招待論文

# AV関連技術の動向と富士通の取組み

*Trends and Fujitsu's R&D approaches on audio-visual signal processing*

株式会社 富士通研究所　フェロー
松　田　喜　一　*Kiichi MATSUDA*

## 1 はじめに

富士通研究所は、長年に渡って、画像・音声・音響の処理技術に関して、基本となる方式開発からLSI、装置、システムまで、幅広く研究開発を行ってきている。どのメディアも、信号をデジタル化して処理を行うことで、多様な用途に向けて、必要とされる機能を実現している。

処理速度の高速化、メモリ集積度の向上、消費電力の低下といったデバイス技術の進歩に伴い、多機能・高性能なサービスを広く安く提供できるようになってきている。この5〜10年を振り返ってみるだけでも、高精細なデジタルテレビやDVDメディアの普及、手のひら静脈で個人認証する機能を持ったATMの登場、携帯電話の高機能化など、加速度的に技術の進歩と実用化が行われていることを身近に実感できる。

生活に密着したこれらの動きの裏では、30〜40年を超える過去からの研究開発活動、その成果としての技術の蓄積がある。以下では、動画像符号化、画像認識処理、音声・オーディオ符号化、音声合成・音声認識の各技術につき、富士通研究所での研究開発、および関連事業部と連携して市場に出してきた製品のいくつかについて概要を紹介する。

## 2 画像・映像処理技術の取り組み

三つの技術分野を取り上げて紹介する。

第一は動画像符号化で、情報量の膨大な動画像を放送したり、各種メディアに記録したり、という目的で広く実用化されている。

第二は画像の認識処理技術である。走行中の車のナンバー読み取り、利用者を特定するための静脈認証など、画像から用途に応じた有用な情報を抽出する技術であり、セキュリティビジネスにおける重要なコア技術となっている。

第三は、印刷情報の機密部分をマスク（暗号化）し、あらかじめパスワードを通知された人しか読むことができないようにして印刷物のセキュリティを高める、紙の暗号化技術である。

### 2.1 動画像符号化

動画像は情報量が膨大である。このため、実用的なサービスの実現には圧縮技術の導入が必須であり、**図1**に示すように、1970年頃から、そのときどきに利用可能な方式、実装技術を用いた製品提供に貢献してきた。

1980年代中頃までは、高速データ通信サービスの一環としてのテレビ会議との位置付けで、NTT主導で開発が進められ、それをベースに各社が独自仕様を加えて製品としていた。

その後は、ISDN（Integrated Services Digital Network）の普及を見込み、高速性（64k〜2Mbps）を生かすサービスとして、国際間のテレビ会議を実現するために方式の国際標準化活動がITU（International Telecommunication Union）で行われ、1990年にH.261が勧告化された。この標準化には、「動き適応ループ内フィルタ」という、再生画像品質の向上に効果のある技術を提案して採用された。また、この規格に準拠した製品FEDIS（Fujitsu Efficient Digital Image transmission System）-Tの開発に携わった。

次の主要な国際標準方式はMPEG（Motion Picture Experts Group）-2であり、DVDや地上デジタル放送に用いられている。この標準には「VBVバッファ制御」という技術を提案して採用された。受信（あるいは再生）側で入力データを蓄積するバッファがオーバーフローを起こさないために送信側が守るべき条件を規定するものである。本技術は、MPEG-2の実装に必須の特許を管理しているMPEG LAという組織に認定され、利用者からのライセンス収入を富士通にもたらすことに貢献できた。製品化では、膨大な演算量を必要とする「動き検出」を実行するLSIの開発、それを導入したFEDIS-M2の開発に寄与した。

