徴を大小で指定すれば、該当するマトリックスの組合せを検索できるようにした。更に、「元気が出る音楽」という感性キーワードや、Webブラウザで閲覧中の画像データを解析した画像特徴量に対して、感性マトリックスの音響特徴に変換する知識ベースを作成し、キーワード入力やWebブラウザ表示画面に連動して最適な再生楽曲を選定できるようにした。また、図7に示すように、映像に相応しいマトリックスの組み合わせを

18) Modegi, Toshio, 'Automatic Synthesis of Background Music Track Data by Analysis of Video Contents', *Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia*, Springer-Verlag Heidelberg, pp.591-598, 2004-11.

19) 茂出木敏雄、「映像コンテンツ解析によるBGMサウンドトラックの自動生成」、『電気学会 電子情報システム部門誌』a, Vol.125-C, No.7, オーム社, pp.1004-1010, 2005-7

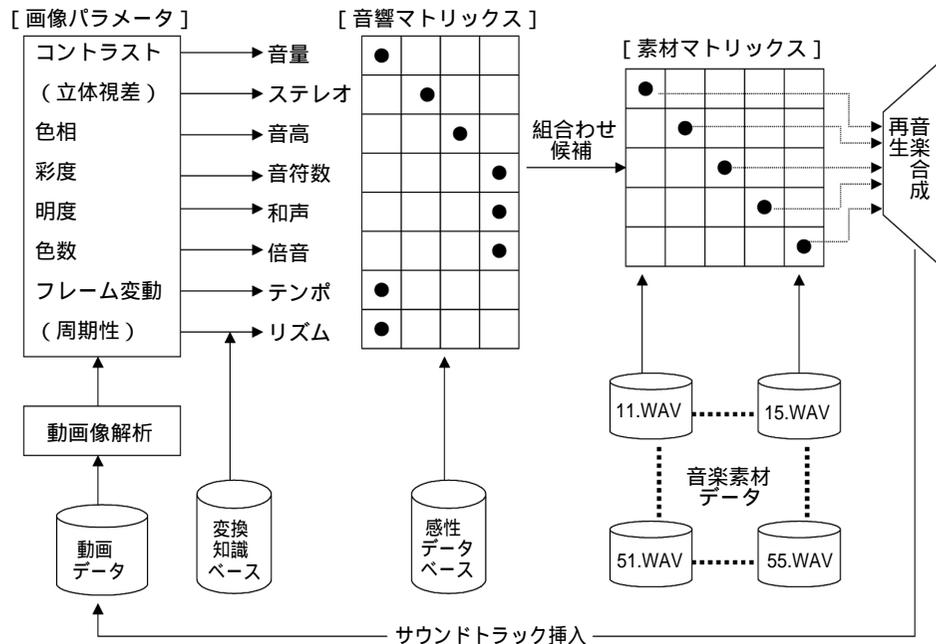


図7 マトリックス音楽プレーヤを用いた動画ファイルへのBGM自動付加システム

推薦し、自動的にBGMを付加する機能も追加試作した。

4. 音楽断片を用いて楽曲メタデータ等を検索可能な「音響フィンガープリント技術」

映像・音楽などのマルチメディア・コンテンツのデジタル信号処理技術の進展と、ネットワーク情報通信技術の進展に伴って、高品質な違法コピーを安価かつ容易に行えるという弊害をもたらしている。これに伴い、コンテンツに暗号化を施し容易に再生できないような工夫も行われているが、再生されている時のアナログ映像信号や音声信号をキャプチャすれば、複製は不可能ではなく万全ではない。これは、デジタルコンテンツの盲点でアナログホールと呼ばれ、その典型例が、映画館におけるビデオカメラによるスクリーン画面の盗撮である。民生用カメラもHDTV対応になった今日では、このような形態でも商品価値が十分ある複製品を作成できてしまう。そこで、手元にある映像・音楽信号の断片から映像・音楽のタイトルや制作者・著作権者（これらを映像・音楽コンテンツ本体に対してメタデータという）を探索し、違法コピーか否かを判定できるようにしたいという要求が出てきた。アプローチとしては映像・音楽コンテンツ本体にあらかじめ細工をする必要があるか否かで2方式があり、細工が不要な方法はフィンガープリント（電子指紋）、細工が必要な方法はウォーターマーク（電子透かし）である。本節では筆者らが提案する音楽のフィンガープリント手法について述べ、次節で音楽のウォーターマーク手法について述べる。

