

音声認識技術に関する特許出願技術動向調査報告

平成15年5月22日
特許庁総務部技術調査課

第1章 音声認識技術の概要

第1節 音声認識技術とは

人間が自然で容易に機械を使用することを可能とするヒューマンマシンインターフェイスを実現するために、

- ・ 音声認識技術：人間の音声の発声内容を機械により認識し入力する技術
- ・ 音声合成技術：機械から言葉を音声によって出力する技術

が昔から研究されている。

音声認識技術とは、要約1-1表に示される 音響処理技術、音響モデル作成・適応化技術、マッチング・尤度演算技術、言語モデル技術、対話処理技術の5つの要素技術の組み合わせにより実現されるものである。

要約1-1表 音声認識の要素技術

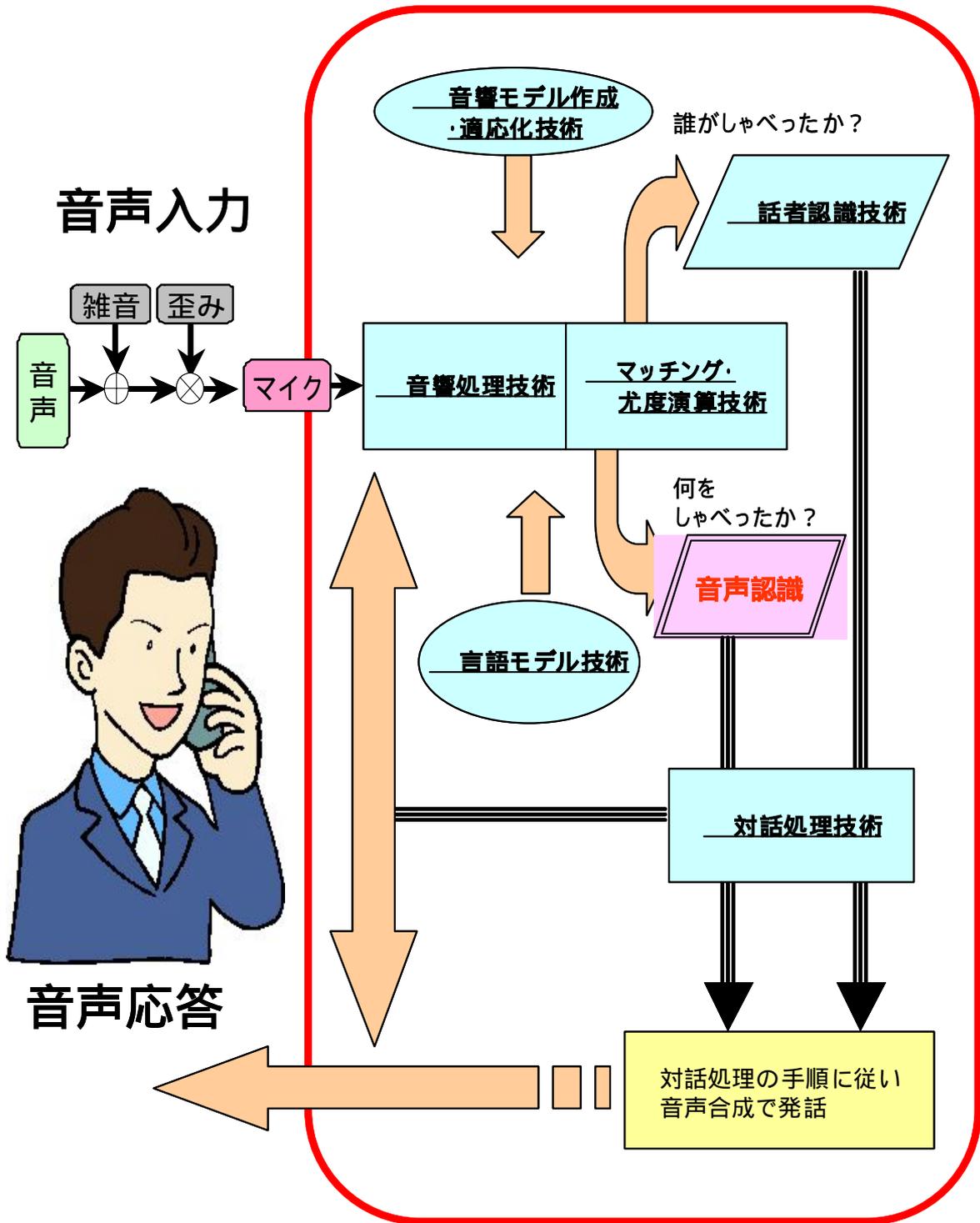
要素技術名	概要
音響処理技術	デジタル化された音声信号から、音響パラメータを抽出する技術。実環境での雑音や周波数歪みを除去する技術、音声発声区間を検出する技術も含む。
音響モデル作成・適応化技術	特定話者、不特定話者の音声を構成する要素（音素、単語など）の音響モデルを作成する技術及び音響モデルを特定話者や環境に適応化する技術。
マッチング・尤度演算技術	音響モデルに対する音響パラメータ系列の距離や尤度（確からしさ）を求め、認識結果を決定する技術。
言語モデル技術	連続する単語間の接続確率や係り受け関係をモデル化する技術。連続単語認識に利用される。異音や単語間のリエゾン、発話様式などの発音変動に対処する技術も含む。
対話処理技術	人間と音声認識装置との間の音声による円滑な対話を実現するための処理技術。音声認識により認識された結果から対話の進行状況を判断し、音声合成などの出力手段を制御して、人間に情報を出力する技術。