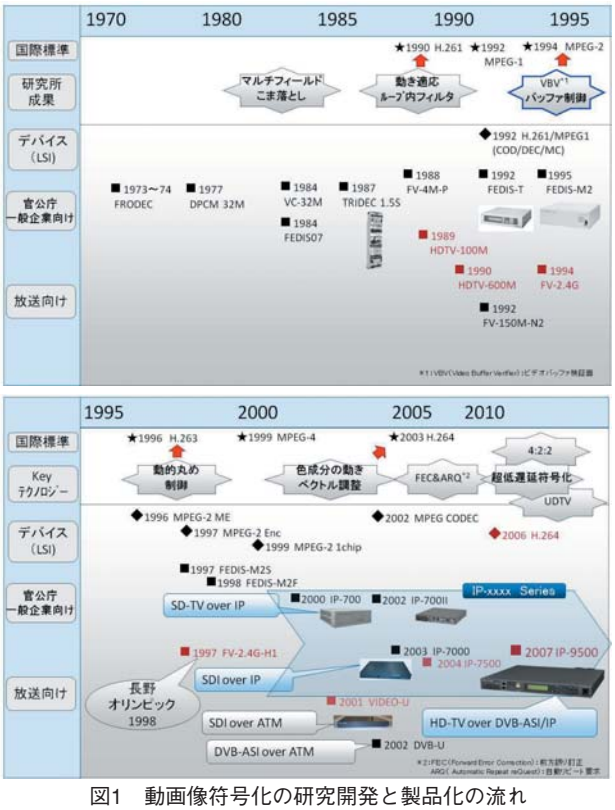


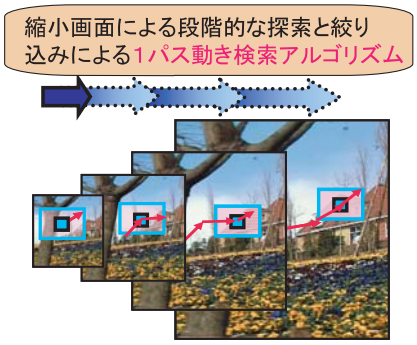

図1　動画像符号化の研究開発と製品化の流れ
Fig.1 R&D Activities and Products for Video CODEC

その後も継続してMPEG-4、H.264といった規格の標準化活動に参加し、自社技術の採用に繋げている。最近は、より精細度の高い映像形式を扱う方式の標準化活動である、MPEGのHEVC（High Efficiency Video Coding）、SMPTE（Society of Motion Picture and Television Engineers）といった場で活動を行っている。

最新の製品化への寄与として、H.264の実現に適用されている二つの技術を説明する。

1）H.264は、符号化性能の高さに伴って、必要な演算量が膨大である。LSI化するには、MPEG-2と比べると10倍以上の回路規模となるため、画質に影響を与えずに演算量を1/5程度に低減できるアルゴリズムを開発した。**図2（a）**に示すように、動き検出の処理において、元の大きさの画面を対象にする代わりに、縮小した画面でおおよその動きを求め、順次精細化して動きの精度を上げる方式である。

2）1Gbps近い情報量のハイビジョン映像が、放送では約20Mbpsまで圧縮されている。このように、1/50、あるいはそれ以下という大幅な情報量削減を画面全体に均一に適用すると、画質の低下が避けられなくなる。そこで、たとえば人の顔のような、視聴者が注目する部分には多めに情報量を割り当てて綺麗に、それ以外の部分は情報量を減らして全体での圧縮率を高める、という**図2（b）**に示す割り当て制御方法を開発した。




（a）小型・低消費電力化技術


（b）高画質化技術

図2　コストパフォーマンスの高いH.264実装

Fig.2 Implementation of H.264 Providing High Cost Performance

## 2.2 画像認識処理

画像認識は、映像や写真、絵の内容を自動的に認識する技術であり、屋内および屋外の監視・認証などの用途で広く利用されている。技術開発により、さらに高度な応用が広がってきている。画像検索やロボットなどへの応用も可能な技術である。

### （1）画像監視

富士通では、1999年には画像認識プラットフォームISHTAR（Image Sequence Hardware for Temporal Analysis in Real time）を空港における飛行機の入出港監視などの画像監視システムとして実用化した。このほかにも鉄道の監視、河川の監視などへの適用を進めてきた。道路を走行する車両のナンバープレートの文字を自動認識するシステムとして、2000年には車両の進行方向に対して道路わきに設置したカメラから撮影するナンバープレート認識技術を業界で初めて実用化した。この技術は、公安などの用途から、駐車管理、店舗などへの進入管理など民需への展開も進められている。2002年には車両後方からの撮影、2004年には2輪車への対応など、業界の先駆的な技術を実現してきた。2003年には降雪により視界不良な状態でも安定したナンバープレート認識が可能な技術を開発し、技術的な優位を築いてきた（**図3**）。これらの技術による高い競争力で、業界トップシェアを獲得している。