フィンガープリント手法とは、あらかじめ図8の楽曲素材1・楽曲素材2のような音楽信号

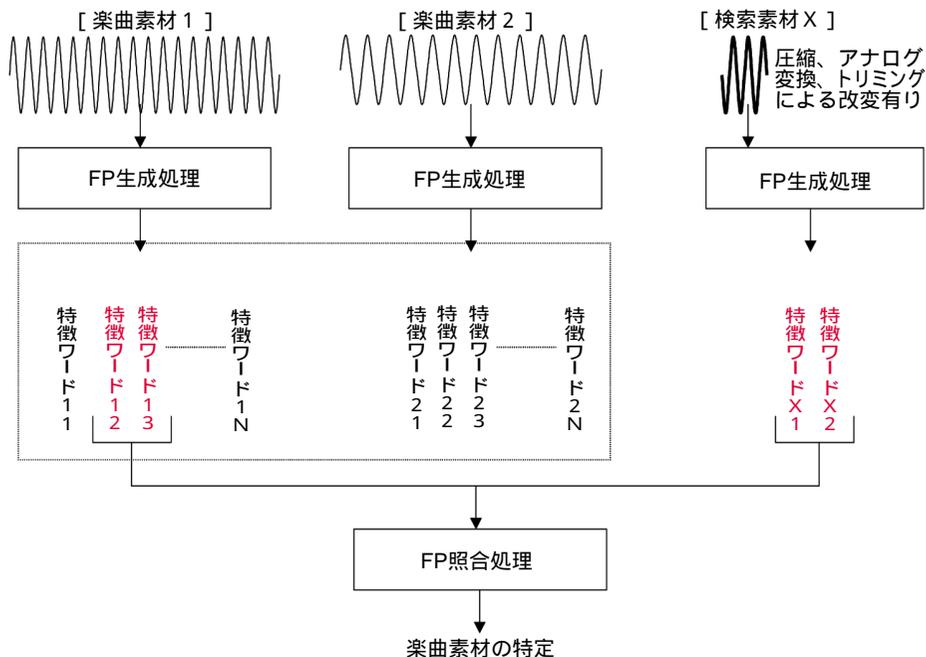


図8 音響フィンガープリントの基本原則

に関するデータベースを構築しておき、手元の信号断片である検索素材Xをデータベースと照合して検索素材Xがどの楽曲素材に属するかを特定する方法である。照合する際に、波形データどうして比較する方法をとると、処理負荷が大きく、波形のデータベース化により著作権侵害になり、検索素材Xに圧縮などの信号処理が施されると同定できなくなるといった問題が生じる。そこで、楽曲素材に対して楽曲素材を復元できない程度に圧縮した特徴パターンであるフィンガープリントで照合を行う方法が考案された。著名な先行事例として、米国Gracenote社があり、日本ではAu/KDDIがこれを用いた携帯電話による楽曲検索サービス「聴かせて検索」を実施している。この方法では、携帯電話のマイクロフォンでキャッチした品質の悪い音楽信号でも、フィンガープリントに変換すれば楽曲を特定できる。

現状では、フィンガープリント技術の詳細やフィンガープリントの生成や照合を行うためのソフトウェアライブラリ等は公開されていないため、自前で開発する必要がある。音響信号において、フィンガープリントとして使用する特徴パターンは一般的にスペクトルである。一定時間間隔ごとに平均的スペクトルを生成して照合する方法がとられる。図9は筆者らが提案するフィンガープリント生成アルゴリズムで、ソース音響信号44.1kHzに対して4096サンプル長のフレームで2048サンプル間隔でスペクトルを計算する。各スペクトルに対して、携帯電話で受信可能な電話音声帯域(300Hz~3.4kHz)を32バンドに分割し、直前フレームとのスペクトル差分値を10フレーム分にあわせて平均化し、2値化して32ビットの1ワードを生成する。同時に32バンドの総和値を音量データとして、10フレーム分にあわせて平均化

