

## 「米国における音声認識・画像認識技術とその応用を巡る動向」

市川類@JETRO/IPA NY

### 1. はじめに

これまで情報技術（IT）は、主として文字データを対象に発展してきたが、近年、ITの進展・データ処理能力の大幅な向上に伴い、これらに加えて、音声や画像データをコンピューターで取り扱うことが可能となってきている。

音声認識技術と画像認識技術については、実世界や人間とコンピューターとのインターフェースの改善に資するものとして、従来から研究開発がなされてきているが、近年、コールセンターでの自動応答やパスポート管理を含む公安分野などの企業向けのアプリケーションが普及しつつあり、また、ベンチャー企業やそれらが合併した中核企業が登場しつつある。

また、最近においては、モバイル化・クラウド化の流れの中で、スマートフォンなど消費者向けモバイルアプリケーションにベンチャー企業が参入し、急速に普及しつつあるとともに、更には、クラウド系インターネット企業の参入により、人の音声や画像以外を対象にした検索アプリケーションなども登場しつつある。

このような問題意識の下、本稿においては、米国における音声認識技術と画像認識技術に係る産業構造とそれらの応用を巡る最近の状況について、報告する。

### 2. 音声認識・画像認識技術の概要

#### （1）音声認識・画像認識技術の位置づけ

本報告書では、音認識（Sound recognition）及び画像認識（Image recognition）に係る技術、とりわけ、音声認識（Speech recognition）及び顔認識（Face recognition）を巡る動向について取り扱う。一般的には、音声認識技術とは、話者が発した言葉が何であるかを認識する機能、また、顔認識機能とは、デジタルイメージにおける人物（顔）を認識・特定する機能を指す。

#### ① 実世界・人間とのインターフェースとしての音声認識・画像認識

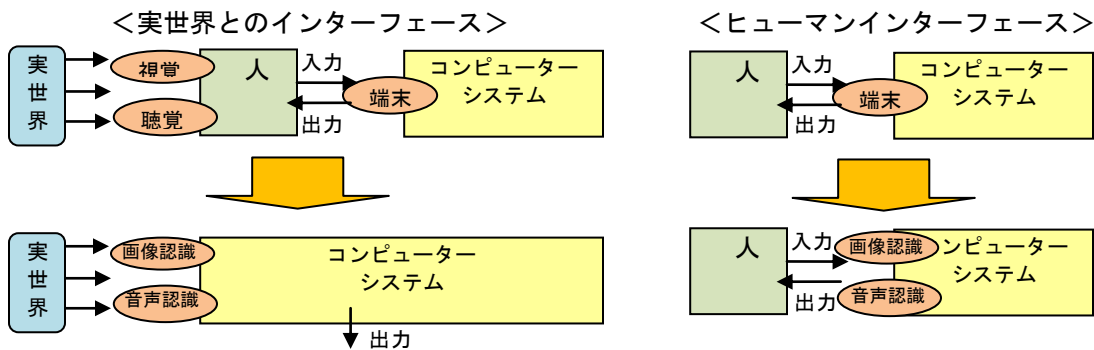
このような音認識や画像認識は、実世界や人間（ヒューマン）とコンピューターとインターフェースに係る技術の一部であると位置づけられる。

すなわち、これまで、人間（利用者）がコンピューターを利用するにあたっては、そのインターフェースとして各種端末機器（キーボード、タッチパネルな

ど)を物理的に操作し、文字を入力することによって動かす場合が多かったが、音認識や画像認識の技術を利用し、コンピューターに人間の五感の重要な要素である聴覚、視覚を持たせることによって、以下のようなインターフェース機能を有させることが可能となる。

- ・ **実世界とのインターフェース**：従来の人間がその視覚や聴覚を利用して判断し、コンピューターへの入力していた行為を、コンピューターに任せることができるようになる。これにより、人手に頼っていた入力作業を排除するとともに、実世界の動きに自動的に即応できる新たなコンピューターシステム<sup>1</sup>を構築することが可能となるものであり、具体的には、コールセンターの自動音声案内や公安系システムなど企業向けアプリケーションに多く利用される。
- ・ **ヒューマンインターフェース**：従来人間が端末を操作して入力していた行為を、音声（マイク）あるいは画像（カメラ）によって入力を可能とする。これにより、コンピューターをよりユーザーフレンドリーにするとともに、検索なども含めて新たなサービスも提供することが可能とするものであり、モバイル系や消費者向け製品などの消費者向けアプリケーションなどに多く利用される。

音声認識・画像認識の位置づけ<sup>2</sup>



② 情報技術（IT）の進展とモバイル化／クラウド化の影響

<情報技術（IT）の進展に伴う音声認識・画像認識技術の普及>

このような音声認識技術や画像処理技術は、当該技術の高度化だけでなく、情報処理能力の大幅な向上が背景となって、普及が加速され、より身近なものとなりつつある。

技術的に見た場合、文字情報が一次元の非連続なデータであるのに対し、音声情報は時系列上の一次元の連続データ（時間＋周波数）、画像情報は二次元の連続データ（縦＋横）であり、文字情報と比較して、情報量が格段に多い。

<sup>1</sup> 実世界の状況を把握して判断するシステムという意味で、(ヒューマノイドである必要はないが、)ロボットの一種であるともみなすことができるとともに、ロボットにとっても不可欠な技術要素である。

<sup>2</sup> 出典：筆者作成

このため、音声認識技術、画像認識技術のいずれの場合においても、（後述するとおり）、その多量のデータの中から、意味ある情報を取り出すべく、予め保存しているデータベースとの照合を含めて、大量のデータ処理が必要である。

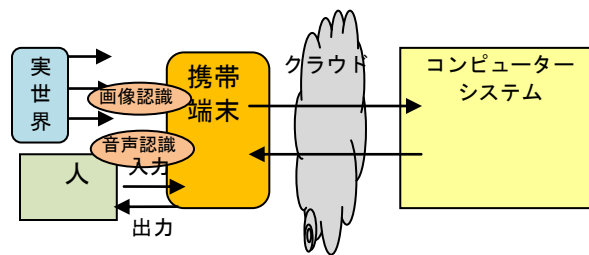
このような中、近年のハードウェアを含む情報技術（IT）の進展、情報処理能力の向上に伴い、これらの多量のデータ処理が容易になってきており、このため、これらの技術そのものの高度化と相まって、これらの技術の普及が進んできているものと位置づけられる。

#### <モバイル化とクラウド化による消費者向け音声認識・画像認識技術の普及>

特に、近年のクラウド化・モバイル化の進展<sup>3</sup>は、これらの技術の消費者向けアプリケーションなど身近なところでの普及促進に大きな影響を与えつつある。

すなわち、スマートフォンなどの携帯（モバイル）端末を利用し、音声を入力または写真を取るにより、音声・画像情報を取得し、一方照合すべきデータベースについては、クラウド上のデータセンターにおくことによって、多量のデータベースを必要とする音声・画像認識のアプリケーションを、身近なモバイル機器で利用することが可能となる。

#### モバイル化／クラウド化と音声認識・画像認識<sup>4</sup>



このような中、モバイル向け音声・画像認識技術に係る消費者アプリ、特に多量のデータベースを必要とするモノ（音楽、商品等）の認識技術に対して、これらのクラウドを有するインターネット企業が参入しつつある。

### （2）音声認識・画像認識技術の概要

#### ① パターン認識プロセスと検索・認証プロセス

音認識・画像認識技術は、概念的には、パターン認識プロセスと、検索・認証プロセスからなるものと考えられ、広義には、音・画像に係る検索プロセスであると位置づけられる<sup>5</sup>。

<sup>3</sup> NYだより2009年12月号(p2)参照。

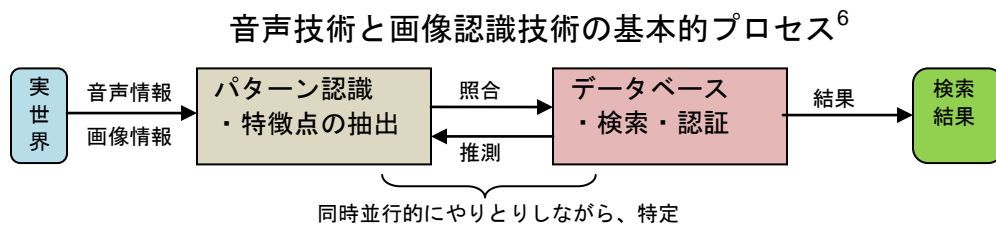
<sup>4</sup> 出典:筆者作成

- a) **パターン認識プロセス**：音情報、画像情報の多量のデータの中から、特徴となる情報をパターンとして抽出・認識し、情報量を減らすべくデータ処理を行うプロセス。具体的には、まずは入力された情報が、対象とする声・顔であることを認識した上で、特徴となるパラメーターに係るデータを、パターン認識に係るアルゴリズムに基づき、抽出することになる。
- b) **検索・認証プロセス**：それらによって抽出されたパラメーター情報について、予め保存しているデータベースとの照合を行い、そのパターンが何であるかを特定（認証）するプロセス。具体的には、データベースの中から、それらのパラメーターと合致するものを検索することになる。

ただし、実際には、認証精度を高めるためにも、アルゴリズムの実施（パターン認識）とデータベースとの照合（検索・認証）は、同時並行的に行われ、互いに情報をやりとりしながら、行われることになる。すなわち、

- ・ 時系列でのこれまでの音声データを、データベース（辞書等）と照合することにより、次に来ることが見込まれる音声データを推測する。
- ・ 画像の一部の情報を、データベースと照合しつつ、画像の他の部分を推測する。

などのプロセスをやりとりしつつ、情報を絞り込み、より精度の高い認証を確保するための手法・技術が採用される。



## ② 音認識・画像認識技術の対象範囲

音認識あるいは画像認識技術で特定する対象は、多岐に亘るが、いずれの場合も、大きく分けると「人」を対象とするものと、「モノその他の対象物」に分けられ、また、文字・話（Speech）などの「人の意思・意味」を対象とするものと、「それ以外」を対象とするものに分けられる。

<sup>5</sup> なお、経済産業省の実施する情報大航海プロジェクト(平成 19～21年度)では、「現在のパソコンによるホームページの中だけのテキスト中心の情報検索ではない、個人の生活、ビジネスなどあらゆる局面で、必要ときに必要な情報を解析できる情報基盤(プラットフォーム)を実現する…」としており、そのために「多種多様な情報の中から必要な情報を的確に検索・解析するための「次世代検索・解析技術」を開発するとともに、…」としている。

[http://www.meti.go.jp/policy/it\\_policy/daikoukai/igvp/index/index.html](http://www.meti.go.jp/policy/it_policy/daikoukai/igvp/index/index.html)

<http://www.igvpj.jp/index/>

<sup>6</sup> 出典：筆者作成

音認識・画像認識のまとめ表<sup>7</sup>

対象	DB	音認識	画像認識
人	文字 文章	<音声認識> ・音声内容（意味）の認識 →テキスト化 →コンピューター操作 →検索、翻訳等	<文字認識> ・文字の認識 →OCR（テキスト化） →バーコード（二次元を含む） （モノ特定、コンピューター操作等）
	個人 の特定 （個人データ ベースとの 照合）	<音声認識> ・人（声）であることの認識  <話者認証> ・個人の特定 →個人認証、犯罪捜査等	<顔認識> ・人（顔）であることの認識 →デジカメ写真の顔認識機能 <顔認証> ・個人の特定 →個人認証、犯罪捜査等 <指紋その他の生体認証（同上）>
その他対 象物	対象物の特 定（対象物 データベー スの照合）	<音の内容の特定> →音楽の特定（音楽検索）など	<モノの特定> →芸術作品の特定など →モノ・商品の特定など →場所、風景の特定など

<音認識技術の対象範囲>

このうち音認識（Sound recognition）技術の対象としては、人の声（音声）が中心になるが、その中でも、音声認識技術（Speech recognition）と話者認識技術（Speaker recognition）がある。また、人以外には、例えば、音楽などを特定する技術も最近登場してきている。

人に係る音声認識技術の概要とそのアプリケーション<sup>8</sup>

	概要	アプリケーション
音声認識 Speech recognition	<ul style="list-style-type: none"> <li>・人の声であることを認識し、音声内容（意味）を特定し、話し手の言葉を認証するもの。</li> <li>・基本的には、音声→テキスト変換がベース。言葉やフレーズに係る辞書データベースと照合。</li> </ul>	<ul style="list-style-type: none"> <li>・①音声のテキスト化（Speech-to-text、音声タイピング等）</li> <li>・②コンピューターの操作（音声ダイヤル、コールルーティング等<sup>9</sup>。）</li> <li>・③コンピューターの利用（検索・返答、翻訳・他言語での自動返答等）</li> </ul>
話者認識 Speaker recognition :	<ul style="list-style-type: none"> <li>・声の主体である話者を特定・認証するもの。</li> <li>・基本的には、声の周波数特性等を分析し、それを既存の個人のデータと照合をした上で、話者を特定。</li> </ul>	<ul style="list-style-type: none"> <li>・個人認証や犯罪捜査など<sup>10</sup>。</li> </ul>

<sup>7</sup> 出典：筆者作成。なお、緑部分は、今回での報告書で主に対象とする部分。

<sup>8</sup> 出典：筆者作成。

<sup>9</sup> <http://searchcrm.techtarget.com/definition/speech-recognition>

<sup>10</sup> <http://www.speaker-recognition.org/navApps.html>

＜画像認識技術の対象範囲＞

一方の画像認識（Image Recognition）の対象としては、以前から、OCR等の文字に係る認識・認証はあるが、それに加えて、近年、人、特に顔（Face recognition）を対象とするものに関心を集めるとともに、さらには、モノ（美術品・芸術品、商品・モノ等）や、風景（道路情報、ランドマーク等）などの対象物（Object recognition）とするものが登場・実用化しつつある。