画像監視技術の新たな応用として、認証済みの利用者とともに未認証の人物が部屋に入ることを防止する入退室管理システム向けの共連れ防止技術や、店舗や施設でのお客様の動線を抽出・分析する技術への展開を積極的に進めている。

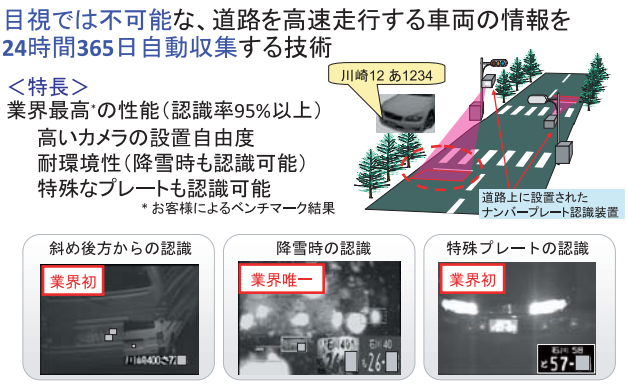



図3　ナンバープレート認識

Fig.3 License Plate Recognition

### （2）手のひら静脈認証

人間の体の特徴を利用して個人を識別・認証する生体認証技術が、銀行のATMをはじめとして、さまざまな分野で利用されてきている。

富士通では、1980年代から培ってきた画像認識技術を使った手のひら静脈認証技術を実用化している。この技術は、確実に本人を識別できる高い認証精度と、非接触で利用可能という特長を持つ方式である。さらに、静脈は体のなかの情報であり、第三者に盗まれにくいという、他の生体認証方式にない特長を有している。

認証の原理は、静脈内の還元ヘモグロビンが波長約760nmの近赤外光を吸収する特性を利用し、静脈の血管パタ

ーンを撮影するものである。この撮影パターンと予め登録したパターンとを照合し認証する。

2002年には世界初のマウス型認証装置を開発し、2003年には非接触型の手のひら静脈認証装置を開発した。これにより、衛生面での配慮が必要な場面でも安心して使えるようになった。2004年には世界で初めて金融機関での本人認証に採用され、ATMに搭載された。高い認証率と使い勝手の良さから、金融機関や教育機関、図書館などの公共の場に採用されてきている。2006年にはセンサが小型化され、適用分野の拡大と海外展開が活発化した。2009年には、動いている手のひらでも本人認証を可能にする高速撮影技術を開発した。この技術では、1,000分の1秒程度の撮影時間とすることで手のひらの動きによる画像のブレを防止するとともに、連続撮影した画像のなかから認証に最適な画像を自動的に識別できる機能を新たに開発した。この結果、駅の自動改札でICカードを近づけるような、手のひらを近づけていく動作のなかで認証を行う使い方への展開が期待されている（**図4**）。

図4　新しい手のひら静脈認証の利用イメージ
Fig.4 Usage Image of New Palm Vein Authentication

## 2.3 紙の暗号化

情報漏洩事件は、なかなか根絶できない社会問題である。電子的な形態のデータは、暗号化などの電子的な手段を適用することで、運用をきちんと行えば守ることができる。しかし閲覧性などの便利さの点で、漏れてはまずい情報であっても印刷する、ということを皆無にすることは難しい。

そこで、漏洩してはいけない情報にマスク（暗号化）処理を施してから印刷し、そのままでは読めないがパスワードを知っている人だけが解読して読むことができる「紙の暗号化」技術を開発した。暗号化の基本的な技術は、画像中の指定された領域を、スクランブルをかけて見えなくするものである。

**図5（a）**に示すように、印刷する機密文書をパソコンのディスプレイに表示し、マウス操作で暗号化する部分を選択する。解読に必要なパスワードを入力して暗号化した後、印刷する、という簡単な操作で、暗号化の作業を行うことができる。

印刷物を復元するときは、スキャナでその文書をパソコンに取り込む、あるいは携帯電話のカメラで撮影した後、暗号化された個所に対応するパスワードを入力すると、元の内容（文書、グラフ、図形、など）を再現することができる。