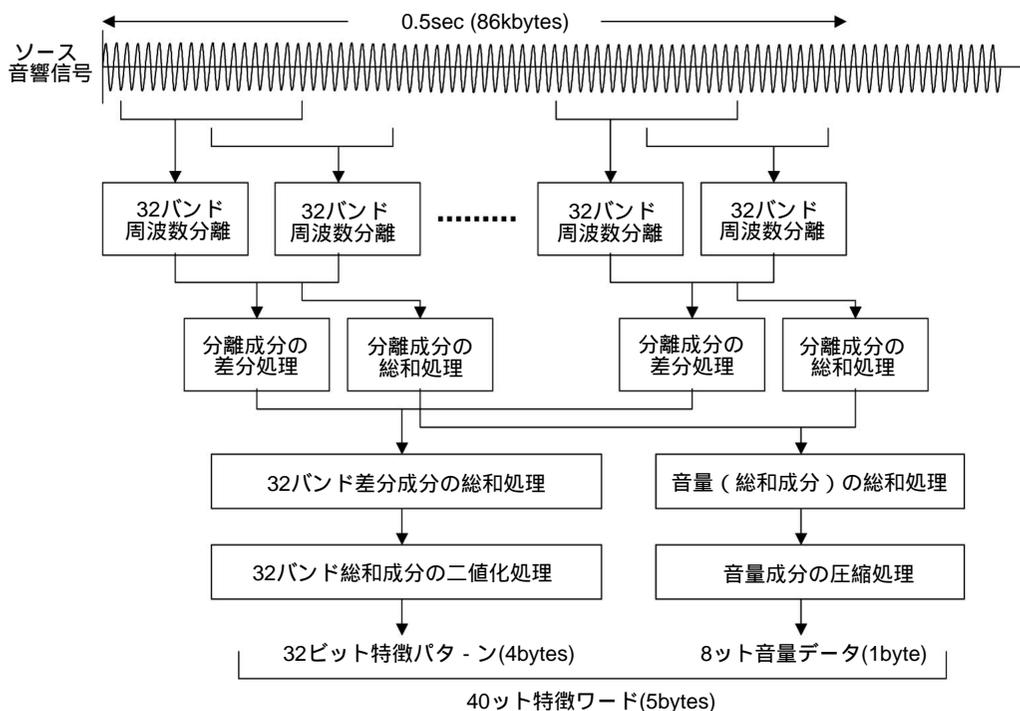


図9 音響フィンガープリント生成アルゴリズムの説

した音量データを8ビットで符号化し、先の32ビットのスペクトルワードとともに10フレーム（約0.5秒、ソースデータ容量86kbytes）あたり5バイトのフィンガープリントを生成する。

図8の検索素材Xに相当する音響信号に対しても図9と同様な方法でフィンガープリントを生成するが、時間軸上の位置ずれを考慮して、1フレーム～4フレーム分だけ位相を遅らせた4通りのフィンガープリントを同時に作成し、全5通りのフィンガープリントとデータベースに登録されている各楽曲別のフィンガープリントを照合する方法をとった。また、データベースに登録されている各楽曲フィンガープリントは原則としてオリジナルのフル尺分存在するため、検索側のフィンガープリントと位置をずらしながら照合を行う必要があり、照合負荷が大きい。そのため、フィンガープリントを時間軸方向に平均化し単一ワードで表現したメタ・フィンガープリントを定義して、位置をずらしながら詳細な照合を行う前に、対象楽曲とメタ・フィンガープリントにより大雑把な照合を行い、該当する可能性が無い楽曲を事前に排除する工夫も行っている。

5. 音楽作品に耳に聞こえないサインを埋め込む音楽電子透かし技術「ゲンコーダ[®] Mark」

前節で述べた手元にある音楽信号の断片から音楽のタイトルや制作者・著作権者等のメタデータを探索するもう一つの手法として、再生音に影響を与えないようにメタデータを音楽信号に埋め込む電子透かし手法が知られている。筆者らは本手法を音楽の著作権管理だけで