このうち、人（顔）の認識技術に関しては、指紋、静脈認証などの生体認証と比較して、一般的には、精度は落ちるものの、よりユーザーフレンドリーであるとともに、実世界の画像からの判断も可能であるため、犯罪捜査など公安面等での利用範囲も広がるという特徴を有する。

文字・人に係る画像認識技術の概要とそのアプリケーション<sup>11</sup>

	概要	アプリケーション
文字認識	<ul style="list-style-type: none"> <li>・ OCR（光学文字認識 Optical Character Recognition）</li> <li>・ バーコード、二次元バーコード</li> </ul>	<ul style="list-style-type: none"> <li>・ ①画像（文字）のテキスト化</li> <li>・ ②コンピューターへの入力・操作（物品管理等）</li> <li>・ ③コンピューターの利用（検索等）</li> </ul>
顔認識 Face Recognition	<ul style="list-style-type: none"> <li>・ 画像の中から、顔（人であること）を特定するとともに、顔に係る画像から、その個人を特定・認証するもの。</li> <li>・ 基本的には、顔に係る画像情報を分析し、それを既存の個人のデータと照合をした上で、個人を特定。</li> </ul>	<p>＜人（顔）であることに認識＞</p> <ul style="list-style-type: none"> <li>・ デジカメの顔認識機能、セキュリティカメラの人物認識機能等</li> </ul> <p>＜個人の認識＞</p> <ul style="list-style-type: none"> <li>・ 個人認証や犯罪捜査など。</li> </ul>
その他生体認証	<ul style="list-style-type: none"> <li>・ 指紋、静脈認証（指、掌）、虹彩認証など</li> </ul>	<p>＜個人の認識＞</p> <ul style="list-style-type: none"> <li>・ 個人認証や犯罪捜査など。</li> </ul>

以下においては、上記のうち、音声認識技術の代表として、音声の意味内容に係る音声認識技術、また、画像認識技術の代表として、顔認識技術に係る技術の内容を中心に紹介するものとする<sup>12</sup>。

（3）音声認識・画像（顔）認識技術の具体的内容

上述の通り、音声認識技術も画像認識技術においても、基本的には、いずれもパターン認識・認証技術の一種であり、共通点を有する。すなわち、いずれの技術においても、対象となる音声情報・画像情報を構成しているデータ要素間の順序関係／位置関係を踏まえて、その特徴となるパターンを抽出しつつ、それらに、合致する類似データをデータベースから検出することで、認証プロセスが行われる。

<sup>11</sup> 出典：筆者作成。

<sup>12</sup> <http://www.face-rec.org/general-info>

ただし、音声認識と顔認識とは、そもそも対象が異なるため、当然異なる技術も利用される。具体的には、時系列に係る情報を対象とする音声認識技術は、もともと時系列特有の技術（線形予測符号化、nグラムモデルなど）から発展してきたのに対し、画像（顔）認識技術では、更に詳細な分析を行うよう、3次元の要素や、あるいは肌質の分析なども、併せて利用されるようになってきている。

音声認識、顔認識に利用される分析手法<sup>13</sup>

音声認識	音声認識技術	画像（顔）認識技術
時系列予測手法 （音声特有）	線形予測符号化（LPC） nグラムモデル など	—
パターン認識技術 （共通）	線状判別分析（LDA） 主成分分析（PCA） 隠れマルコフモデル <sup>14</sup> など	線状判別分析（LDA） 主成分分析（PCA） 隠れマルコフモデル など
3Dその他 （画像特有）	—	伸縮パンチグラフマッチング法 肌質分析 など

① 音声認識技術の具体的内容

音声認識技術では、まず、取得したデータから雑音などを排除した上で<sup>15</sup>、データベースと照合することになる。その際、人間が話す言葉には話し方に個人差がありばらつきがあるために、このようなばらつきも含めた大量のデータをコンピューターに学習させ、言葉や音声の特徴を抽出し、データとして補完している。

また、その照合にあたっては、単に音声データのの一つ一つを文字情報に置き換えるというのではなく、時系列の情報（文字等）、あるいは、その前後の情報（文字等）を踏まえた上で、そのように学習されたデータベースと照合し、データが、最も特徴に近く意味を成すような言葉が、統計的な手法によって選別されるような各種手法（LPC、n-gram、隠れマルコフなど）が開発されている。具体的な技術としては、以下の通り。

主要な音声認識技術の概要<sup>16</sup>

技術	概要
線形予測分析 （Linear	・過去に集められたデータを基に、ある時点以前、もしくは以後の音声波形を予測するというパターン認識手法の一種 <sup>17</sup> 。

<sup>13</sup> 出典:筆者作成。

<sup>14</sup> <http://www.isle.illinois.edu/pubs/2004/geirhofer04report.pdf>

<sup>15</sup> 雑音などへの対策としては、認識の推定に幅をもたせるミッシングフィーチャー理論(Missing Feature Theory)や、音源の位置情報を元に、複数の音源を分離させるという幾何学的音源分離法(Geometric Source Separation)など、音声認識の性能を向上させる可能性のある技術の開発も進められている。

<sup>16</sup> 出典:以下、その他各種資料より筆者作成。

[http://www.caip.rutgers.edu/~lrr/Reprints/354\\_LALI-ASRHistory-final-10-8.pdf](http://www.caip.rutgers.edu/~lrr/Reprints/354_LALI-ASRHistory-final-10-8.pdf)

<sup>17</sup> [www.cqpub.co.jp/hanbai/books/30/30901/30901\\_4Syo.pdf](http://www.cqpub.co.jp/hanbai/books/30/30901/30901_4Syo.pdf)

Predictive Coding:LPC)	・その後 1970 年代中盤までに LPC 手法に基づき、パターン認識を音声認識技術に応用するための様々な手法が考案され、その後の技術の発展に大きく貢献。
n-gram モデル <sup>18</sup>	・言語モデルの一種であり、データベースをもとに、(n-1) 個の言葉を元に、n 番目目にくるであろう言葉を確率的に予測する手法。 ・1980 年代に登場して以降、n-gram モデル (とその垂流モデル) は、音声認識システムになくなくてはならないものとなった。
線形判別分析 (LDA) <sup>19</sup>	・既にあるデータを元に、新しいデータが既存のどのデータ群に属するかを判断するものであり、音声認識では、入力された音声、言葉・フレーズのどのデータに属するかを判断する。
主成分分析 (PCA) <sup>20</sup>	・特徴となる主成分を抽出する技術であり、音声認識に利用される際には、話者に特有の特徴や残響などを排除し、音声の特徴となる主成分を抽出する。
隠れマルコフモデル <sup>21</sup>	・現在の音声認識技術の基礎となる統計的手法。マルコフ過程に従う確率モデル (直前の状態によって、次の状態が予測できる統計モデル <sup>22</sup> ) のうち、直前の状態が明らかになっていないモデルのことで、時系列のデータの処理に適しているとされる。 ・1980 年代に開発されたこのモデルは、音声認識だけでなく画像認識にも利用可能であり、現在に至るまで利用されている。

なお、現在の音声認証システムの性能<sup>23</sup>は、現在の技術では、ノイズなどを排除した理想的な環境で、特定の話者による限定された語彙であれば、高い認識率が達成できるとされるが、引き続き多くの課題も存在する<sup>24</sup>。

## ② 画像認識技術の具体的内容

画像 (顔) 認識技術においても、まずは、入力された画像情報に関し、データの歪みやノイズを取り除き、補正処理を行った後、その画像 (顔) に特有の特徴<sup>25</sup>を抽出した上で、データベースと照合することになる。

<sup>18</sup> <http://www.shuiren.org/chuden/teach/n-gram/index-j.html>

<sup>19</sup> <http://portal.acm.org/citation.cfm?id=1570285&dl=GUIDE&coll=GUIDE&CFID=86362009&CFTOKEN=33843003>、

[http://homepage2.nifty.com/nandemoarchive/GLM/tahenryou\\_03\\_discrim.htm](http://homepage2.nifty.com/nandemoarchive/GLM/tahenryou_03_discrim.htm)

<sup>20</sup> <http://www.sciencedirect.com/science/article/B6V14-3SNV03G-N/2/70093ed213e2ede9c01dc4529f4d1fc1>

<http://blog.goo.ne.jp/fuchik/e/b56a14a3462c640ba9083b87869a87e0>

<sup>21</sup> <http://hmm.shokuji.org/?p=5>、<http://hmm.shokuji.org/?p=4>

<sup>22</sup> 例えば、「Qu」の後には「estion (question)」、「iz (Quiz)」、「otation (Quotation)」がくると予測するなど。

<sup>23</sup> 一般的に、音声認証システムは、正確度と速度によって規定され、うち、正確度は単語誤認率 (word error rate: WER) で、また、速度は実時間係数 (real time factor: RTF) で表される。

<sup>24</sup> 例えば、以下の点が課題とされる。

- ・会議での発言等の複数の話者による会話の認識は困難。
- ・人間の通常の話し言葉 (会話) は、同じ単語でも話し手や会話の内容によって異なる意味を帯びる場合があるなど多様性があるため、上記で紹介した技術や理論だけで精度を大幅に向上させることは難しい。
- ・ソフトウェアを開発しても、実際に利用されるデータを収集したり、言語の利用パターンを習得したりするのに多くの労力と時間が必要とされる。



その照合にあたっては、単に一つ一つの特徴をデータベースと突合するだけでなく、音声認識と同様、各部分ごとの特徴とその周辺の特徴との相関関係も踏まえつつ、総合的な判断により、認識されたデータと最も近いものが特定される。具体的な技術としては、線状判別分析、主成分分析、隠れマルコフモデルといったアルゴリズムが利用されている<sup>26</sup>。

また、それに加えて、顔画像という特徴を踏まえた認識手法が登場してきている。具体的には、もともと顔は3次元であることを考慮した、伸縮バンチグラフマッチング法(EBGM)などの手法の利用が拡大しつつあるとともに、最近注目されている技術としては、画像による肌質分析技術があり、次世代顔認識システムの一環として、ここ2~3年で導入されるようになってきている<sup>27</sup>。

主な画像(顔)認識技術の概要<sup>28</sup>

技術	概要
線状判別分析 (Linear Discriminant Analysis : LDA <sup>29</sup> )	<ul style="list-style-type: none"> <li>線形判別分析では、既に集められているデータセットに基づき、新しいデータがどのデータセットに属するかを判別する。</li> <li>顔認識に利用される場合のデータセットは、複数の人物の顔データ(データ群)で構成され、各個人の様々な表情や角度から取った画像が含まれる。各データは直面データの形で多重に分解され、分解されたピースの中で最も多くの特徴を持ったピース(目の周りや口など)が分析に利用される。</li> </ul> <div data-bbox="721 997 1101 1276" style="text-align: center;"> <p style="text-align: center;">データ群</p> </div> <ul style="list-style-type: none"> <li>しかし、実際の顔サンプルは表情に変化があるのに対し、LDAで収集された情報ではデータが2次元情報として処理されるため、データベースに保存・学習させられる情報の多様性が少なく、誤差が起こりやすいという欠点がある。</li> </ul>
主成分分析	<ul style="list-style-type: none"> <li>画像認識における主成分分析では、顔の画像から特徴だけを抽出した後、その</li> </ul>

<sup>25</sup> なお、顔認識における特徴としては、両目の間隔、頬骨の高さ、鼻のサイズ、額の形状や眼窩の深さなどの特徴やそれぞれの位置関係といった情報があげられる。

なお、顔以外の認証であっても、認証の段階で照合されるデータセットが異なるだけで、認証そのものを利用される技術の土台は同じである。

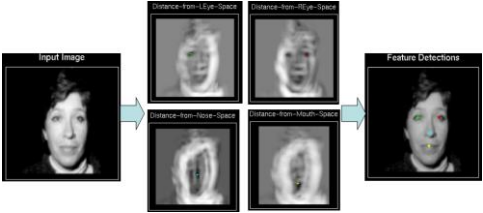
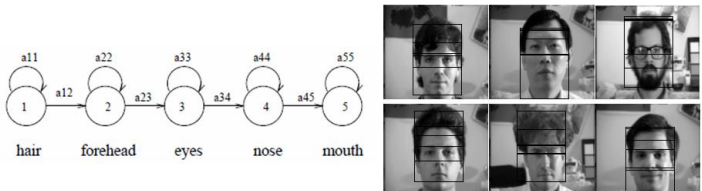
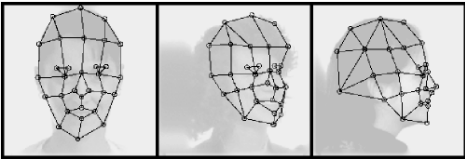
<sup>26</sup> <http://vis-www.cs.umass.edu/~vidit/Al/index.html>  
<http://www.howstuffworks.com/gadgets/high-tech-gadgets/facial-recognition3.htm>  
<http://www.aclu.org/technology-and-liberty/qa-face-recognition>

<sup>27</sup> <http://www.howstuffworks.com/gadgets/high-tech-gadgets/facial-recognition.htm>  
[http://www.animetrics.com/library/NextGenFaceRecognition\\_LL42406.pdf](http://www.animetrics.com/library/NextGenFaceRecognition_LL42406.pdf)

<sup>28</sup> 出典: 以下を含め、各種資料より筆者作成。

<http://www.biometrics.gov/Documents/FaceRec.pdf>

<sup>29</sup> [http://www.yobology.info/text/Markov\\_source/Markov\\_source.htm](http://www.yobology.info/text/Markov_source/Markov_source.htm)

