

# 音楽からメディア統合へ

— 新しいメディア意味論のパラダイムを目指して —

メディア情報研究部 特別研究員

平田圭二

## 1 古い IT の枠組からの脱脚

IT の世界では、ビジネスや技術開発の進み方が大変速く目まぐるしいという意味でドッグイヤーという言葉が良く使われる。しかしそのような IT でも、基礎的な研究に関してはむしろ閉塞気味と言えるのではないだろうか。ムーアの法則に従う PC の速度向上、メモリ・ディスク等記憶媒体の性能向上、インターネットのブロードバンド化、モバイル・ユビキタス環境の浸透、AIBO や ASIMO に代表されるロボットの登場、Web 上に溢れ出る新しいビジネス等々。一見華やかな IT の世界ではあるが、では 5 年後、10 年後に必要なとされる技術に関して、我々は今どのような問題に取り組むべきなのだろうか。

計算機が誕生してほぼ 60 年が経過し、IT の分野で簡単に解ける問題は殆ど解かれてしまったとも言われている。新しい方向性は、単により高い数値、低い数値を目指すだけでは見えてこないであろう。米国大統領の科学技術に関する諮問機関である NSTC が発行する Networking and IT R&D という年次報告書 (一般に Bluebook と呼ばれている) からも読み取れるように、まず気づくのは、Internet/Web の発明・普及に伴い、従来の計算機という概念が拡張されている (あるいは崩壊している) ことである。つまり、計算機という概念は通信やコンテンツも含み、大規模・複雑であることが前提となっている。こうして、ここ 10 年ほどの間に IT という分野の全体像は変質してしまい、古い IT の枠組では現実の問題が解けなくなっているのである。特に、人が直に接するシステムを支えるソフトウェアに関しては、大きな技術革新が期待されていることを感じる。

解けない問題、今取り組むべき将来の課題はいろいろある筈だが、新しい IT を特徴付けるものの一つという意味で、古い IT の枠組には登場していなかったコミュニケーション、メディア、コンテンツを取り上げたいと思う (無論、これが NTT コミュニケーション科学基礎研究所の講演だからという理由もある)。

## 2 音楽を形式化できるのか

これまで音楽というものは一部の専門家だけが作曲、編曲、演奏などに従事でき、そうでない人々は単に聴いて楽しむことしかできなかった。しかし計算機による支援が得られるのであれば、一般のユーザでも音楽を自由に処理することが可能となろう<sup>1</sup>。実

際、計算機による支援が最も進んでいるのは自然言語の分野であるが、そこでは多くのユーザが様々な状況で様々なスタイルの文章を自由にやりとりし、コミュニケーションを楽しんでいる。音楽に関しても同様の状況が実現できれば、大変有意義なことであると思われる。

しかし、音楽という対象は、計算機上で形式化 (数学的に表現し操作すること) するには非常に手強い相手である。従来の IT の枠組では善し悪しの判断基準が精度、効率、構造、再現、予測、制御である。この判断基準の元で上手く取り扱うことができる対象のみが研究されてきたとも言える。一方、音楽の特徴として良く挙げられるのは主観、経験、暗黙、曖昧、個性、感性、情動、芸術性等である。IT とは水と油と言ってしまって良いくらいであり、とても科学の俎上に乗るとは思えない。実際、音楽で (知的な) コミュニケーションを行う試みや音楽コンテンツを創る試みの多くは、システムを適切に動作させるために、扱う音楽のジャンルや様式を限定したり様々なヒューリスティクスを導入せざるを得ないのが現状である。もし、そのような制限を避けようとするなら、できる操作は音楽を単にコピー (配信) することだけになってしまう。

筆者が音楽を選んだのは幸運であったように思う。なぜなら音楽には音楽理論があるからである。音楽理論の目的は、楽譜に書かれた楽曲を分析・理解し音楽的な構造を解明することである。音楽的な構造とは、楽譜上に見える音符や休符といった表層的な情報に対し、音符どうしがどのような結び付き方をしているのかという深層的な情報を意味する。音楽理論は、異なる音楽構造は聴取者に異なる認識を与え、その際聴取者に与える認識には、段落感、終止感、進行感、浮遊感、調性感等があるという前提に立ち、そのような認識を支配する重要な音やグループの同定方法 (知覚的ゲシュタルトに基づく楽曲の体制化) を種々提案している。また、音楽理論がこの体制化に関連させる認識が、段落感、終止感、進行感、浮遊感、調性感等という低次の感覚であり、感情、情動、感性という高次の感覚ではないという点にも留意しておきたい。

筆者は、この体制化と is\_a 関係 (抽象化/具体化関係あるいは包摂関係) の対応付けが音楽の形式化への足掛りだと考えている (これ以上は本稿のスコープを越えるので割愛させていただきます)。