なく、ユビキタス音響空間^{20)・22)}への応用を提案している。これは空間に分布する音源の情報を携帯端末等で非接触に抽出し、音源に関連するサイバー空間と連動できるようにする構想で、具体的には、携帯電話を音源に近づけるだけで、音響信号に埋め込まれた情報を抽出できる新規な音楽電子透かし手法の開発を検討した。既提案の音楽著作権保護を目的とした電子透かし手法では、埋め込みノイズが聞こえないようにヒトの聴感特性が鈍感な音域を改変し、音質を余り劣化させない程度の各種圧縮・変調処理などに対する耐性を備えていれば良かった。これに対し、音源に埋め込まれた電子透かしの携帯電話で非接触に抽出する機能を実現するためには、アナログ空間を経由することに対する耐性と、携帯電話の音声入力信号処理系に対する耐性の双方を備えている必要がある。

現行の3G携帯電話では電話回線帯域と同様に4kHzを越える音は収録できず、更に録音された音声データは3GPP準拠圧縮により振幅側も顕著な歪みが発生する。更に、電子透かしが埋め込まれたソース音源は、アナログ/デジタル放送やストリーミングで配信される場合もあり、同様に4kHzを超える周波数成分は変調や圧縮による劣化が加わることも考慮する必要がある。従って、開発すべき電子透かし技術の要件として、4kHz未満の聴感特性が最も鋭敏な音域に、振幅歪みが加わっても抽出可能なように顕著な改変を加え、かつ再生時にノイズが聴取されないようにするという矛盾する課題を解決する必要がある。手始めに、筆者らは2チャンネル・ステレオ再生を応用して、データを埋め込んだL側スピーカで顕著に発生するノイズをR側スピーカで相殺する手法を提案した²³⁾。空間分割法、周波数分割法、時分割法の3つの基本方式を順次提案し、特に時分割法を用いれば、携帯電話による非接触抽出が可能であることを確認した^{22)・23)・24)}。しかし、これらのステレオ方式の提案手法はモノラル再生やヘッドフォン再生環境ではノイズが目立ち運用が難しいこと、受信端末を再生音源のL側スピーカから遠ざけたり、室内残響が豊かであるとR側信号の影響を受けやすく、抽出精度が低下するという問題があり実用化が困難であった。

そこで、聴覚心理学の分野で知られているヒト聴覚系の錯覚現象である音脈分凝^{おんみやくぶんぎょう²⁵⁾・26)}に着目した。これは音声の分野ではカクテルパーティ効果として知られている現象と同様な原理に基づくもので、ヒトの会話音声や音楽のメロディー・和声進行などの音素の流れ(音脈)が、途中で雑音等で遮られ分断されても、ヒトの脳聴覚野で補間し音脈を自然と再構築する現象である。図10はこれを電子透かし埋め込みに応用する手法を示したものである。横軸

20) 茂出木敏雄、「電子透かし技術の概要と今後の展望--携帯電話社会に到来した第3の波」、『マテリアルステージ』,No.7(2)、技術情報協会、pp.70-76、2007-5。

21) 茂出木敏雄、「音響空間のユビキタス化に向けた電子透かし埋め込み容量の拡大技術」、『電気学会 電子情報システム部門誌』, Vol.127-C, No.7、オーム社、pp.1013-1021、2007-7。

22) Modegi, Toshio, 'Construction of Ubiquitous Acoustic Spaces Using Audio Watermark Technology and Mobile Terminals', *IEEJ Transactions on Electrical and Electronic Engineering*, Vol.2, No.6, pp.608-619, Wiley, 2007-11.

23) 茂出木敏雄、「携帯電話で非接触抽出可能な音楽への電子透かし埋め込み技術の開発」、『電気学会 電子情報システム部門誌』, Vol.126-C, No.7、オーム社、pp.825-831、2006-7。

24) Modegi, Toshio, 'Increasing the Audio Watermark Data Rate in the Construction of Ubiquitous Acoustic Spaces', *Electrical Engineering in Japan*, Vol.165, No.1, pp.42-51, Wiley InterScience, 2008-10.