<p>(Principal Components Analysis : PCA)<sup>30</sup></p>	<p>データをベクトルに変換し、距離や比率を測定し主成分を抽出するというもの。この場合、得られたデータを一次元の情報として扱うため、元のデータの1,000分の1という情報量の削減ができる</p>  <ul style="list-style-type: none"> <li>しかしその反面、目や口の位置を照準と合わせなければならず、顔前面の画像が少しでも欠けていると、認識精度が落ちるという欠点もある<sup>31</sup>。</li> </ul>
<p>隠れマルコフモデル<sup>32</sup></p>	<ul style="list-style-type: none"> <li>隠れマルコフモデルを、画像認識に利用する場合、例えば、顔イメージを、①髪の毛、②額、③目、④鼻、⑤口のブロックに分割し、各ブロックの特徴を分析し、予測するという形で利用する。</li> </ul> 
<p>伸縮バンチグラフマッチング法<sup>33</sup> (Elastic Bunch Graph Matching : EBGM)</p>	<ul style="list-style-type: none"> <li>線的に変換した顔の情報では失われてしまう光の加減や顔の向き、表情などの詳細な要素も顔認識の重要な要素として取り入れる手法。この方法では、顔から頭部にかけての情報を、伸縮性を持たせた標高グリッドで把握し、各結節点の特徴を画像処理で抽出する点が特徴。</li> </ul>  <ul style="list-style-type: none"> <li>なお、この方法では、顔認識の特徴となる部位の場所の正確な記録が難しいため、上記のPCAやLDAと併用されることが多い。DARPAによると、この方法は、それまでの技術と比較して認識の信頼性が大幅に高くなるとしている。</li> </ul>
<p>肌質分析<sup>34</sup></p>	<ul style="list-style-type: none"> <li>肌質分析技術では、顔の皮膚表面のサンプル写真を撮り、アルゴリズムを利用して皺や毛穴などの情報を抽出し、これを、位置関係や距離、サイズといった数的なデータに置き換え、皮膚表面のテクスチャーを分析する技術である。</li> <li>同技術を利用すると、既存の顔認識ソフトウェアでは正確に認識できなかった、双子など見た目がそっくりな対象でも違いを認識できるようになるほか、メガネの有無などに認識結果が左右される可能性も少ないという。</li> </ul>

<sup>30</sup> <http://vismod.media.mit.edu/vismod/demos/facerec/basic.html>

<sup>31</sup> <http://et.wcu/aids/BioWebPages/eigenfaces.htm>、  
<http://vismod.media.mit.edu/vismod/demos/facerec/basic.html>

<sup>32</sup> [http://www.anefian.com/research/nefian98\\_face.pdf](http://www.anefian.com/research/nefian98_face.pdf)

<http://www.stat.psu.edu/~jjali/hmm.html>

<sup>33</sup> [http://www.neuroinformatik.ruhr-uni-bochum.de/ini/VDM/research/computerVision/graphMatching/identification/faceRecognition/content\\_s.html](http://www.neuroinformatik.ruhr-uni-bochum.de/ini/VDM/research/computerVision/graphMatching/identification/faceRecognition/content_s.html)

<sup>34</sup> <http://www.howstuffworks.com/gadgets/high-tech-gadgets/facial-recognition.htm>

画像認識技術ではこれまでのところ、認証精度は必ずしも高いとは言えない<sup>35</sup>。そのため、近年では、従来の2次元的な画像認識で得られたデータに加え、より精度の高い、上述の3次元画像認識や肌質認識<sup>36</sup>などを併用するといった方法が取られてきている。

なお、顔前面の情報（2次元の情報）については、ANSI<sup>37</sup>やISO<sup>38</sup>が既に規格が策定されてきている。一方の、EBMGや肌質認識技術については、まだ、互換性のあるデータフォーマットや認識方法が設定されていない。このうち、3次元の画像認識については、システムや技術の標準化に向けた研究が進められており、国際的に統一された方法にするための取り組みが進められている<sup>39</sup>。

### 3. 音声認識技術の主要企業とアプリケーションを巡る動向

#### (1) 音声認識技術の歴史と市場・産業構造

音声技術の歴史は古く、また、大企業や政府の研究開発が起源となっているが、その後、企業や大学からのスピンオフがビジネスを開始し、その後、これらのスピンオフは合併により中規模の企業に育ってきている。

#### ① 音声認識技術の歴史<sup>40</sup>

音声認識技術の起源はもともと、AT&TのBell研究所が1940年代に開発した、発せられた単音を認識するという言語認識機器に遡る<sup>41</sup>。その後の1960年代には、1964年にIBMが初期の音声認識機器であるShoeboxを発表した<sup>42</sup>ほか、同年代後半に線形予測分析（Linear Predictive Coding）の音声認識への応用が確立され

<sup>35</sup> [http://www.nyu.edu/ccpr/pubs/Niss\\_04.08.09.pdf](http://www.nyu.edu/ccpr/pubs/Niss_04.08.09.pdf)

<sup>36</sup> 肌認識技術を利用した代表的な製品としては、Identix社（現L-1社）のFaceIT Argas、Animetrix社（後述）のFACEngine ID Set Lightなどがあげられ、前者の場合、肌質認識と既存の顔認識を併用することで、顔認識の精密度を20～25%向上する事に成功したと同社はしている。

<http://www.l1id.com/pages/71-facial-screening>

<sup>37</sup> ANSI NIST ITL 1-2000、ANSI/INCITS (M1)385-2004、385-2004 ANSI INCITS INCITS 398-2005 Common Biometric Exchange Formats Framework (CBEFF)

<http://www.itl.nist.gov/ANSIASD/sp500-245-a16.pdf>

[http://www.itl.nist.gov/div893/biometrics/documents/April%206\\_%20FP\\_Published\\_INCITS\\_M1\\_Standards.pdf](http://www.itl.nist.gov/div893/biometrics/documents/April%206_%20FP_Published_INCITS_M1_Standards.pdf)

<sup>38</sup> ISO SC37 19794-5 Face Recognition Data Interchange Format、<http://www.iso.org> ISO/IEC 19794-5:2005, 2004

<sup>39</sup> <http://www.biometrics.gov/Documents/FaceRec.pdf>

<sup>40</sup> [http://www.caip.rutgers.edu/~lrr/Reprints/354\\_LALI-ASRHistory-final-10-8.pdf](http://www.caip.rutgers.edu/~lrr/Reprints/354_LALI-ASRHistory-final-10-8.pdf)

<sup>41</sup> <http://www.lumenvox.com/resources/tips/historyOfSpeechRecognition.aspx>

<sup>42</sup> [http://www-03.ibm.com/ibm/history/exhibits/specialprod1/specialprod1\\_7.html](http://www-03.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html)

いる。また、同時期には、DARPA（当時の名称はARPA）がSpeech Understanding Researchプログラムを立ち上げ、音声認識研究を本格的に開始するきっかけともなっている。

その後においても、IBMやAT&TのBell研究所も音声認識技術の研究を継続しており、この結果、音声認証の基盤となる技術が複数開発されている。

- ・ 1980年代に、IBMの研究で開発されたn-gramモデル（およびその亜流）は、その後の音声認識技術にとって不可欠なものとなった。
- ・ AT&Tのベル研究所において、消費者に音声を利用した音声ダイヤルなどの自動サービスの提供に向け、様々な声の種類や訛りなども認識できるシステムが開発された。このようなシステムは、現在でも広く利用されている。
- ・ 続く1980年代にも、カーネギーメロン大学（CMU）が、隠れマルコフモデル<sup>43</sup>と呼ばれる統計的手法を音声認識に応用する<sup>44</sup>など、技術の進化が続いた。

このモデルも、現在でも音声認識に利用されている。

1990年代に入ると、大学やそのスピンオフを中心に、音声認識技術を利用したアプリケーションや製品が数多く登場するようになった。この中には、CMUが開発したSphinxシステム<sup>45</sup>、SRI研究所のDecipherシステム<sup>46</sup>、BBNのBYBLOSシステム<sup>47</sup>などが含まれる。また、AT&Tも各種ボイスコマンド・アプリケーションを発表・導入しており<sup>48</sup>、その後、これらのアプリケーションは現代の生活にも浸透するに至っている。

## ② 音声認識技術を巡る産業構造（主要企業）と市場動向

### <産業構造と主要企業>

このような中、米国においては、音声認識技術に係る多くのスピンオフが登場してきたが、その中で、Nuance Communication<sup>49</sup>（マサチューセッツ州Burlington）が、多くのスピンオフの技術を吸収し、主要企業として成長してきている。

<sup>43</sup>隠れマルコフモデルは元々、1960年代後半に防衛分析研究所(Institute for Defense Analysis)のLeonard Baum氏によって発表された統計手法。<https://www.ida.org>

<sup>44</sup>[http://www.caip.rutgers.edu/~lrr/Reprints/354\\_LALI-ASRHistory-final-10-8.pdf](http://www.caip.rutgers.edu/~lrr/Reprints/354_LALI-ASRHistory-final-10-8.pdf)

<sup>45</sup><http://cmusphinx.sourceforge.net/>

<sup>46</sup><http://www.speech.sri.com/projects/decipher/>

<sup>47</sup><http://www.bbn.com/resources/pdf/bbn-slp-w-call-021212.pdf> <http://www.bbn.com/speech/>

<sup>48</sup>AT&Tは1992年、かかってきた電話の音声を認識して適切な部署に配分するシステムであるVoice Recognition Call Processingを導入。

また、1990年後半には、コールセンターのオペレーターに代わる自動音声応答システムであるVRCPやHow May I Help You?システム(HMIHY)を発表。

[http://www.research.att.com/evergreen/about\\_us/tech\\_timeline.html](http://www.research.att.com/evergreen/about_us/tech_timeline.html)

<sup>49</sup><http://www.nuance.com/>、<http://www.technologyreview.com/communications/13212/page1/>

同社の原型は、MITからのスピンオフで、OCRソフトウェアの企業である ScanSoft である<sup>50</sup>が、その後これまでに、Dragon Systems（現在の音声認識技術の主力製品を開発）を始めとする多くのベンチャー企業の合併、IBMの音声認識技術の獲得、医療向け音声認識技術分野で大手であった Philips Speech Recognition Systemsの合併等を行い（特に2005年以降、5年間で20以上の企業・技術を獲得）<sup>51</sup>、現在では、音声認識、画像認識に係る総合ソリューション提供企業として成長してきている。2009年の売上は9.5億ドル<sup>52</sup>。

<sup>50</sup>同社の原型は、MITからスピンオフして1974年に設立されたOCRソフトウェア企業である Kurzweil Computer Products である。[http://americanhistory.si.edu/archives/speechsynthesis/ss\\_kurz.htm](http://americanhistory.si.edu/archives/speechsynthesis/ss_kurz.htm)  
同社は、その後、Xerox、次いで Visioneer という企業（1992年設立）に買収され、ScanSoftと名称を変更した。[http://en.wikipedia.org/wiki/Nuance\\_Communications](http://en.wikipedia.org/wiki/Nuance_Communications)

<sup>51</sup> ScanSoftは、2001年12月、不正経理等により倒産した Lernout & Hauspie 社（ベルギー系）の Speech & Language 部門を買収し（0.4億ドル）、同社が保有する音声技術を手に入れた。

Lernout & Hauspie 社は、1987年に設立されたベルギー系の音声認識企業であり、1996年には Text-to-Speech の開発企業である Berkeley Speech Technologies を、2000年に米国の音声認識技術で歴史を有する Dictaphone を（10億ドル）、また、その直後には、Dragon Systems を買収していた。この Dragon Systems は、1982年に設立された企業で、連続音声認識を行う世界初のソフトウェア NaturallySpeaking 1.0 を1997年6月に発売するなど、同分野における草分け的存在であり、Dragon の名称は、現在でも Nuance Communications 社の製品に冠されている。

[http://en.wikipedia.org/wiki/Lernout\\_%26\\_Hauspie](http://en.wikipedia.org/wiki/Lernout_%26_Hauspie)  
<http://en.wikipedia.org/wiki/Dictaphone>

Scan Soft は、2005年に、Nuance を買収（2.2億ドル）し、社名を Nuance に変更。Nuance は、1994年に、音声認識技術を主力技術として SRI International の Speech Technology and Research (STAR) Laboratory (加州 Menlo Park) からスピンオフした企業。

その後、Nuance は、2005年の合併以降も、音声認識関連技術関連企業を数多く（少なくとも22以上）買収あるいは技術取得を行っている。[http://en.wikipedia.org/wiki/Nuance\\_Communications](http://en.wikipedia.org/wiki/Nuance_Communications) その中には、以下の通り、IBM や Philips などが含まれる。

- ・eScription（音声認識での医療診断報告作成システム：2008年3.4億ドル）
- ・Philips Speech Recognition Systems（2008年1.0億ドル）（なお、同社の買収は、独占禁止法上の問題を引き起こした。）[http://www.nuance.com/news/pressreleases/2008/20081001\\_philips.asp](http://www.nuance.com/news/pressreleases/2008/20081001_philips.asp)
- ・Jott（留守番電話のメッセージをテキストに変換：2009年）  
[http://www.businesswire.com/portal/site/home/permalink/?ndmViewId=news\\_view&newsId=20090714005664&newsLang=en](http://www.businesswire.com/portal/site/home/permalink/?ndmViewId=news_view&newsId=20090714005664&newsLang=en)
- ・IBMの特許を取得（2009年）  
[http://japan.nuance.com/news/20090116\\_ibm.asp](http://japan.nuance.com/news/20090116_ibm.asp)

<IBMでの音声認識技術に係る経緯>

IBMでは1950年代から音声認識技術の研究開発を行っており、1996年に音声認識製品 VoiceType 3.0をWindows95向けの音声テキスト変換ソフトウェアとしてリリースすると共に、IBMの独自OSであるOS/2 Warpにも統合する形で展開している。[http://www.research.ibm.com/hlt/html/body\\_history.html](http://www.research.ibm.com/hlt/html/body_history.html)