スキャナ用の復元技術を携帯電話に適用するにあたっては、紙とカメラの位置関係を固定できないことから生じる三次元的な歪みの補正（**図5（b）**）、光が暗号化領域全体に均一に当たらないために復元画像に生じる光学ノイズの除去（**図5（c）**）、などの技術を新たに開発するが必要があった。

ビジネスでもプライベートでも、多数のID（ユーザ番号、口座番号）やパスワードを持っている、あるいは使わざるをえない状況の人が増えていることを考えると、それらを安全にメモしておくような用途にもこの技術を適用できる。

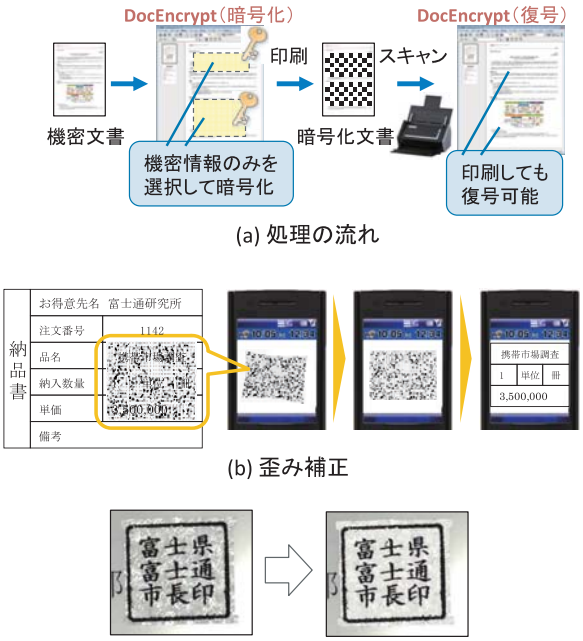



図5　紙の暗号化技術
Fig.5 Encryption Technology for Paper

# 3 音声処理技術

人間の最も基本的な伝達手段である音声を用いた情報交換を支援する技術として、富士通研究所では音声・オーディオ符号化、音声合成、音声認識の先端技術を1965年あたりから継続して開発してきた。音声・オーディオ符号化は主に人と人との情報交換の際に伝送や蓄積の効率を上げるための情報圧縮を担い、音声合成と音声認識は人とマシンの間の情報交換のためのインタフェースとなる。人間が話す内容をマシンが理解するために音声認識が、また逆にマシンからのテキスト情報を音声として人間に伝えるために音声合成が利用される。

これらの技術のルーツは1700年代の機械式音声合成器に端を発する音声生成過程の研究にあるといわれているが、1940年あたりのPCM（Pulse Code Modulation）の発明以降のデジタル化の発達で、デジタル音声処理技術として急激な進展を遂げた。

### 3.1 音声・オーディオ符号化技術

音声・オーディオ符号化はデジタル信号処理で音声信号を圧縮する技術であり、携帯電話や携帯音楽プレーヤなどに幅広く適用されている。

富士通研究所では、音声符号化の原理に人間の音声生成モデルを活用した情報圧縮技術であるCELP（Code Excited Linear Prediction）方式を採用し、圧縮性能と再生音質の両方をさらに高める独自の先端技術を開発してきた。そして、4kbps（圧縮率1/16）の低ビットレート条件で32kbps のADPCM（Adaptive Differential PCM）相当の高音質を実現する独自アルゴリズムを開発し、ITUの厳しい音質基準を満たしたため、1999年に方式提案まで進めた。

一方、楽音などを対象としたオーディオ符号化については、対象信号を周波数領域に変換した上で人間が知覚できない帯域を心理聴覚モデルで抽出することで情報圧縮する変換符号化について、独自技術を開発してきた。富士通研究所ではデジタルテレビ放送にも採用されたMPEG AAC（Advanced Audio Coding）の高音質化に2003年から取り組み、心理聴覚モデルの改良などにより、64kbps（圧縮率1/22）の低ビットレート条件でステレオCD音質に匹敵する高音質符号化方式を実用化した。