25) 茂出木敏雄、「音脈分凝を活用した携帯電話で抽出可能な音楽電子透かし技術 - ゲンコーダMark - 」、『電気学会

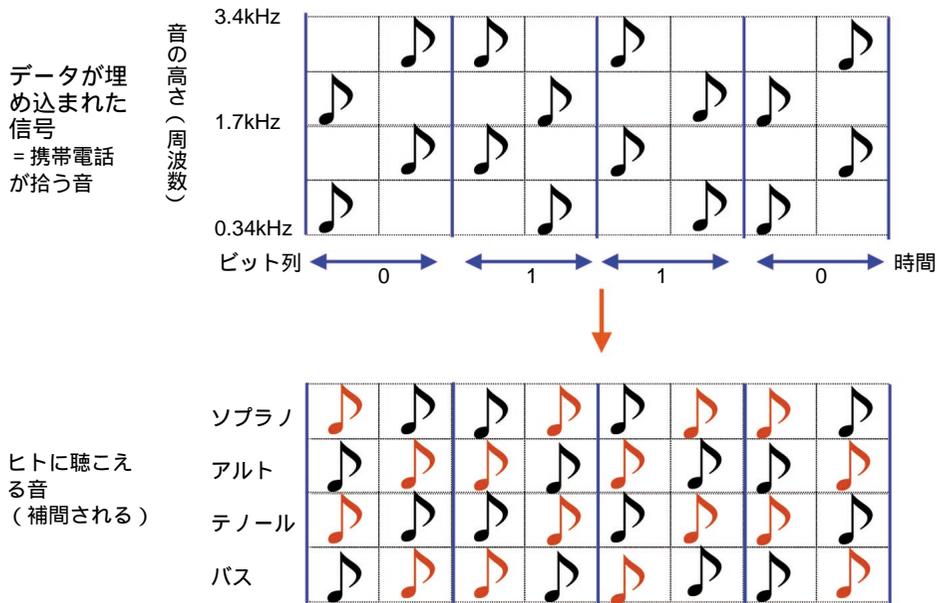


図10 音脈分凝を用いた音楽電子透かし埋め込み手法の概念

は時間で埋め込み対象の音楽信号を 50[msec]程度のフレーム間隔で分割し、各フレームに1ビットのデータを埋め込む様子を示している。図の縦軸は周波数で埋め込み対象の周波数を携帯電話で受信可能な電話回線帯域 (0.34 ~ 3.4kHz) とし、この帯域を4分割する。この4つの分割バンドは音楽でいうソプラノ・アルト・テノール・バスにはほぼ対応する。各フレームに対して時間軸方向に更に2分割し、周波数方向を含め8分割の信号成分に対し、図10上に表示するように市松模様のように信号成分を抜いてゆく。このとき、埋め込むビット値に基づいて、2通りの方向で信号成分を抜いてゆく。図10上に示されるパターンをヒトが聴取すると、図10下のように抜けた信号成分が補間されて聞こえるため、原音とあまり差異無く聞こえてしまう。しかし、マイクロフォンや携帯電話は図10上に示される市松模様のパターンをそのまま読み取るため、埋め込まれたビットパターンを抽出することが可能になる。このような方法で、視聴者には気づかれずに 20 [bps]のレートで音楽信号に情報を埋め込むことができることを確認した^{25) 26)}。

図10の方式で埋め込まれた電子透かしを抽出するソフトウェアは、パソコン以外に携帯ゲーム機やWindowsMobile搭載のPDA (Personal Digital Assistant) に実装し、博物館の展示ガイドへの応用を試みた²⁷⁾。音楽電子透かしでは空間的な位置情報とともに時間的な位置情報を携帯端末で空間伝送可能で、図11は空間的な位置情報を受信して静止画で構成される展示説

電子情報システム部門誌_α, Vol.128-C, No.7, オーム社, pp.1087-1095, 2008-7.

26) Modegi, Toshio, 'Audio Watermark Embedding Technique Applying Auditory Stream Segregation: G-encoder Mark, Extractable by Mobile Phone', *Electronics and Communications in Japan*, Vol.92, No.12, pp.1-12, Wiley InterScience, 2009-11.