その後、1997年8月には当時まだ市場で活躍していた Dragon Systems の音声認識ソフトウェア製品に対抗すべく、同様の機能を持つ ViaVoice を99ドルで発売。ViaVoice の発売を受け、Dragon Systems は Naturally Speaking を、初期の価格である\$700から\$299へ、その後も\$199まで値下げしている。  
<http://www.technologyreview.com/communications/11740/?a=f>

このような数々の買収・合併の結果、同社が展開する音声認識ソフトウェアは、音声認識して文書を作成する「Dragon NaturallySpeaking<sup>53</sup>」、同ソフトウェアのカルテ作成バージョンである「Dragon Medical<sup>54</sup>」を含めて多岐に亘る。なお、同社は、それ以外にも、PDFコンバーターなどのなど、全く別分野のソフトウェアアプリケーション製品も取り扱っている。

#### <音声認識技術に係る市場規模>

Opus Research の調査（2007年3月頃）<sup>55</sup>によると、2006年の音声認識技術関連市場全体の規模は、2004年からの2年間で100%増加し、10億ドルを突破したとしている。また、分野別では、特にあげられている分野は以下の通り。

- ・ コールセンター向けのサーバーベース音声認識システム市場は、2006年に6億ドルに到達、2009年までに2倍に成長する見込み。
- ・ 医療分野における市場規模は、2008年に1.7億ドルに達し、2013年にかけて同市場は倍増する見込み<sup>56</sup>。
- ・ 携帯電話や自動車などに搭載される音声認識技術の市場規模は、2006年に1.25億ドルとなり、2010年までに5億ドル市場に拡大する見込み<sup>57</sup>。

このような中、近い将来、音声認識システムはアプライアンスや機器など、人々が利用するあらゆる機器に搭載されるようになるだろうとしており、市場が拡大していくと予測されていた。

しかしながら、その後、同社は2009年1月、ViaVoiceを含む、音声認識関連の特許権やライセンスをNuanceに売却(金額は非公開)<http://online.wsj.com/article/SB123202456169485393.html>

また、同年7月には、両社は10産業分野における音声認識ソリューションの開発・改良に向けて協力体制を構築すると発表した。[http://www.nuance.com/news/pressreleases/2009/20090727\\_ibm.asp](http://www.nuance.com/news/pressreleases/2009/20090727_ibm.asp)

一方、Nuanceに吸収されたViaVoiceの生産は、同年8月には、生産中止となっている。

<http://japan.nuance.com/viavoice/>

ただし、コンピューターに組み込まれた形でのViaVoiceは、現在でもIBMより提供されている模様。

[http://www-01.ibm.com/software/pervasive/embedded\\_viavoice/](http://www-01.ibm.com/software/pervasive/embedded_viavoice/)

<sup>52</sup> [http://www.nuance.com/news/pressreleases/2009/20091123\\_q4FY09.asp](http://www.nuance.com/news/pressreleases/2009/20091123_q4FY09.asp)

<sup>53</sup> <http://www.nuance.com/naturallyspeaking/products/editions/preferred.asp>

<sup>54</sup> [http://www.nuance.com/healthcare/products/dragon\\_medical.asp](http://www.nuance.com/healthcare/products/dragon_medical.asp)

<sup>55</sup> 2007年3月14日付のCNN.comの記事。

[http://money.cnn.com/magazines/business2/business2\\_archive/2007/02/01/8398978/index.htm](http://money.cnn.com/magazines/business2/business2_archive/2007/02/01/8398978/index.htm)

<sup>56</sup> Healthcare IT News誌(2008年6月30日付け)。2008年のPCおよびサーバーベースの音声認証技術市場の85%が、医療記録転写を中心とした医療分野における利用。また、将来的には、法務関係の文書作成や、言語習得支援分野への音声認識技術の利用が拡大すると予想。

<http://www.healthcareitnews.com/news/market-healthcare-speech-recognition-technology-hit-340-million-2013?page=0,1>

なお、オンライン業界誌Digital HealthCareによる、Philips Speech Recognition System社(2008年10月にNuance社が買収)のマーケティング・マネージャーに対するインタビュー(2008年7月)によると、米国の医療分野における音声認識技術の普及率は、分野によって差はあるものの、平均すると約10~20%であるとしている

<http://www.bio-itworld.com/2008/07/22/speech-recognition.html>

<sup>57</sup> 調査会社DataMonitorの調査結果。

(2) 音声認識技術に係る企業向けの主要アプリケーションを巡る状況

このうち、音声認識技術の企業向けアプリケーションとしては、その範囲は、コールセンター、自動音声案内を筆頭に、医療関係や、軍事・航空関係などの分野があげられ、これらの分野に対して、上述の Nuance や、旧 Bell 研の技術を含むベンチャーなどの企業が積極的に参入をしている。

音声認識技術に係る企業向け主要アプリケーション<sup>58</sup>

分野	概要	主要ベンダー
コールセンター、自動音声案内	<ul style="list-style-type: none"> <li>・ 音声認識技術のアプリケーション分野の中でも最も大きい分野。コールセンターに電話をかけてきた顧客に対し、必要情報の音声入力を求め、その音声を認証・認識して、顧客が求める情報を提供したり、適切な部門へ通話が転送されたりするというシステム<sup>59</sup>。</li> <li>・ 自動音声システムを導入することで、企業は、オペレーターの人件費、コールセンター維持に係るコストを削減することが出来る。</li> </ul>	<ul style="list-style-type: none"> <li>・ 同分野における有力ベンダー<sup>60</sup>は、Alcatel-Lucent傘下（旧 AT&amp;T Bell 研を含む）の Genesys や、ベンチャー企業の Voxeo、Voxify など。</li> <li>・ 一方、現 AT&amp;T は、2006年11月、音声認識技術を利用した無料電話案内サービスを提供する InFreeDA を買収<sup>61</sup>。</li> </ul>
医療分野	<ul style="list-style-type: none"> <li>・ 音声認識技術を、診断報告書作成業務などの分野において応用。医師らは、診療結果などを音声で発声するだけで、ソフトウェアが自動的に診断報告書の書式に合わせた文書を作成。出力された文書は、診断報告書の下書きとして使用できるため、最終報告書の作成時間を著しく短縮できるとされている<sup>62</sup>。</li> </ul>	<ul style="list-style-type: none"> <li>・ 主要ベンダーとしては、以下の企業があった。</li> <li>▽eScription<sup>65</sup>の AutoScript<sup>66</sup>（自動診断報告書作成ソフトウェア）</li> <li>▽Philips Speech Recognition Systems（Philips SRS）<sup>67</sup>の Speech Magic<sup>68</sup>の実績。</li> </ul>

<sup>58</sup> 出典：各種資料より筆者作成。

<sup>59</sup> <http://www.callcentrehelper.com/the-top-five-uses-of-speech-recognition-technology-1536.htm>  
<http://www.em-t.com/Speech-Recognition-Technology-in-the-Call-Center-s/194.htm>  
[http://www.wizzardsoftware.com/att\\_nv\\_landing.php](http://www.wizzardsoftware.com/att_nv_landing.php)

<sup>60</sup> <http://ivr.tmcnet.com/topics/ivr-voicexml/articles/54645-datamonitor-selects-genesys-as-top-ivr-vendor.htm>、<http://ivr.tmcnet.com/topics/ivr-voicexml/articles/67433-ovum-names-voxify-top-speech-application-vendor.htm>

Genesys(1990年設立、Daly City, CA)は、2000年に Alcatel(現 Alcatel Lucent)の傘下になっている。Voxeoは1999年設立(Orlando, FL)、Voxifyは2001年設立(Alameda, CA。2008年 Genesysとも連携)。<http://www.genesyslab.com/>、<http://www.voxeo.com/about/home.jsp>、<http://www.voxify.com/>

<sup>61</sup> InFreeDAは、無料の番号案内サービスを提供し、該当ビジネスへの電話転送サービスを提供(利用者に音声広告を聞かせることにより収入)するベンチャー企業。

<http://techcrunch.com/2006/12/26/att-acquires-infreeda-gets-into-free-411-business/>  
<http://investing.businessweek.com/research/stocks/private/snapshot.asp?privcapId=26559321>

なお、同ライバル企業としては、Jingle Network、TellMe(後に Microsoft が買収。後述)、Google-411(後述)などがある。(NYだより2008年2月号参照)

<sup>62</sup>例えば、Health Alliance of Greater Cincinnati では作業効率が二倍になったとの報告がある。

[http://escription.net/customer\\_stories\\_hal.htm](http://escription.net/customer_stories_hal.htm)

	<ul style="list-style-type: none"> <li>・今後、電子カルテシステムが導入されるにつれ、利用が促進されると考えられている<sup>63</sup>。例えば国防総省では、2011年までに500以上の医療施設に導入予定<sup>64</sup>。</li> </ul>	<ul style="list-style-type: none"> <li>・両社とも前述の通り、Nuanceに買収されている。</li> </ul>
<p>軍事分野 航空分野</p>	<ul style="list-style-type: none"> <li>・軍事分野での音声認識技術のアプリケーションとしては、音声認識機器を戦闘機に搭載したシステム<sup>69</sup>の開発や、空軍オペレーションセンターでの手順入力の応用などが試みられている<sup>70</sup>。また、その他分野において研究が進められている<sup>71</sup>。</li> <li>・また、航空分野<sup>72</sup>では、航空交通管制官養成訓練システムにおいて音声認識技術を応用<sup>73</sup>。既に、空軍や海兵隊、陸軍、FAAなどの軍関係・政府機関で既に使用されている。</li> </ul>	<ul style="list-style-type: none"> <li>・軍事分野での実際の製品例としては、NuanceのSmart Speech<sup>74</sup>（開発者はAmcom<sup>75</sup>）など<sup>76</sup>。</li> <li>・また、航空交通管制官養成訓練システムとしては、UFA<sup>77</sup>やAdacel<sup>78</sup>などのシステムが導入されている。</li> </ul>

<sup>65</sup> <http://www.healthtechwire.com/The-Industry-s-News-unb.146+M5ebbc47ca8a.0.html>

<sup>66</sup> [http://www.escription.com/news\\_AutoScript7.htm](http://www.escription.com/news_AutoScript7.htm)

<sup>67</sup> 同社は、1994年に世界初の話し言葉をPC上でテキストに変換する音声認識エンジンを開発・販売。

[http://findarticles.com/p/articles/mi\\_m0EIN/is\\_2007\\_Dec\\_10/ai\\_n21148729/](http://findarticles.com/p/articles/mi_m0EIN/is_2007_Dec_10/ai_n21148729/)

<http://www.dictation.philips.com/index.php?id=start>

<sup>68</sup> <http://www.myspeech.com/>

Speech Magicは2007年、Frost & Sullivan社のGlobal Excellence Awardに選定され、世界的に優れていると認定された。同賞は、国際的見地から、革新的なビジネスやテクノロジー、品質、マーケティング方法を有すると判断された企業に与えられる賞。

<http://www.frost.com/prod/servlet/press-release.pag?docid=122809609>

クライアント・サーバーベースの医療用アプリケーション(25ヶ国語対応)を、世界50カ国、8,000件導入(ユーザー数は65,000人)の実績(当時)。

[http://findarticles.com/p/articles/mi\\_m0EIN/is\\_2007\\_Dec\\_10/ai\\_n21148729/](http://findarticles.com/p/articles/mi_m0EIN/is_2007_Dec_10/ai_n21148729/)

<sup>63</sup> <http://www.informationweek.com/news/software/enterpriseapps/showArticle.jhtml?articleID=207800986>

<sup>64</sup> AHLTA(Armed Forces Health Longitudinal Technology Application)と呼ばれる電子カルテシステム。医師がカルテにタイプすることなく、患者の診察内容を記録することができるようにするとの計画。なお同システムの一部にも、NuanceのNaturally Speaking Medical Speech-recognitionソフトを使用予定。

<sup>69</sup> ラジオ周波数の設定、自動操縦装置の作動、爆弾の投下地点の照準あわせなどを口頭による指令で行うことができるシステム。

<sup>70</sup> 戦域内空域戦術命令策定システム上での音声認識技術の応用や、無人飛行機オペレーションセンターでの手順入力。<http://www.springerlink.com/content/k00rv66563751454/>

ただし、同分野については、これまでのところ、成功例は非常に限定的であるとされる。

<sup>71</sup> 通信・ナビゲーション分野に、航空電子工学技術と音声認識技術を融合させる研究が重視されている。

[http://www.f-16.net/news\\_article2571.html](http://www.f-16.net/news_article2571.html)

また、DARPAでは2010年2月12日に音声認識技術のさらなる向上、特に雑音などの要因により音声信号の劣化がある場合の認識力向上をめざし、新規の研究プログラムの開始を発表している。

<https://www.fbo.gov/index?s=opportunity&mode=form&id=ccc472998ed4fa09aaad60e9f732e98b&ab=core&cvview=1>

<sup>72</sup> [http://www.voicesearchbar.com/Speech\\_voice\\_Recognition\\_in\\_Training\\_air\\_traffic\\_controllers.html](http://www.voicesearchbar.com/Speech_voice_Recognition_in_Training_air_traffic_controllers.html)

<sup>73</sup> 具体的には、訓練シミュレーションにおいて、管制官とのやりとりするパイロット役をする人員を、音声認識と音声合成機能をもったコンピューターシステムに置き換えるもの。なお、パイロットと管制官の間の会



その他に、例えば、公判記録の分野では、裁判での公判を音声から自動的にテキスト変換するアプリケーションである Dragon NaturallySpeaking (Nuance 社) などが販売されている<sup>79</sup>。

### (3) 消費者向け・モバイルアプリケーションを巡る動き

一方で、近年、IT 技術の進展に伴い、このような音声認識技術は、コンパクトな機器に収めることができるようになってきており、また、タッチスクリーンなど、消費者がハイテク製品に対する「自然なインターフェース（かかわりかた）」を求めてきていることを背景に、音声認識技術の応用は、今後、更に家電をはじめ、消費者向けに身の回りに広く浸透していくと見られている<sup>80</sup>。