### 3.2 聞きやすさ向上技術

富士通研究所では、携帯電話向けの聞きやすさと使い勝手を追求した他社差別化の新技術開発で、携帯電話製品の市場競争力を強化してきた。ここでは独自の先進技術として二つを紹介する。

#### (1) はっきりボイス

周囲が騒がしいときに、相手の声を強調して聞きやすくする音声処理技術である（**図6**）。音声を周波数情報として表したスペクトルにおいて周囲騒音で邪魔されやすい高い周波数成分を増幅させることで、明瞭度を向上させて聞きやすい音声に加工することができる。富士通研究所の技術は、この強調処理を周囲騒音の特性に合わせて細かく調整することで最も聞きやすい音質を適応的に作り出す他社にはない長所を有している。

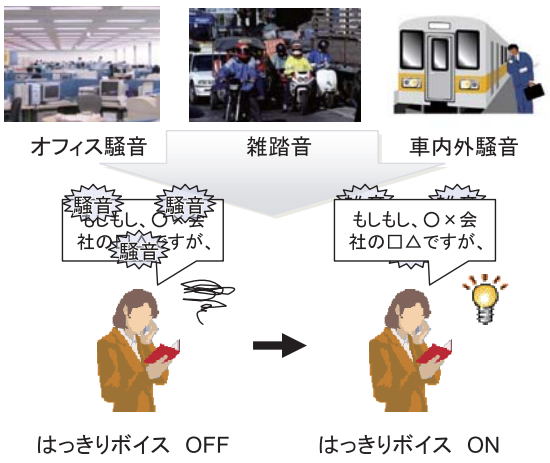



図6　はっきりボイスの概要

Fig.6 Outline of HAKKIRI VOICE (Clear Voice)

#### (2) ゆっくりボイス

受信側で相手の声をゆっくり出力することで、聞きやすさを向上させる話速変換技術である。基本技術は、音声の高さを変えずにゆっくり再生させることができる音声伸長技術であるが、この伸長による時間の遅れを会話中の無音部分を検出して回復させる独自の遅延制御技術と組み合わせることで、世界に先駆けて双方向通信である携帯電話への適用を実現している（**図7**）。

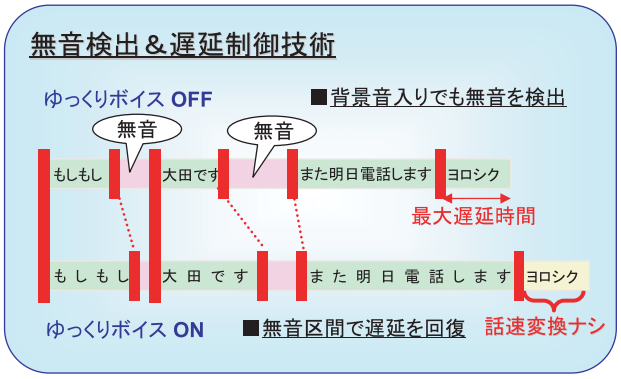



図7　「ゆっくりボイス」の原理

Fig.7 Principle of YUKKURI VOICE (Slow Voice)

### 3.3 音声合成・認識技術

富士通研究所では、業界をリードする音声インタフェースの実現に向けた音声合成、認識技術の先端技術を開発している。

#### (1) 音声合成

テキスト情報を音声で明瞭に伝える技術として「音声合成」技術を開発し、大手銀行のテレホンバンキングの自動音声応答システムや、富士通携帯機製品のらくらくホン（F671i以降）のメールの読み上げ機能として数多く採用されてきた。近年では、従来の音声の明瞭性に加え、極めて自然で人間の声に迫る音声合成を目指した研究開発を行っている。人間が日本語を話す際に、無意識のうちに自然に作り出すリズムに着目し、独自の発話リズム制御モデルを導入した。さらに、さまざまなイントネーションを網羅した使用頻度の高い数万個のフレーズを格納した大規模音声波形データベースを構築した。これらにより、プロのナレータに迫る高品質な音声合成が可能になった。この結果、自然な語り口が求められるテレビ、ラジオやeラーニング教材などにおいて、プロのナレータの代替としての利用が可能になった。