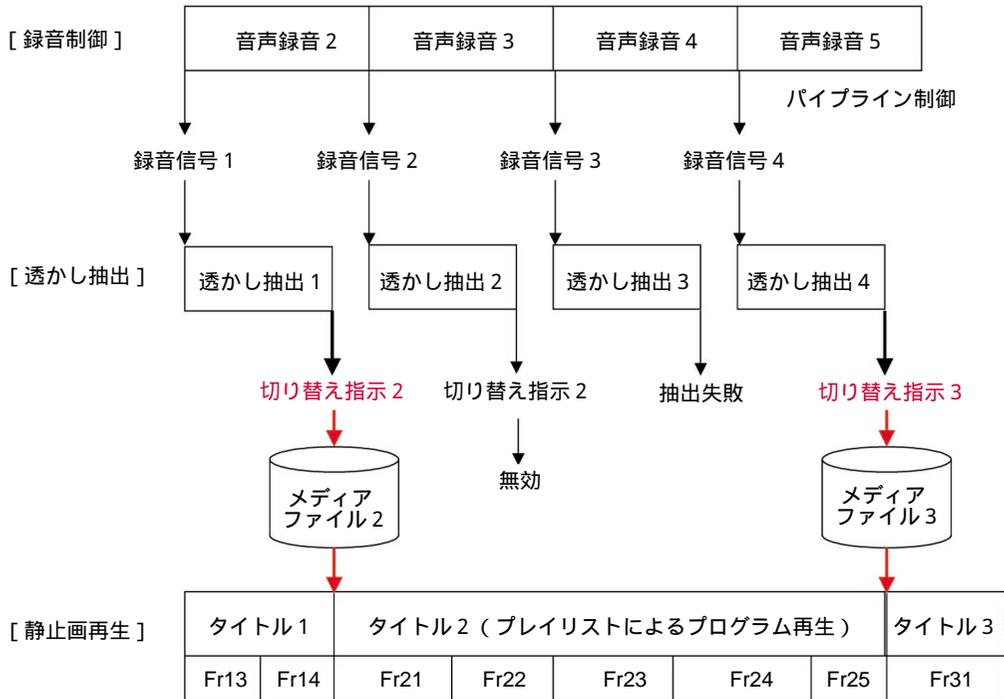


図11 博物館展示ガイド用のPDAソフトウェアの動作説明

明コンテンツを切り替えるPDAソフトウェアの動作を示している。博物館の主要スポットに音楽プレーヤーが設置されており、スポットのIDが電子透かしで埋め込まれたBGMが絶えず流れている。PDAは図11に示されるように連続的に3秒間隔で録音制御と透かし抽出が行われ、スポットのIDの受信に成功したら、図11下部に示されるようにスライドショーで流れているガイドコンテンツを切り替える。このような構成で、(株)音楽館(カシオペア・向谷実氏が社長)と共同で、2008年2月～4月に江戸東京博物館で開催された「川瀬巴水展」にて実証実験を実施した²⁸⁾。

図12は音楽電子透かしで再生中のコンテンツの時間的な位置情報を携帯端末に伝送する応用例を示している。例えば、シアターで流れている映像の音声トラックに電子透かしでタイムコードを埋め込んでおくと、携帯端末は随時タイムコードを受信し、タイムコードの受信に成功した時点でシアター映像と同期して外国語音声等を再生することができる。本手法の特徴は、電子透かし抽出に多少失敗しても再生開始時刻が遅れるだけで、シアター映像との同期再生に支障をきたさない点にある。

27) Modegi, Toshio, 'Spatial and Temporal Position Information Delivery to Mobile Terminals Using Audio Watermark Techniques', *Proceedings of Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IHHMSP2009*, pp.499-502, 2009-9.

28) ハイテック・ティービー「ゲンコーダMark」の動画紹介(2008年3月制作) <http://www.81tech.tv/product/70>

[主シアター・オーディオ再生設備]