具体的な例としては、電話や PC での利用だけではなく、特に音声入力为前提とされ、近年急速に進展しつつあるスマートフォンなどのモバイル分野を中心に、大きく普及が進みつつある。具体的に搭載が進みつつある消費者向け機器とそのアプリケーションとしては、以下のようなものがあげられる。

音声認識技術の消費者向けアプリケーション (例) <sup>81</sup>

端末・機器	アプリケーション
電話・PC	<ul style="list-style-type: none"> <li>・留守番電話の録音内容のテキスト変換、ハンズフリー・タイピング</li> <li>・ハンズフリー・コンピューティング<sup>82</sup> (音声認識によってパソコンを操作)</li> <li>・音声キーワード検索 (Podcast などで公開されたスピーチデータから特定のキーワードを照合させる検索等<sup>83</sup>)</li> </ul>

話の形式は比較的明確に決まっているため、音声認識アプリケーションで問題が発生する可能性はそう高くないとのことである。

<sup>74</sup> 戦場で、兵士が通話先の部隊名を発声することで、自動的に電話を繋いだり、ダイヤルしなくとも、電話番号案内や本国への電話などを音声発信したりすることができるというもの。

<sup>75</sup> <http://www.amcomsoftware.com/index.aspx>

<sup>76</sup> [http://findarticles.com/p/articles/mi\\_m0EIN/is\\_2002\\_Sept\\_9/ai\\_91196701/](http://findarticles.com/p/articles/mi_m0EIN/is_2002_Sept_9/ai_91196701/)

<sup>77</sup> <http://www.ufainc.com/> マサチューセッツ州ボストンに拠点を置く企業。2008年9月に音声合成、認識機能をもった、ATVoice Voice Recognition and Response System を発表している。

<sup>78</sup> 1987年創設。北米の本拠地はフロリダのオーランド。<http://www.adacel.com/ATCIntro.html>, [http://www.adacel.com/press/2007/ada\\_AUSairsvcs\\_contract.pdf](http://www.adacel.com/press/2007/ada_AUSairsvcs_contract.pdf)

<sup>79</sup> [http://www.nuance.com/naturallyspeaking/pdf/cs\\_JenniferSmith\\_rev2\\_PF1.pdf](http://www.nuance.com/naturallyspeaking/pdf/cs_JenniferSmith_rev2_PF1.pdf)

<sup>80</sup> <http://www.physorg.com/news167499933.html>

<sup>81</sup> 出典: 各種資料より筆者作成。

<sup>82</sup> <http://www.lumenvox.com/resources/tips/historyOfSpeechRecognition.aspx>

また、ハンズフリー・コンピューティングや声を機械的に変えたりするフリーの音声認識ソフト等も販売されている。<http://www.mymusictools.com/download/voice-recognition/>

<sup>83</sup> 例えば、ハイテク大手 Raytheon 社グループの BBN テクノロジー社による Podzinger 音声認識サーチエンジンなどがある。

携帯電話などの モバイル端末	<ul style="list-style-type: none"> <li>・音声による自動発信、受信音声のテキスト変換。</li> <li>・音声によるモバイルインターネット、各種検索等。</li> <li>・自動通訳・翻訳<sup>84</sup></li> </ul>
自動車の各機器	<ul style="list-style-type: none"> <li>・音声による自動車内機器の操作</li> </ul>
ビデオゲーム、 ペットロボット 等	<ul style="list-style-type: none"> <li>・音声によるゲーム機器の操作<sup>85</sup>。</li> <li>・ペットロボットへの音声での指示・コミュニケーション</li> <li>・語学学習の発音評価システム</li> </ul>
ホームオートメ ーション	<ul style="list-style-type: none"> <li>・音声により各家庭機器（エアコン、セキュリティシステム、ホームシアター、テレビや電話）を操作・制御<sup>86</sup>など。</li> </ul>

このような中、音声認識に関する専門誌である Speech Strategy News の編集を手がける Bill Meisel 氏（2009年7月）<sup>87</sup>によると、ボイス・コマンドの分野は今後さらに発展・普及する見込みであるとしていた他、音声認識技術はナビゲーションシステムから目覚まし時計、家電に至るまで、遠隔操作ができるあらゆる製品に応用されていくだろうと予測している。

以下では、このうち、特にモバイル系及び自動車向けのアプリケーションに係る動きとその事例を紹介する。

#### ① モバイル系の音声認識アプリケーション

<ベンチャー企業とスマートフォン（iPhone）アプリ>

近年の iPhone を初めとするスマートフォンの普及と App Store ビジネスの拡大の中で、音声認識技術に係る多くのベンチャーが、モバイル向けの各種のアプリケーションを発表してきている。

また、この中でも、最近では、単に音声に係る認識技術だけではなく、音認識技術を活用した音楽検索技術が、ベンチャー企業によって登場してきている。

---

<http://www.ramp.com/2006/01/podzinger-makes-podcast-searching-fast-easy-and-accurate-%E2%80%93-advanced-speech-recognition-technology-delivers-relevant-search-results-for-multimedi/>、<http://websearch.about.com/od/imagesearch/a/podzinger.htm>  
[http://www.mymusictools.com/developers\\_kit\\_17/research\\_lab\\_speech\\_engines\\_suite\\_23473.htm](http://www.mymusictools.com/developers_kit_17/research_lab_speech_engines_suite_23473.htm)

<sup>84</sup>有力企業は Voxtec。同社は DARPA の研究開発プロジェクトで開発した自動同時通訳技術を用い、2002年にハンドヘルドタイプの Phraselator を開発した。現バージョンの Phraselator は 40 以上の言語をサポートしている。<http://www.voxtec.com/>

<sup>85</sup>有力ベンダーは Fonix Speech で、Microsoft の Xbox や Sony の Playstation に導入。  
<http://www.eweek.com/c/a/VOIP-and-Telephony/Xbox-360-Video-Game-to-Feature-Voice-Recognition-Technology-from-Fonix/>

また、音声認識技術を応用したフライト・シミュレーションゲームが比較的多く市場に出回っている。  
<http://www.speechtechmag.com/Articles/News/News-Feature/Voice-Joins-Gaming-Battle-24544.aspx>、<http://www.atcsimulator.com/info.htm>

<sup>86</sup>有力ベンダーは Home Automated Living など。<http://www.automatedliving.com/>

<sup>87</sup> 2009年7月の PhysOrg 誌 <http://www.physorg.com/news167499933.html>

スマートフォン（iPhone 等）アプリを提供する音声認識技術系ベンチャー企業<sup>88</sup>

企業名	概要	アプリの販売
Vlingo <sup>89</sup> 2007年設立 <sup>90</sup> Cambridge, MA	<ul style="list-style-type: none"> <li>音声テキスト変換アプリケーションを開発・提供するベンチャー企業。</li> <li>同社のアプリでは、音声での通話発信やウェブ検索ができる。</li> </ul>	<ul style="list-style-type: none"> <li>iPhone、Blackberry、一部の Nokia 携帯向け等に無料で提供。</li> <li>Blackberry では無料版に加え、有料版の Vlingo Everywhere も提供<sup>91</sup>（音声認識範囲を拡大）。</li> </ul>
VoiceBox Technologies <sup>92</sup> 2001年設立 <sup>93</sup> Bellevue, WA	<ul style="list-style-type: none"> <li>音声認識に関する技術やこれらの技術を利用した製品を提供する企業。</li> <li>VoiceBox の技術は会話認識プロセッサと音声検索エンジンからなり、後者は発音された音声を認識するだけでなく、会話の内容によって意味を判断する能力もある<sup>94</sup>。</li> </ul>	<ul style="list-style-type: none"> <li>同社の技術は、自動車、家電等の様々な業界で使用されている<sup>95</sup>。</li> <li>iPhone 向けアプリとしては、VoiceBox Dialer（音声認識で電話帳内に記録した通話発信先を呼び出す）を無料で<sup>96</sup>、また、VoiceBox Dialer Plus（高機能版）を 2.99 ドルで提供<sup>97</sup>。</li> </ul>
Siri <sup>98</sup> 2007年設立 <sup>99</sup> San Jose, CA	<ul style="list-style-type: none"> <li>Google, Yahoo, Apple, Xerox, SRI, NASA のエンジニアらにより立ち上げられた企業。</li> <li>2010年2月、音声認識を利用した Siri Personal Assistant を発表。</li> </ul>	<ul style="list-style-type: none"> <li>同アプリ（無料）では、では、音声によりレストランや映画、天気、イベントなどを検索、予約する事ができる。</li> <li>現在、iPhone アプリのみ。</li> </ul>
Fonix Speech <sup>100</sup> Lindon, UT	<ul style="list-style-type: none"> <li>Fonix Corporation（1994年設立、Salt Lake City, UT）の 100%子会</li> </ul>	<ul style="list-style-type: none"> <li>iPhone 用の音声コマンドアプリの Fonix iSpeak（9.99 ドル<sup>101</sup>）、</li> </ul>

<sup>88</sup> 出典：各種資料より筆者作成。

<sup>89</sup> <http://www.vlingo.com/>

<sup>90</sup> シリーズ A 投資（2006年12月）で 6.5 百万ドル（Charles River Ventures、Sigma Partners）、シリーズ B 投資（2008年4月）で計 13.5 百万ドル（Yahoo、Charles River Ventures、Sigma Partners）を取得。

<http://www.crunchbase.com/company/vlingo>、<http://www.crunchbase.com/company/vlingo>

<sup>91</sup> <http://www.vlingo.com/products/blackberry/>

<sup>92</sup> <http://www.voicebox.com/>

<sup>93</sup> 設立資金は計 21 百万ドル（InfoSpace とエンジェル投資家）。2008年、13 百万ドルの追加出資を獲得（AutoNavi、Iventec Appliance、MiTAC、Morningside など）

[http://www.techflash.com/seattle/2009/09/speech\\_recognition\\_startup\\_voicebox\\_lands\\_13m\\_looks\\_for\\_more.html](http://www.techflash.com/seattle/2009/09/speech_recognition_startup_voicebox_lands_13m_looks_for_more.html)

<sup>94</sup> [http://money.cnn.com/magazines/business2/business2\\_archive/2007/02/01/8398978/index.htm](http://money.cnn.com/magazines/business2/business2_archive/2007/02/01/8398978/index.htm)

<sup>95</sup> 例えば、Toyota、Pioneer、MAGELLAN、RENAULT、HUGHES TELEMATICS、Sirius XM Radio、Johnson Control など。

<http://www.voiceboxtechnologies.com/pressroom/releases/release-5.php>、

<http://www.voicebox.com/pressroom/releases/release-6.php>

<sup>96</sup> <http://voicebox.com/products/iphone.php>

<sup>97</sup> <http://itunes.apple.com/us/app/id295878175?mt=8>、

<http://itunes.apple.com/us/app/id318955292?mt=8>

<sup>98</sup> <http://siri.com/>

<sup>99</sup> 2008年10月に、Morgenthaler Ventures、Menlo Ventures から、8.5 百万ドル、また、2009年11月にも、両社から、15.5 百万ドルの資金提供を受けている。

[http://www.readwriteweb.com/archives/semantic\\_stealth\\_startup\\_siric.php](http://www.readwriteweb.com/archives/semantic_stealth_startup_siric.php)

<http://venturebeat.com/2009/11/24/siri-raises-15-5m-more-for-virtual-personal-assistant/>

<sup>100</sup> <http://www.fonixspeech.com/>

	<p>社。</p> <ul style="list-style-type: none"> <li>音声コマンドアプリに加え、Speech-to-TextのAPI開発のためのSDKであるFonixTalkを販売。</li> </ul>	<p>Windows Mobile向けソフトのFonix VoiceDial (15.95ドル)<sup>102</sup>。</p> <ul style="list-style-type: none"> <li>英語、ドイツ語、イタリア語、フランス語、スペイン語に対応。</li> </ul>
<p>Yap<sup>103</sup> 2006年設立 Charlotte, NC</p>	<ul style="list-style-type: none"> <li>North Carolina州の2007年Early Stage Companyに選出、また、Tech CrunchのイベントのFinalistなど、設立当初から注目を集める。</li> <li>音声認識技術を利用し、Speech-to-Textプラットフォームを開発。</li> </ul>	<ul style="list-style-type: none"> <li>Yap自体ではスマートフォン用アプリを発表していないものの、iPhoneやBlackberryのアプリ開発に必要なSDKを提供<sup>104</sup>。</li> </ul>
<p>Shazam Entertainment<sup>105</sup> 2002年設立 (音楽検索) London, 英国</p>	<ul style="list-style-type: none"> <li>Shazam<sup>106</sup>とは、楽曲認識アプリケーション。マイク経由で音楽をインプットし、楽曲のタイトルやアーティスト名を検索することが可能<sup>107</sup>。</li> <li>なお同社の音楽認識技術は2005年にBMIに売却されたが<sup>108</sup>、BMIから引き続きライセンスを受けている。</li> </ul>	<ul style="list-style-type: none"> <li>現在、無料版と有料版(Twitterなどと連動。Shazam Encore<sup>109</sup>)の2種類が用意されている。</li> <li>両バージョン共に、現在、英語、ドイツ語、日本語などの8ヶ国語に対応しており、22カ国で提供。</li> </ul>
<p>Melodis社<sup>110</sup> 2005年設立 San Jose, CA (音楽検索)</p>	<ul style="list-style-type: none"> <li>同社のSoundHoundは、歌や鼻歌などで楽曲を検索できる。</li> <li>具体的には、①歌、鼻歌、②探している歌手や曲名を吹き込む、③楽曲のタイトルや曲名を打ち込む、④流れている音楽をインプットする(Shazamと同様の仕組み)、という4つの方法で楽曲を検索可能<sup>111</sup>。</li> </ul>	<ul style="list-style-type: none"> <li>2010年2月、iPhone用音声認識アプリとして4.99ドルで販売開始<sup>112</sup></li> <li>なお、同アプリの前身は、2008年に発表されたMidomi(歌および鼻歌による楽曲検索のみに対応。当時は無料<sup>113</sup>。)</li> </ul>