#### (2) 音声認識

音声認識は、人間が音声を理解する過程と同様に、まず話者の声から音素情報を抽出し、言語モデルを用いて単語や文章として理解（= 認識）する。富士通研究所では、入力音声から周波数領域の特徴量を抽出し、音声のさまざまなばらつきを表現できる確率モデル（HMM: Hidden Markov Model）と言語モデルの両方で最も確からしい文章を見つけ出す原理を採用している。さらに電話回線の伝

送周波数の制約や周囲雑音への耐性に優れた方式とすることで、電話を介した自動応答システムや、騒音環境下で作業員が音声で車両検査を進めるシステムなど、幅広い適用を可能としている。また、予めリストに設定したキーワードを検出するワードスポッティングにおいて、キーワードに似た発話を除外する独自モデルによる高精度化で、コールセンターや金融窓口業務に適用できる優れた実用性能を実現した。

このように、音声合成におけるナレータに迫る高音質性能の追求と、音声認識における使いやすいワードスポッティングへの進化により、機器のユーザ・インタフェースからコンプライアンス管理や内部統制まで、適用領域を大幅に広げている（**図8**）。

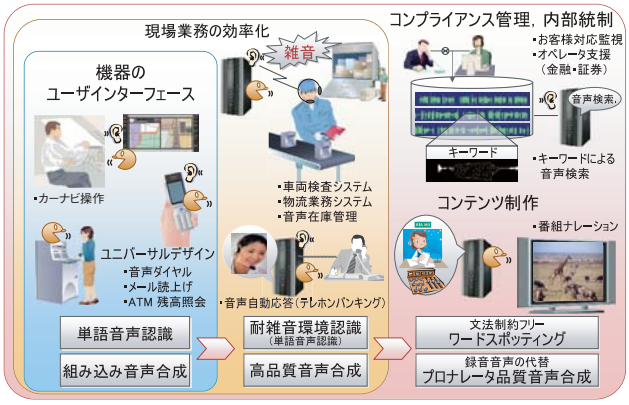



図8　音声合成・認識技術のトレンド

Fig.8 Trend of Speech Synthesis and Recognition Technology

## 4 おわりに

前章までに述べた、最近の成果とそれに到る開発の経緯を受けて、ここでは画像および音声の処理技術が、今後どういう方向に向かうのかを概観する。

まず画像は、普及しつつあるHDTV（High Definition Television）の延長線上に、HDTVの4倍の解像度のUDTV（Ultra high-definition TV）や16倍のSHV（Super Hi-Vision）がある。処理能力やメモリ容量の増強が必要になるだけでなく、信号の特性が変わるため、それに適した符号化方式の開発が必要と考えている。また、立体感（奥行き）を感じられる3D映像のサービスも始まったところであり、従来の映像にはない新規性は、今後の目玉のひとつと捉えている。

画像認識処理は、さまざま環境下で、より高い認識性能を安定して実現するとともに、利用者の自然な動きの中での認識・認証を実現することで、安心・安全と利便性を両立する技術として、日常生活に密接にかかわっていくと考えている。

音声符号化・音響信号処理においては、これまで、ネットワークや端末など伝送系からユーザ周囲の音響環境までの特性を取り込みながら、聞きやすさの向上を実現してきた。今後は聴覚から感性までを対象として、ユーザ個々の聞きやすさを追求していくことが、技術の優位性を継続させるために必要と考えている。

音声合成・認識においては、言語としての音声を対象に、耐環境性に優れる認識や明瞭で自然な音質の合成を実現してきた。今後は、意図やニュアンスなどのパラ言語 (1)、さらには感情などの非言語までを対象とした究極の音声インタフェースを目指した技術開発を行っていく。

＊(1) Paralanguage
話し手が聞き手に与える言語情報のうち、イントネーション、リズム、ポーズ、声質といった文字では表現できない情報。

社外執筆者紹介

松田　喜一
（まつだ　きいち）

1974年㈱富士通研究所入社。動画像と音声の信号処理技術、システム化技術、バイオメトリクス認証技術、ITSに関わる機器、ソフトウエア、およびシステム技術の研究開発に従事。現在、㈱富士通研究所フェロー。