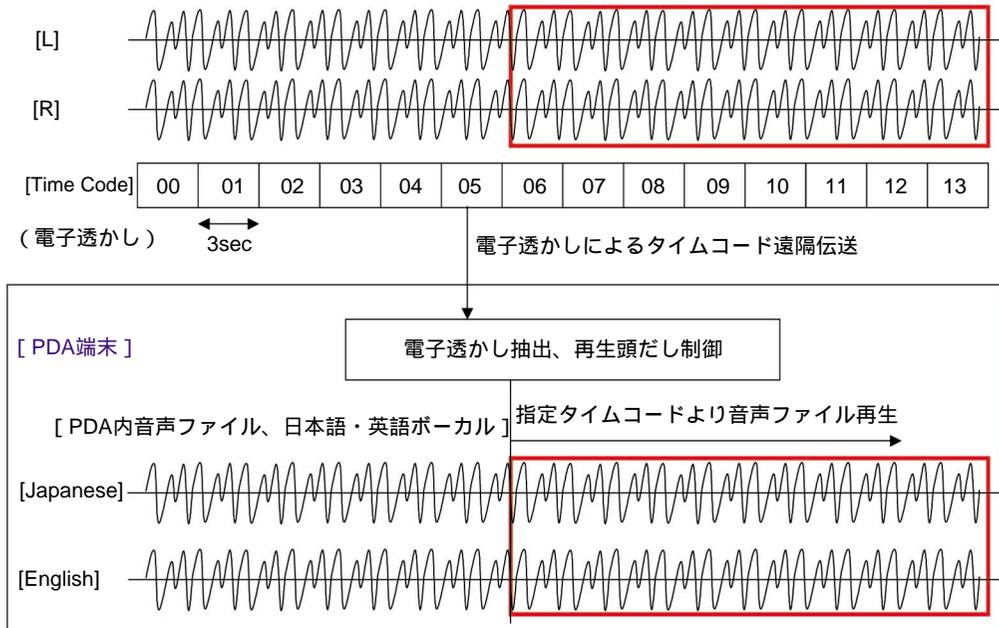


図12 シアター映像とPDA内コンテンツとの同期制御への応用例

図13はNTTドコモのFOMA3G携帯電話で電子透かし抽出を実現する構成を示す。スマートフォンの場合は、前述のPDAと同様に透かし抽出ソフトウェアを実装可能であるが、3G携帯電話の場合、Javaアプリ(ドコモのiアプリ)ベースでは録音制御が困難で、処理能力的にも透かし抽出ソフトウェアを実装することは困難である。そこで、携帯電話とサーバーを連携させて透かし抽出を実現するようにし、音楽に埋め込まれた透かし情報をもとに携帯電話を所定のサイトへ誘導させることができることを確認した。まず、携帯電話のJavaアプリ側では動画撮影カメラを起動し、3秒程度の撮影(実際は音声録音)を終了したら、3GPP形式で動画ファイルを保存して所定のサーバーへ送信する。サーバー側では受信した動画ファイルより音声データのみを復号化し、透かし抽出を行い、抽出した情報をもとにURLに変換し、待機中の携帯電話側のJavaアプリへ返送する。これにより、携帯電話側は所定のサイトへ接続できる。本システムを試作した結果、携帯電話からサーバーへ動画ファイルを伝送する時間がネックになることが判明し、ダミーの映像フレームを伝送する時間が予想外に冗長であった。そこで、動画ファイルより携帯電話側でAMR形式の音声データ部のみ抽出して送信するように改善し応答性が倍増した。