なお、大手である、Nuance Communicationsも、2009年12月8日、音声によるテキスト入力を実現するiPhone向けのアプリケーションDragon Dictation<sup>114</sup>を発表し、無料での提供を開始している<sup>115</sup>。

<sup>101</sup> <http://www.fonixspeech.com/ispeak.php>

<sup>102</sup> [http://www.fonixspeech.com/mp\\_voicedial21.php](http://www.fonixspeech.com/mp_voicedial21.php)

<sup>103</sup> <http://yapme.com/>

<sup>104</sup> 同社の技術を利用したアプリの例は、ReQall、IM+、MyCaptionなど。

[http://www.regall.com/about/regall\\_iphone\\_native](http://www.regall.com/about/regall_iphone_native)、<http://www.mycaption.com/home/product>、

<http://www.shapeservices.com/en/products/details.php?product=im&platform=none>

<sup>105</sup> <http://www.shazam.com/>

<sup>106</sup> <http://www.shazam.com/music/web/pages/iphone.html>

<sup>107</sup> <http://www.crunchbase.com/company/shazam>

<sup>108</sup> <http://www.bmi.com/press/entry/534559>

<sup>109</sup> <http://www.shazam.com/music/web/newsdetail.html?nid=NEWS088>

2009年の11月9日にApp Storeに登録。

<sup>110</sup> <http://www.melodis.com/>

<sup>111</sup> [http://download.cnet.com/SoundHound-for-iPhone/3000-2170\\_4-10863038.html](http://download.cnet.com/SoundHound-for-iPhone/3000-2170_4-10863038.html)

<sup>112</sup> [http://download.cnet.com/SoundHound-for-iPhone/3010-2170\\_4-11239451.html](http://download.cnet.com/SoundHound-for-iPhone/3010-2170_4-11239451.html)

<sup>113</sup> [http://www.wired.com/listening\\_post/2008/07/midomi-iphone-a/](http://www.wired.com/listening_post/2008/07/midomi-iphone-a/)

<sup>114</sup> <http://ax.itunes.apple.com/us/app/dragon-dictation/id341446764?mt=8>

<Googleによるモバイル向けアプリケーション>

Googleでは、自社で開発した音声認識技術を利用して、これまで GOOG-411 などのサービスを展開してきた<sup>116</sup>が、最近においては、GoogleのAndroid系のスマートフォンに音声検索機能を（無料）標準搭載する等の取り組みを進めている。

Googleの音声認識関連サービス<sup>117</sup>

サービス	概要
GOOG-411 (2007年)	<ul style="list-style-type: none"> <li>・音声認識技術を利用したGoogle独自の無料電話番号案内サービス。</li> <li>・電話をかけて、自動音声による質問（ロケーションや事業タイプなど）に答える形で電話番号の検索が行え、目的とする電話をかけることができるというもの。</li> <li>・Google-411は、自社開発音声認識の精度を上げるために行われたプロジェクトであり、同サービスで得られる多くの種類の音声データが、同社が開発する音声認識力の向上に際し、重要な役割を担っているとしている<sup>118</sup></li> </ul>
Google Mobile Apps (2008年11月)	<ul style="list-style-type: none"> <li>・Googleのモバイル用アプリケーションの一つに、音声検索機能（Voice Search）を導入（当初はiPhone向けとして配布）。</li> <li>・音声検索では、音声を入力することによって、検索等各種のサービスを行うことができる。日本語版は、2009年12月に発表<sup>119</sup></li> <li>・現在では、GoogleのAndroid系のスマートフォンに標準装備。</li> </ul>
Google Voice <sup>120</sup> (2009年3月)	<ul style="list-style-type: none"> <li>・Google Voice（2009年3月発表）では、音声ベースのボイスメールをテキストに変換した上で通知する機能を有する。</li> <li>・具体的には、サーバー上の留守番電話への録音メッセージをテキストに変換し、電子メールやショートメッセージとして転送したり、Google Voiceアプリケーションインターフェースからテキストで確認したりできる。</li> </ul>

② 自動車内での音声認識の活用とマイクロソフト

<自動車内での音声認識技術の利用>

近年、自動車などの移動機器に通信装置を取り付け、どこでもリアルタイムで情報サービスを提供するというサービス（テレマティクス：Telematics）が注目されているが、そのような中で、音声認識技術を利用して、ユーザーが音声で操

<http://www.labnol.org/software/iphone-voice-recognition-software/11518/>

<sup>115</sup>同アプリを利用して作成したテキストは、メールやSMSの作成文書として利用したり、クリップボードに貼り付けて他のアプリで使用したりすることもできる。現在の対応言語は英語のみであり、米国のみで展開。なお、音声認識から文書への変換は、Nuanceのサーバー上で実行される。

<sup>116</sup> <http://blogs.zdnet.com/Google/?p=852>

<sup>117</sup> 出典：各種資料より筆者作成。

<sup>118</sup>同社のSearch Products & User Experience部門の副社長であるMarissa Mayer氏。

<http://www.infoworld.com/t/data-management/google-wants-your-phonemes-539?page=0,0>

<sup>119</sup> <http://plusd.itmedia.co.jp/mobile/articles/0912/07/news089.html>

<sup>120</sup> Google Voiceでは、ユーザーに専用の電話番号が無料で与えられ、ユーザーはこの番号を利用して各種サービスを利用する。米国内の通話においては、自らのGoogle Voice番号に電話をかけ、そこから無料で通話することもできる(国内)。

作できる機能が追加されてきている<sup>121</sup>。本分野では、大手では、Microsoft<sup>122</sup>が積極的に取り組んでいる。

自動車分野における音声認識技術の活用<sup>123</sup>

製品企業	概要	プラットフォーム
SYNC (Ford)	<ul style="list-style-type: none"> <li>・ Ford は 2007 年の自動車ショーにて、自動車向けエンターテインメント・情報システムである Ford SYNC を発表。ハンドフリー電話等 9 種類の機能を装備<sup>124</sup>。</li> <li>・ これらの機能の一部は音声認識技術を利用しており、ドライバーは車内から音声コマンドを通して、モバイル端末やカーナビなどの自動車システムを操作する事ができる。<sup>125</sup></li> </ul>	<ul style="list-style-type: none"> <li>・ Microsoft Auto プラットフォーム<sup>126</sup> (Microsoft に買収された TellMe とのパートナーシップ) の上に構築。</li> </ul>
UVO (Kia)	<ul style="list-style-type: none"> <li>・ Kia Motors は 2010 年 1 月に開催された CES にて、SYNC と同様の機能を持つ UVO システムを発表<sup>127</sup>。</li> <li>・ UVO の音声認識機能では、同時に 2 人までの声を識別、各利用者 (トレーニングを受ける人物) 音声コマンド認識をカスタマイズすることができるほか、認識できる言語も数種類に渡っており、使用言語もユーザー毎に管理できるといった機能もある。</li> <li>・ 同製品は 2010 年夏に発売開始される Kia Sprento (2011 年モデル) に搭載される予定<sup>128</sup>。</li> </ul>	<ul style="list-style-type: none"> <li>・ Microsoft の Windows embedded auto Software をプラットフォームとして利用。</li> <li>・ ただし具体的な機能面については未発表。</li> </ul>

<Microsoft の動き>

Microsoft では、従来から自社内で音声認識技術の研究を行っており、これまでに数々の論文を発表している他<sup>129</sup>、音声認識技術を利用した製品も発表してきた。また、2007 年 3 月には、当時音声認識技術市場の有力企業であった Tellme Networks を買収<sup>130</sup>することによって、音声認識分野における取り組みを拡大して

<sup>121</sup> [http://www.onstar.com/us\\_english/jsp/plans/tbt.jsp?seo=goo | 2008 OnStar Upfront | OnStar Telematics | Telematics | telematics](http://www.onstar.com/us_english/jsp/plans/tbt.jsp?seo=goo | 2008 OnStar Upfront | OnStar Telematics | Telematics | telematics)

<sup>122</sup> <http://www.microsoft.com/presspass/press/2006/jun06/06-22TelematicsAwardPR.msp>

<sup>123</sup> 出典: 各種資料より筆者作成。

<sup>124</sup> ハンズフリー通話、911 アシスト、ナビゲーション、音楽検索、自動車安全レポート、ニュース・天気・スポーツ情報、ビジネス検索、リアルタイム渋滞情報、音声ショートメッセージ。

<http://www.microsoft.com/auto/ford.msp>

[http://informationweek.in/Hardware/10-01-12/CES\\_Ford\\_Rewires\\_Cars\\_for\\_Connectivity.aspx](http://informationweek.in/Hardware/10-01-12/CES_Ford_Rewires_Cars_for_Connectivity.aspx)

<sup>125</sup>例えば、SYNC では、ドライバーが車内に入った時点で携帯電話と接続し、ドライバーの携帯電話がテキストメッセージを受信した際、内容を音声で読み上げたり、また、音声コマンドで情報サービスにアクセスし、ニュースや交通情報などにリアルタイムでアクセスすることも可能。

<http://www.fordvehicles.com/innovation/sync/features/#page=Feature4>

<sup>126</sup> 同プラットフォームの技術の一部に TellMe が含まれている。

<sup>127</sup> <http://www.4wheelsnews.com/kia-uvo-the-best-competitor-to-fords-sync-technology/>

<sup>128</sup> <http://www.4wheelsnews.com/kia-officially-unveils-its-uvo-infotainment-system/>

<sup>129</sup> <http://research.microsoft.com/en-us/groups/srg/>

<sup>130</sup>TellMe は、1999 年に設立された企業で、音声認識技術を利用した番号案内サービスを提供。具体的には、携帯電話からの音声入力に対して、テキストメッセージで回答を返す「Speech Self-Service

おり、Microsoft Office、家庭用・自動車用アプリケーション、携帯電話向けの音声認識ソリューションを開発している<sup>131</sup>。

Microsoft の音声認識技術を利用したサービス<sup>132</sup>

サービス	概要
Windows Speech Recognition	音声タイピングやボイス・コマンドなどの機能を持つ Windows 向けの音声認識ソリューション。Windows7、Windows Vista 向け <sup>133</sup> に発表
Microsoft Auto	自動車向けの音声認識プラットフォーム（Ford の SYNC などに搭載）
Tellme Voice Services	音声案内サービス（1-800-555-TELL）（TellMe の技術を利用 <sup>134</sup> ）
Tellme Mobile <sup>135</sup>	携帯電話向けの音声認識機能（TellMe の技術を利用）

#### 4. 画像認識技術の主要企業とアプリケーションを巡る動向

##### （1）画像認識技術の歴史と市場・産業構造

画像認識技術も歴史は古いものの、音声認識技術と比べると、比較的新しい技術である。同技術についても、主に政府（主に諜報機関、軍関係）の研究開発や、大学での研究開発が起源となり、その後、大学等からのスピンオフが、主に政府公安向け等を中心のビジネスを展開し、合併等を進めることによって規模を拡大しつつあるものの、引き続き多くのベンチャー企業が存在する。

##### ① 顔認識技術の歴史

顔認識技術の開発は、米国の諜報機関の依頼を受け、パターン認識を研究していた研究者らが 1964～65 年にかけて取り組んだ研究が始まりである<sup>136</sup>。

---

Platforms」と呼ばれるサービスを提供。2006 年の時点で年商 1 億ドルの事業規模を誇るなど、音声認識市場においては草分け的なベンチャー企業。調査会社である Forrester Research が 2006 年 8 月に発表した市場調査でも、同社は市場のリーダーとの評価。2007 年 3 月に Microsoft に買収された。

[http://www.businessweek.com/technology/content/mar2006/tc20060313\\_913136.htm?chan=technology\\_technology+index+page](http://www.businessweek.com/technology/content/mar2006/tc20060313_913136.htm?chan=technology_technology+index+page)、<http://www.msnbc.msn.com/id/17611739/http://www.tellme.com/files/forrester/ForresterWave.pdf>

<sup>131</sup> <http://www.microsoft.com/presspass/press/2007/mar07/03-14PowerOfSpeechPR.msp>

<sup>132</sup> 各種資料より筆者作成。

<sup>133</sup> 対応言語は英語（英米）、ドイツ語、フランス語、スペイン語、日本語、中国語（繁体字、簡体字）。

<sup>134</sup> <http://www.microsoft.com/speech/technology.aspx#none>

<sup>135</sup> 携帯電話向けの音声認識機能は、まず Blackberry 向けに発表され、その後 Windows Mobile 版が登場している。

<sup>136</sup> 初期の技術は、技術者が写真から特徴を抽出し、ペンタブレットで顔写真にある瞳孔の中心や額の中心などの特徴点をコンピューターに入力し、座標軸に変換して、目の間隔や鼻と口など、顔のパーツ間の距離や比率を計算し、その結果を大規模な画像データベースと照らし合わせるという、非常に手間のかかるマニュアル的手法であった。<http://vismod.media.mit.edu/tech-reports/TR-516/node7.html>

その後、同分野における研究はスタンフォード大学や MIT といった大学研究機関にも広がり、1990 年代以降、同技術は目覚ましい発展を見せた。この背景には、1988 年に線状判別分析 (LDA) が顔認識のアルゴリズム応用されるようになったことに加え、カーネギーメロン大学 (CMU) やメリーランド大学 (UMD<sup>137</sup>) といった学術研究機関による研究の結果、認識・認証の精度が格段に向上した。

1993 年には、DOD の Counterdrug Technology Development Program Office が国防・諜報・法執行機関に利用されうる顔認証システムの開発に向けた FERET プログラム (Face Recognition Technology) を設立し、ここでは、1994~96 年の各年に渡り、顔認証システムの技術評価が行われた<sup>138</sup>。その後、FERET の取り組みは 2000 年に人物認識ベンダー試験 (Face Recognition Vendor Test : FRVT) に引き継がれ、近年、人物認識技術は大きく進歩してきている<sup>139</sup>。