<sup>1</sup>筆者が音楽情報処理の研究を始めたことは「研究補償説 (研究者は、その研究者自身の弱い能力を補償するような研究テーマ

を選ぶという定説。例えば、コミュニケーションの苦手な研究者はコミュニケーションを研究する)」で説明できる。

### 3 音楽ではないメディアの形式化

まずここでは、メディアをそれによってコミュニケーションを行うものと定義しよう。我々にとって代表的なメディアは自然言語(テキスト)である。その自然言語においては、まず語に関して signifiant/signifie があり、さらに統語論、意味論、語用論があり、自然言語の文章の意味を解釈する。そして、ほぼ当たり前の事柄として、現実世界の事物と語の意味との対応付け(grounding)を前提にして計算機システムを構築する。このような自然言語の形式的処理体系が構築されているが故に、自然言語処理は多大なる恩恵を我々にもたらしている。これより、他の画像、音、ジェスチャなどのメディアに対しても、同様の枠組の構築を試みたくなるのが人情である。だがそもそも、これらのメディアには signifiant/signifie があるかどうか怪しい。逆に、自然言語のような処理体系が規定できるメディアは他にも存在するののかという疑問すら湧いてくる。むしろ音楽のように signifiant/signifie (音と意味)が不分離(同型)であるようなメディアの方がより一般的なのではないだろうか。

例えば、現在 NTT で研究されている画像、思い出、視線は、コミュニケーションの手段という意味ではメディアである。これらのメディアは、自然言語のような形式的処理体系に従うよりも、筆者が音楽において行ったようなゲシュタルトに基づく体制化と抽象化/具体化関係との対応付けによって解釈することが適切ではないかと思える。例えば、体制化に基づいてより抽象的な思い出、より具体的な視線というものを上手く規定するのである。ここで体制化された思い出や視線に関連する認識は、終止感、進行感、浮遊感等に対応する何らかの低次の感覚になる筈である。

### 4 メディアを統合する方法

近年、テキスト処理の分野では、メタデータやアンテーションの支援による検索、翻訳、要約という汎用性の高いタスクに注目が集まっている。小説や論文の執筆や、HP の作成といった大きなタスクの構成要素になるという意味で検索、翻訳、要約タスクは中粒度であるとしよう。中粒度のタスクがインタフェースと自然に考えられるようになった理由は、中粒度までのタスクなら計算機がテキストの意味を比較的正確に解釈し実行できるからである。

ここにメディア統合のヒントが隠されている。というのは、画像、思い出、視線、音楽等テキスト以外のメディアでも体制化に対応付けられた抽象化/具体化関係を用いて検索、翻訳(スタイル変換あるいは言い換え(paraphrase))、要約のアルゴリズムを構成すれば、意味を正確に解釈したタスク実行が期待される。さらに、メディア固有な認識や処理を低減できることも期待される。

これまでのテキストを除くメディア処理の多くは、メディアごとにバラバラの表層的な特徴量を用いて、バラバラのインタフェースを提供していただけである。そして、この特徴量を用いる手法は一般

に合成に弱く、コンテンツ生成には不向きでもある。

従来のメディア統合、情報統合は、共通フォーマットでコンテンツデータを構造化することのみを指していた。筆者はこれに加えて、新しく中粒度の汎用タスク群のインタフェースを導入すべきであると考えている。そのタスク群には、検索、要約、翻訳(スタイル変換)を初めとして、類似、分類、模倣、冗長、誇張、混合、補間、予想等が含まれる(下図)。

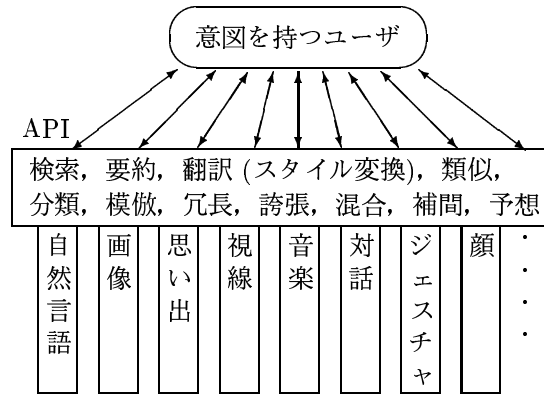


図: メディア統合プラットフォーム

新しい IT の枠組の元で今後我々が取り組むべき研究課題の 1 つは、中粒度の汎用タスクを自由に組み合わせられる環境やリソースを一般のユーザに提供するプラットフォーム(あるいはミドルウェア)を構築することである。このメディア統合プラットフォームは、専門家でない一般のユーザでも各自のセンスに従って大粒度のタスクをデザインし、様々なメディアを使い分けたコミュニケーションを可能にする。

例えば、道案内をする状況を考えてみよう。通常、道を聞かれた人は相手の様子を見ながら説明を試みる。言葉だけの説明では不十分な場合は身振り手振りを交えたり、指差しと頷きだけで十分の場合もあるだろうし、相手によって丁寧あるいはインフォーマルになり、時には分かり易い所まで相手を誘導する。道を聞く人にとっては、様々なメディアを通して伝えられた情報を統合して、相手の意図を再合成する問題になる。一方、道を聞かれた人にとっては、相手に伝えたい道順(つまり意図)を様々なメディアをどのように使い分けて効率良く正確に伝えるかという問題になる。このような状況に対応できるシステムを構築するには、メディア統合プラットフォームという枠組が必要ではないかと思う。

本稿で提案するメディア統合プラットフォームがコミュニケーション科学のための新しいメディア意味論への一歩ではないか考える。

平田 圭二 Keiji Hirata

メディア情報研究部 メディア表現研究グループ  
特別研究員

hirata@brl.ntt.co.jp