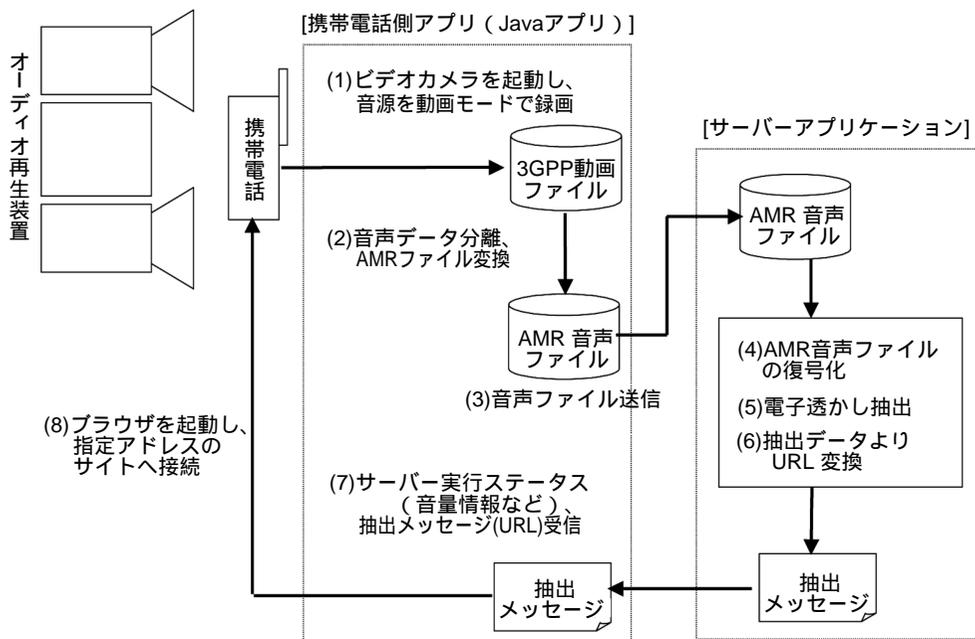


図13 3G携帯電話を用いた音楽電子透かし抽出システムの構築例

おわりに

本稿では、筆者が過去15年間にわたって推進してきた音楽情報分野における各種ツール開発について概観してきたが、丁度耳の鼓膜を破って内耳の蝸牛に到達し、永遠に出口が見えない螺旋階段に迷い込んでしまった段階といえる。さしあたり、蝸牛の入り口に戻って、聴覚構造の全体像を俯瞰し、新たな戦略を検討したいという意図もあって本稿を作成している。

まず、あらゆる音を音符に変換できるMIDI符号化技術「オート符[®]」に関しては、実用的な自動採譜ツールを実現することが当面の目標といえるが、複数の楽器音をパート分離する音源分離技術、生成されたMIDI符号から倍音などの非演奏音を分離する技術、生成されたMIDI符号から五線譜に変換する技術など多くの課題が残されており、これらは本研究に着手した段階で既に提示されており、結果的に殆ど進展が無かったと言わざるを得ない。また、楽器音でボーカルを再現する技術に関しては、具体的なアプリケーションを定めて、それに要求される明瞭性に近づけるよう向上させる必要がある。

次に、音響信号に対して劣化を加えないロスレス情報圧縮技術「ゲンコーダ[®]」については、情報理論上の圧縮率の限界に近づいていると考えられ、ロスレスでこれ以上の圧縮率は期待できない。むしろ、偶然に発見したロスレス圧縮率とヒット曲との相関など音響統計解析的な応用の方が期待できると思われる。

楽曲部品の組み合わせによりBGMを合成できる自動作曲技術「マトリックス音楽」につい

では、ループ合成ツールとは異なる新規な自動作曲・マッシュアップツールとしての期待がもてるが、今回の試作システムのように、楽曲部品が互いに音楽的なリズム進行や和声進行を崩さないようにヒトの手によって作成されたものである以上は実用に供さない。任意の既製楽曲を放り込めば自動的に音楽的に同調させるような技術開発が今後望まれる。

音楽断片を用いて楽曲メタデータ等を検索可能な「音響フィンガープリント技術」については、既に世の中では著作権管理業務などで実運用されており、技術的には完成に近づいており、検索速度についてもサーバーの並列化などで実用に耐えるレベルを実現可能といえる。最大の問題は著作権管理などの用途以外に産業上有望なアプリケーションが開拓されていないことにあると考えられる。

最後に、音楽作品に耳に聞こえないサインを埋め込む音楽電子透かし技術「ゲンコード Mark」については、「音響フィンガープリント技術」と同様に技術的には完成に近づいており、埋め込み容量の拡大、室内残響への対策と受信可能な距離拡大などの課題は残されているものの、これらはアプリケーションが特定されないと技術目標も定まらない。同様に、最大の問題は著作権管理などの用途以外に産業上有望なアプリケーションが開拓されていないことにあると考えられる。

今後の音響分野における取り組むべき研究テーマとして、筆者が現在興味をもっている分野は、音楽情報分野とほぼ対極に位置する騒音分野である。「騒音」については、これまで如何にして削減するかが技術目標であったが、今後は「騒音快音化」といって、「騒音」と心地よく共生するには、どのような技術開発が必要かといった新たなアプローチが求められ、これまで培った音楽情報分野の成果が一部活かすことができると考えている。

以上、本稿では筆者が過去15年間にわたって推進してきた音楽情報分野における各種ツール開発について紹介したが、筆者はほぼ同等の期間を本学でお世話になっており、本学の学生や教職員の皆様とのインタラクションがこれらの研究開発の推進に多大な影響を与えたものと考えており、改めて本学関係者の皆様に謝意を示す。また、本稿で紹介したツールについては、筆者が担当している「マルチフィールド体験演習」等の授業で既に活用しているが、教育・研究・その他にご活用されたい場合はソースコードを含め提供可能ですので、ご連絡ください。