また、そのような中、2000 年代以降、画像認識技術を利用したアプリケーションやベンチャー企業が登場するようになり、特に 2001 年のテロ事件を受け、セキュリティ強化の必要が高まったことから、主にセキュリティ面への利用や個人認証に使われるようになった。また、最近では、セキュリティ以外にも、デジタル写真に写った人の顔が誰のものであるかを認証する機能がウェブアルバムに搭載されるなど、顔認識技術は徐々に身近なものになりつつある。

## ② 顔認識技術に係る主要企業と市場規模

### <主要企業とベンチャー企業>

画像認識技術を利用した製品の大手ベンダーとしては、L-1 Identity Solutions<sup>140</sup> (本社、Stamford, CT) があげられる。

同社は、L-1 Investment Partners 社の主導のもと、2006 年 1 月、顔認証ソリューション提供企業の Viisage (1993 年マサチューセッツ州に設立) と、指紋などの生体認証システムのベンダーである Identix<sup>141</sup>との合併によって設立された、個人認証技術を総合的に扱うソリューション企業である。なお、同社は設立以降、6 つほどのベンチャー企業を買収している<sup>142</sup>。

<sup>137</sup> <http://www.face-rec.org/research-groups/#USA>

<sup>138</sup> <http://www.frvt.org/FERET/default.htm>、<http://www.frvt.org/default.htm>

<sup>139</sup> 2000、2002 年、2006 年の 3 回に渡り、人物認識技術研究のスポンサーや人物認識技術のパフォーマンス評価が実施された。2000 年のソフトウェア比較では、人物認識関連ソフトの性能に懸念が示されたが、2006 年の試験結果では、静止画像、3 次元、虹彩比較分野で新たに開発されたシステムにより、人物認識技術が大きく進歩したことが判明している。Face Recognition Vendor Test 2006 Report

<http://www.frvt.org/>、<http://www.frvt.org/FRVT2006/default.aspx>

<http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf>

参加した機関は、大学研究所や民間企業など全 22 機関。<http://www.frvt.org/FRVT2006/>

<sup>140</sup> <http://www.l1id.com/>

<sup>141</sup> Identix 社は、1982 年にカリフォルニア州に設立された、指紋認証 (指紋採取、指紋データの分析など) に関する技術の開発・販売を行う企業。<http://www.answers.com/topic/identix-inc>

<sup>142</sup> <http://www.l1id.com/pages/204-history>



同社は主に政府向けソリューションを主力サービスとしており、現在、米国パスポートの全てと、米国で発行される運転免許証の85%が、同社の製品によって作られているとしている<sup>143</sup>。同社の2009年の売上は34.2億ドル（従業員数は2200人強）で、このうち、顔認証分野での同社の売上は、全体の約12%の約4億ドルとなっている。なお同社は、顔認証分野の売上は、2014年には14億ドル（全体の15%程度）に拡大すると予測している<sup>144</sup>。

同社は、顔認証の分野では、2D、3D、肌認識技術が盛り込まれた顔認識ソフトウェア開発キットである Facelt SDK や、顔認識ソフトウェアの評価ソフト開発ツールである Facelt Quality Assessment SDK を提供しており、これらは、オクラホマ州政府やフロリダ州の地方政府で利用されている。

また、同社以外にも、近年、公安・セキュリティ向け分野等を中心に、多くのベンチャー企業が生まれている。

顔認識技術に係るベンチャー企業（例）<sup>145</sup>。

企業	概要	製品
Pittsburg Pattern Recognition <sup>146</sup> 2004年設立	<ul style="list-style-type: none"> <li>顔を含む対象物（物体）を認識する技術の開発企業。カーネギーメロン大学の Henry Schneiderman 博士がスピンオフして設立。</li> <li>同社は、顔認識関連の製品とて、ソフトウェアの開発キットを販売。</li> </ul>	<ul style="list-style-type: none"> <li>Face Tracking SDK は、顔を検知・追跡すると共に、目や鼻など、顔としての特徴物を検知するというもの。</li> <li>Face Recognition SDK は、上記の機能に、顔認証の機能を加えたもの。</li> </ul>
3VR Security <sup>147</sup> 2002年設立 <sup>148</sup> San Francisco, CA	<ul style="list-style-type: none"> <li>Inktomi、USWeb、TiVo の3企業の元幹部により設立。</li> <li>インターネット上での検索技術を応用し、政府、行政機関、民間企業で撮影された膨大なビデオデータから特定の人物を検索する製品を開発。</li> </ul>	<ul style="list-style-type: none"> <li>同社製品である SmartRecorder では、①顔データを集積・データベース化し特定の容疑者を追跡可能、また、②要注意人物リストを構築・管理し、リスト化された人物が撮影された場合、通知することも可能。</li> </ul>
Animetrics <sup>149</sup>	<ul style="list-style-type: none"> <li>次世代の2D to 3Dの顔認識技術の</li> </ul>	<ul style="list-style-type: none"> <li>同社製品の1つである Animetrics90</li> </ul>

[http://files.shareholder.com/downloads/LONE/0x0x50260/d80d8409-e55b-4528-9d6e-41b0854066e1/L1ID\\_FAQs\\_shareholder.pdf](http://files.shareholder.com/downloads/LONE/0x0x50260/d80d8409-e55b-4528-9d6e-41b0854066e1/L1ID_FAQs_shareholder.pdf)

<sup>143</sup>2010年3月17日の Westport News 誌の報道

<http://www.westport-news.com/news/article/Terrorist-attacks-spawn-idea-for-L-1-Identity-409892.php>

<sup>144</sup>[http://files.shareholder.com/downloads/LONE/869652878x0x349971/e50ce964-3fae-4599-b1b9-d5e4359ab332/Corporate\\_Presentation\\_February\\_2010pdf.pdf](http://files.shareholder.com/downloads/LONE/869652878x0x349971/e50ce964-3fae-4599-b1b9-d5e4359ab332/Corporate_Presentation_February_2010pdf.pdf)

<sup>145</sup> 出典：<http://www.frvt.org/FRVT2006/default.aspx> や各種資料を参考にしつつ、筆者作成。

<sup>146</sup> <http://www.pittpatt.com/>

<sup>147</sup> <http://www.3vr.com/>

<http://www.altassets.net/private-equity-news/by-pe-sector/venturegrowth-capital/article/nz16580.html>

<sup>148</sup> 同社ウェブサイトによると、Klein Perkins Caulfield & Byers (KPCB)、DAG Ventures、Menlo Ventures、In-Q-Tel、Vantage Point Venture などが、同社に資金を提供している。

<sup>149</sup> <http://www.animetrics.com/About/overview.php>, <http://www.animetrics.com/Products/FACER.php>

<p>2002年設立 Conway, NH</p>	<p>開発企業。 ・同社の技術は、米国の軍や諜報機関等で利用されている。</p>	<p>では、2D、3Dの顔認識技術を利用し、同社が保有するFIMSシステムのデータベースと照合することで、監視カメラに写った人物の特定やログイン管理などを行う。</p>
<p>Cognitec Systems<sup>150</sup> 2002年設立 Dresden, 独</p>	<p>・顔認識による生体認証技術を専門とする企業。 ・米国、ブラジル、香港にも支店を持ち、世界展開。</p>	<p>・認識を行うためのデータベース、ビデオに写った人物の認識、顔認証によるログイン管理システム、無人の入国審査にも利用可能な顔認識システムを提供。</p>

### <顔認識技術の市場規模>

顔認識技術に係る市場規模は、現時点では音声認識技術や、他の生体認証技術よりも小さいものの、今後拡大が見込まれている。

例えば、調査会社 Frost & Sullivan の調査報告書（2009年6月頃）<sup>151</sup>によると、人物の照合に利用される顔認識技術市場の規模は、2008年に1.86億ドルであるが、その後年間27.5%で成長し、2012年には10億ドルに達する見込みであるとしていると報告されている。

また、ABI社が発表した2008年の世界における生体認証全体の市場調査結果（2009年11月）<sup>152</sup>によると、全体で5.61億ドルであり、このうち、指紋認証が全体の60%以上の3.49億ドルを占め、次に多く利用されているのが顔認証の0.92億ドル（16.4%）である<sup>153</sup>。また、同社によると、顔認証市場は2008～13年にかけて22%成長し、2.51億ドルに成長するとしている。

### （2）画像認識技術分野における企業向け主要アプリケーションを巡る状況

画像認識技術は、これまで、指紋や虹彩などをもとにした個人認証と同様、主に、公安面（犯罪者の捜査、行方不明者の捜索、カジノ等でのブラックリスト人物の抽出<sup>154</sup>等）、パスポートの管理（入国審査等）、アクセスセキュリティシステム（社員の出退勤の管理<sup>155</sup>等）などを中心に利用されてきている。

このうち、公安面での管理等については、2000年代前半においては、必ずしも十分機能を果たせなかった事例も少なくないが、現在では、最近の技術の進展に

<sup>150</sup> <http://www.cognitec-systems.de/>、<http://www.cognitec-systems.de/Products.12.0.html>

<sup>151</sup> SecurityInfoWatch.com 誌の2009年6月2日の記事。

<http://www.securityinfowatch.com/Biometrics/facial-recognition-looks-future>

<sup>152</sup> XID Technologies 社の発表（2009年10月21日付）の引用より。

[http://www.xidtech.com/our\\_news\\_2009.htm](http://www.xidtech.com/our_news_2009.htm)

<sup>153</sup> 次いで、虹彩が0.47億ドル。

<sup>154</sup> Griffin Investigation など。ブラックリストに載っている人物を識別する。

<sup>155</sup> <http://www.iqt.org/technology-portfolio/a4vision.html>

Bio Script 社による A4Vision など。

に伴い、セキュリティの強化が重視される場所を中心とした試験導入例が、世界的に再度増加しつつある<sup>156</sup>。以下、米国におけるいくつかの事例を、アクセス・セキュリティでの利用も含めて、紹介する<sup>157</sup>。

米国における顔認証技術の応用事例<sup>158</sup>

分野	事例
公安面	・2001年1月、スーパーボウルの会場でセキュリティシステムとして試験的に利用 <sup>159</sup> 。

<sup>156</sup> <http://www.aclu.org/technology-and-liberty/ga-face-recognition>

<sup>157</sup> なお、海外の導入事例は、以下の通り。

<公安面(パスポートを含む)>

・ロンドンのニューハム地区で、地区全体の監視カメラに顔認識システムを導入したが、2001年から2004年までで、システムが犯罪者を一人も認識していないとのことである。ただし、認識システムを導入することで、犯罪者の心理面に働きかけ、結果として犯罪率が減少しているとのことである。

<http://www.howstuffworks.com/gadgets/high-tech-gadgets/facial-recognition.htm>

<http://www.aclu.org/technology-and-liberty/ga-face-recognition>

・英国の警察では、治安の確保に向け、国民の顔写真のデータベース構築に着手している。データベースに記録される写真は、街中に設置されている監視カメラの映像と映像と照合される。2008年3月時点では、一部地域(Lancashire、West Yorkshire、Mersyside)で試験導入が完了済みで、2009年には全国的に正式導入される予定であった。<http://www.guardian.co.uk/uk/2008/mar/19/ukcrime.humanrights>

・オーストラリア政府は、国内の国際空港に到着する同国およびニュージーランド市民(18歳以上のeパスポート保持者)が、eパスポートに記録されたデータと顔認証技術を利用し、税関や入国審査の手続きをセルフサービスでできる制度を、2007年8月より正式に開始している。また、ニュージーランド政府も同様の制度を2010年1月から開始。

<http://www.theaustralian.com.au/australian-it/smartgate-passport-check-goes-national/story-e6frgamf-1111115999497>

・南アフリカサッカー協会も、2010年の南アフリカ・ワールドカップにて、制限区域に立ち入る際の入場審査にて、顔認証技術とビデオ監視などを総合した警備体制を導入するとしている。

<http://www.sourcesecurity.com/news/articles/co-4415-ga-co-4468-ga.2988.html>

<アクセス・セキュリティ面>

・ロンドンに拠点を置くHSBC銀行は、2008年11月、同社が英国内に保有する2つのデータセンターに、顔認証技術を活用した警備システムの導入することを決定した。このような取り組みは英国の銀行としては初めて、従業員の入退室を管理するとともに、英国内で急増しているなりすましなどによる情報漏えいの防止に向けたもの。

<http://www.info4security.com/story.asp?sectioncode=10&storycode=4120893&c=1>

・メキシコでは、2000年の大統領選挙時に、別人による投票や1人による複数投票などの不正を防ぐために顔認識ソフトを導入、成果を上げたとのことである。


[http://74.125.113.132/search?q=cache:PTRWz4gqrckJ:www.cs.unc.edu/~pozefsky/COMP380\\_F06/facerecognition2.ppt+face+recognition,+mexican+election&cd=3&hl=en&ct=clnk&gl=us](http://74.125.113.132/search?q=cache:PTRWz4gqrckJ:www.cs.unc.edu/~pozefsky/COMP380_F06/facerecognition2.ppt+face+recognition,+mexican+election&cd=3&hl=en&ct=clnk&gl=us)

<sup>158</sup> 出典:各種資料より作成。

なお、顔認証以外の画像認識の分野としては、例えば、自動車分野において、小型車載カメラと画像認識技術を組み合わせ、危険が迫っていることをドライバーに警告するといった面で、画像認識技術が利用されており、ここでは、白線などの走行路、車両や歩行者などの障害物、道路標識などの交通情報、ドライバー自身の4種類の対象物が認識されている。

<http://e2a.jp/review/081028.shtml>、<http://journal.mycom.co.jp/articles/2009/05/21/aee04/index.html>

<sup>159</sup> <http://www.aclu.org/technology-and-liberty/ga-face-recognition>

	<ul style="list-style-type: none"> <li>・また、2001年の同時多発テロ事件を機に、ボストンのローガン空港、ロードアイランド州プロビデンスのT.F Green 空港、サンフランシスコ国際空港などでも人物認識技術に基づく顔認証システムが導入<sup>160</sup>。</li> <li>・FBI<sup>161</sup>は、顔認証システムを使い、ノースカロライナ州の高速道路で逃亡中の犯罪者を特定するという取り組みを実施。このシステムでは、陸運局で管理されている多数の運転免許証用の写真と、高速道路で撮影された顔写真を照合し、犯罪者と似ている人が発見された場合に、捜査当局が出動するもの。</li> <li>・なお、2009年には、同取り組みを通じて、殺人容疑者を逮捕した実績がある<sup>162</sup></li> <li>・国土安全保障省は2003年、US-VISITと呼ばれる出入国管理プログラムを導入し、非移民ビザやビザ免除プログラムで入国する外国人に対して、入国の際に指紋や顔写真などを提供する義務を課した。</li> <li>・提供されたデータは、入国可否の審査のため、テロリストや犯罪者データベースの情報と照合されるほか、不法滞在者の取り締まりや滞在条件の変更管理などを目的としたデータベースに保存され、関係機関で共有されることとなっている<sup>163</sup>。</li> </ul>
<p>アクセス・セキュリティ</p>	<ul style="list-style-type: none"> <li>・建物の出入り監視では、従業員の出入りに関し、顔認証によって扉を施錠するというシステムがある。</li> <li>・例えば、Chinavisionが開発したシステムでは、2台のカメラを使って作成される3D画像を、入り口の施錠開錠に利用<sup>164</sup>。現状では500人を登録可能であるが、将来的には15万人まで拡大することを目標。</li> </ul>  <ul style="list-style-type: none"> <li>・コンピューターへのログイン時の利用例としては、Lenovoのノートパソコンが挙げられる。同製品では、顔認証をログインパスワード代わりに使用しており、複数の人物がコンピューターを共用している際には、コンピューターの前に座るだけで、適切なアカウントにログインすることができるようになっている<sup>165</sup>。</li> </ul>
<p>その他（デジタルサイネージ）</p>	<ul style="list-style-type: none"> <li>・米国サムソン電子は、自社のアウトオブホームデジタルサイネージサービスであるSamsung PROMにおいて、TVの視聴者の規模や性別・年齢層などのデータを、市場調査会社ニールセンにリアルタイムで送り、デジタルディスプレイ前を通行する視聴者のデモグラフィックに合わせて広告を変更できるサービスを提供すると発表した<sup>166</sup></li> <li>・同サービスでは、CMを見ている視聴者の顔を検知してその動きを追うことができ、視聴者層の情報により配信内容を常時変更できるようにしている。</li> </ul>

<sup>160</sup> ただし、ボストン、ローガン空港でのテロリスト識別実験や2001年フロリダ州タンパ市警察での実験では、顔認識システムの成果は上がっていないとする結果が出されている。

<http://www.howstuffworks.com/gadgets/high-tech-gadgets/facial-recognition.htm>

<http://www.aclu.org/technology-and-liberty/qa-face-recognition>

<sup>161</sup> FBIは、2001年9月にViisage(現L-1)から顔認識技術の無料提供のオファーを受けている。

<http://www.allbusiness.com/crime-law-enforcement-corrections/law-biometrics/6174516-1.html>

<sup>162</sup> <http://www.popsci.com/technology/article/2009-10/fbi-facial-recognition-software-scan-highway-fugitive-motorists>、<http://www.fbi.gov/thisweek/archive/thisweek070209.htm>

<sup>163</sup> <http://www.ice.gov/pi/news/factsheets/visit051903.htm>

<sup>164</sup> <http://www.geek.com/articles/gadgets/face-recognition-security-system-keeps-robbers-but-not-doppelgangers-out-20091211/>

<sup>165</sup> <http://lenovoblogs.com/insidethebox/?p=132>

なお、過去には、IBMがThinkPad T23のモデルで同様の機能を搭載したが、当時のカメラの仕様が十分ではなかったなどの理由から、この機能はあまり利用されなかった。しかし、現在のLenovoの製品では、内蔵カメラの画像が向上したため、認識の信頼性が高まっており、幅広い利用が期待されている。

<sup>166</sup> <http://trumedia.co.il/inside.asp?ID=73>

(3) 画像認識技術を利用した消費者向け・モバイルアプリケーション

一方、最近においては、画像認識分野においても、消費者向けオンラインサービスへの利用が急速に進みつつあるほか、モバイルアプリケーションとして、クラウド上のデータベース活用し、顔認識以外の分野での画像認識技術のアプリケーションが登場しつつある。

① 消費者向けオンライン画像サービス

消費者向けのオンライン画像サービスとしては、各インターネット企業が、ベンチャー企業の技術等を吸収・活用することにより、各種のウェブアルバムサービスに対して、顔であることの認識技術から始まり、最近では個人を特定(分類)する認識技術も含めて、顔認識技術の導入が進みつつある。

オンライン画像サービスにおける顔認識技術の利用<sup>167</sup>

企業	概要
Flickr <sup>168</sup> (Yahoo!)	<ul style="list-style-type: none"> <li>・ Flickrは、2009年4月、これまでのウェブアルバムサービスに、新しく顔認識機能を追加。同機能には、スウェーデンの新興企業である Polar Rose<sup>169</sup>が開発した技術が使われており、Flickrのユーザーは、写真に写っている人物を人物別に分類することができる。</li> <li>・ また、ユーザーが写真にその人物の名前を入力すると、写真内にデジタルラベルが貼られて顔の横に表示され、Flickrのアカウントにある写真にも反映される。</li> </ul>
Picasa (Google)	<ul style="list-style-type: none"> <li>・ GoogleのPicasaのうち、ウェブアルバム版には2008年9月に顔認証機能が追加。また、デスクトップ版については、2009年9月、バージョン3.5へのアップデートに伴い、顔認証機能が追加された<sup>170</sup>。</li> <li>・ Picasa3.5では、画像データに写真に写っている人物の名前のタグを付けると、ユーザーのフォトライブラリ全体を調べ、同じ人物が移っている写真に自動的にタグが付く仕組みとなっている<sup>171</sup>。</li> <li>・ 同技術は、2006年にGoogleが買収した Neven Vision<sup>172</sup>の技術に基づいている。</li> </ul>

<sup>167</sup> 出典:各種資料より筆者作成。

<sup>168</sup> 2005年3月、Yahoo!が買収。

<sup>169</sup> 2004年にスウェーデンの数学者である Jan Erik Solem 氏によって設立された企業。同社は今後、様々なウェブサイトに Flickr に提供したのと同様の技術を供与し、同サービスを拡大するとしている。

<sup>170</sup> <http://techcrunch.com/2009/09/22/picasa-adds-facial-recognition-and-geo-tagging-to-its-desktop-app/>、<http://techcrunch.com/2008/09/02/picasa-refresh-brings-facial-recognition/>  
<http://picasa.google.com/support/bin/answer.py?hl=en&answer=53209>

<sup>171</sup> 同じ顔かどうかははっきりしない時は、タグを付けるかどうかをユーザーが選択する事ができる。また、タグが付けられた顔写真ではその人物の顔写真のコラージュや動画を作成でき、Picasa ウェブアルバムとタグを同期することもできる

<sup>172</sup> 同社は、同社は、元々、生体認証技術を取り扱う NevenEngineering を基に、2003年に Naven Vision として設立された企業で、後に生体認証技術の画像認識への応用を通じて成長した。  
<http://blogscoped.com/archive/2006-08-15-n52.html>

Apple	<ul style="list-style-type: none"> <li>Appleでも、MacBookなどのコンピューターに、写真に写っている人物を認識し、人物毎に写真を管理できる機能を搭載している<sup>173</sup>。</li> <li>また、最近では、2009年1月6日には一般ユーザー向け写真管理ソフトウェア「iPhoto09」に、また、2010年2月10日にはプロフェッショナルを対象とした写真管理ソフトウェア「Aperture3」に顔認識技術を統合。</li> <li>これにより、利用者が写真に写っている人に名前を付けると、その情報がデータとして蓄積され、ライブラリー全体から似ている人が写っている写真が検出されるほか、ユーザーは人物毎に写真を管理することができる<sup>174</sup>。</li> </ul>
-------	---

## ② モバイル向け各種画像検索サービス

さらに、近年においては、Google、Amazon等は、上記同様ベンチャー企業の画像認識技術を吸収し技術力を強化するとともに、モバイル端末（スマートフォン）を利用し、顔以外のモノ（対象物）を画像認識・検索するサービスが提供し始めている。これらのサービスの前提として、Google、Amazonは自社内データセンターを多く有し（クラウドサービスも提供）、これらに係る多量の画像データを保有していることがあげられる。

### モバイル向けの顔以外の画像認識・検索サービス<sup>175</sup>

企業	概要
Google	<ul style="list-style-type: none"> <li>Googleは、2009年6月22日、画像内の世界の名所を高精度で見分けられる場所認識アプリケーション Google Goggles を発表。同技術は、イメージ検索、オブジェクト認識、クラスタリング技術を活用し、ウェブ上の膨大な画像の場所をベースにインデックス化するというシステムである<sup>176</sup>。</li> <li>本システムは、2009年12月7日、携帯電話のカメラで撮影した写真で検索できるツールとして、Androidスマートフォン向けにリリースをされている<sup>177</sup>。</li> <li>また、2010年2月には、書籍の表紙やCD等をカメラで撮影した写真で、商品を検索し、販売サイトで購入できる Google Shopper を、同じく Androidスマートフォン向けに公開<sup>178</sup>。</li> </ul>

同社は、i-Scout Visual Mobile Search Neven Vision（携帯電話機のカメラに対象物を認識させ、関連コンテンツをユーザーに提供するプラットフォーム）、Mobile Identifier Platform（携帯機器用の生体認証プラットフォーム（顔認証を含む））、Mobile-I Facial Feature Tracking Software（リアルタイムまたはオフラインで、顔の特徴を追跡しながらビデオに映し出す）などのサービスを有していた。

<http://www.linkedin.com/companies/neven-vision>

<sup>173</sup> また GPS 機能とジオタグ (Geotag) もサポートしているため、iPhone や GPS 内臓を内蔵した機器から写真を読み込む際に、撮影場所のデータを追加し、撮影場所に応じた写真管理や検索を行うこともできる。

<sup>174</sup> <http://japan.cnet.com/news/tech/story/0,2000056025,20408339,00.htm>

<http://www.apple.com/pr/library/2009/01/06ilife.html>

<http://www.apple.com/aperture/whats-new.html>

<sup>175</sup> 出典：各種資料より作成。また、NY だより 2009 年 12 月号参照。

<sup>176</sup> 同機能では、写真に写った名所がデータベースに記録されている 5 万箇所に相当すれば、80% の精度で特定できるとしている。なお、発表当時は研究段階

<sup>177</sup> <http://www.itmedia.co.jp/news/articles/0912/08/news027.html>

	<ul style="list-style-type: none"> <li>・なお、2010年4月12日、Googleは、英国のベンチャー企業（創業4ヵ月目）でモバイル向け画像検索技術を手がけているPlinkを買収した<sup>179</sup>。同社が市場に出している製品は、モバイル端末で有名な芸術作品の情報を検索するアプリケーションPlinkArtであり、Gogleは、上記Google Gogglesに同技術を統合する予定としている。（また、開発を予定していたiPhoneアプリは中止することになる。）</li> </ul>
Amazon	<ul style="list-style-type: none"> <li>・Amazon.comは、店頭で見かけた商品をBlackberryのカメラで撮影すると、似たような商品をAmazonの商品リストから検索してくる、Amazon Remembers：Blackberry向けのAmazon App）がある<sup>180</sup>。</li> <li>・なお、同技術は、もともと画像認識技術を活用したモバイルマーケティング企業のSnap Tell（2006年設立、Palo Alto, CA）によるもの。同社は、iPhone向けやAndorid向けのソフトを提供していた<sup>181</sup>が、2009年6月にAmazonの検索部門子会社であるA9.comが買収<sup>182</sup>。</li> </ul>

なお、本レポートは、注記した参考資料等を利用して作成しているものであり、本レポートの内容に関しては、その有用性、正確性、知的財産権の不侵害等の一切について、執筆者及び執筆者が所属する組織が如何なる保証をするものでもありません。また、本レポートの読者が、本レポート内の情報の利用によって損害を被った場合も、執筆者及び執筆者が所属する組織が如何なる責任を負うものでもありません。

<sup>178</sup> [http://internet.watch.impress.co.jp/docs/news/20100219\\_350127.html](http://internet.watch.impress.co.jp/docs/news/20100219_350127.html)

<sup>179</sup> <http://www.plinkart.com/blog/?p=77>  
[http://news.cnet.com/8301-1023\\_3-20002254-93.html](http://news.cnet.com/8301-1023_3-20002254-93.html)  
<http://japan.internet.com/finanews/20100413/11.html>

なお、これ以外にも、Googleは、2010年3月には、オンライン写真編集サービスのPicnikを買収。  
<http://www.itmedia.co.jp/enterprise/articles/1003/02/news027.html>  
 また、2010年4月には、オンラインビデオプラットフォームを手がけるEpisodicを買収。  
<http://www.itmedia.co.jp/enterprise/articles/1004/05/news012.html>

<sup>180</sup> <http://www.amazon.com/gp/anywhere/sms/bbapp>

<sup>181</sup> <http://www.snaptell.com/apps/>

<sup>182</sup> <http://jp.techcrunch.com/archives/20090616image-recognition-startup-snaptell-acquired-by-amazon-subsiary-a9com/>