要約1-1図は要約1-1表の要素技術から構成される音声認識の構成を示した俯瞰図である。

一方、音声を発声した話者を認識する技術を話者認識（あるいは話者認証）と呼ぶ。話者認識は要約1-1表に示される音声認識の要素技術のうち、音響処理技術、音響モデル作成・適応化技術、マッチング・尤度演算技術について音声認識と技術を共有することが多い。

本調査では、音声認識の要素技術である、音響処理技術、音響モデル作成・適応化技術、マッチング・尤度演算技術、言語モデル技術、対話処理技術に加えて、話者認識技術と、音声認識と話者認識についての 実用化技術を調査対象とした。

要約 1-1 図 音声認識の俯瞰図



第2節 音声認識技術の用途・市場

音声認識技術の産業分野における用途・市場は要約1-2表の通りである。

要約1-2表 音声認識の産業分野

産業分野	用途・市場
A) コマンド制御	組み込み機器(カーナビゲーション、情報家電、ロボットなど)の音声による指示。手や目が塞がった状況(ハンズビジー、アイズビジー)で、機器を制御するニーズから音声認識が利用される。雑音のある実環境での認識性能の向上がポイント。
B) 口述筆記(ディクテーション)	音声からテキストへの自動変換。PCソフトが製品化されている。医療分野での電子カルテのニーズにも使用されている。PDA やカーナビゲーションでのメール作成などの分野でのニーズも顕在化している。話し言葉の認識性能の向上がポイント。
C) データ入力	業務系機器(PDA など)への音声によるデータ入力。
D) 介護/福祉	介護/福祉機器への音声による指示。ハンズビジーな状況や高齢者、身体障害者への支援などのニーズがある。
E) 教育	語学教育における発音チェックや e-Learning における音声利用のニーズが顕在化しつつある。
F) コールセンター	ユーザからの問い合わせや予約などのオペレータ業務を自動化し、人件費を削減するニーズ。米国では、人間による電話受付の人件費5.5ドルを10分の1以下に削減する効果が確認されている。
G) 音声ポータル	音声認識、音声合成を利用したインターネットコンテンツアクセスサービス。コンテンツ開発に関して、Voice XML、SALT などの標準化が進んでいる。
H) 音声ブラウザ	音声認識、音声合成を利用した音声によるインターネットコンテンツアクセス機能を有するマルチメディアブラウザ。コンテンツ開発に関して、VoiceXML、SALT などの標準化が進んでいる。
I) 索引付け	TV プログラム、ビデオカメラ、IC レコーダの音声部を利用した索引付けによる検索自動化。コールセンターへの問い合わせ音声の自動分類というニーズもある。
J) 書き起こし	講演音声などのテキスト書き起こし。
K) 放送	聴覚障害者のためのニュースなどのクローズドキャプション。
L) 自動翻訳	会話を認識し、他の言語に翻訳し、テキスト表示又は音声合成で出力。
M) 話者認識	セキュリティシステムにおける音声での認証。

音声認識の産業分野は実装の形態によって、サーバ型/組み込み型、PC実装型/RISC・DSP実装型、オンライン型/オフライン型に分類することができる。

サーバ型とは、多量の組み込み機器(携帯電話、カーナビゲーションなど)で収録された音声をネットワーク上のサーバに送り、サーバに設置された音声認識装置で一括して認識する方式を言う。組み込み型とは、組み込み機器(携帯電話、カーナビゲーションなど)で収録された音声を組み込み機器内に実装された音声認識ソフトウェアで認識する方式を言う。

オンライン型とは、発声された音声を即座に認識する方式を言い、オフライン型とは、発声された音声を一度保存し、後で保存した音声を認識する方式を言う。

この分類に従って、要約1-2表の用途・市場を分類した結果が要約1-3表である。この表では、日本で既に製品化されている、又は開発中の製品やサービスの例も示しておいた。

要約 1-3 表 音声認識の産業分野の分類

産業分野	サーバ型	組み込み型		オンライン型	オフライン型	製品、サービス例
		PC 実装型	RISC・ DSP 実装型			
A) コマンド制御						ケリオ、パオニアなどのカーナビゲーション
B) ディクテーション						IBM「ViaVoice」 ScanSoft「Dragon Speech」 NEC「SmartVoice」 東芝「LaLaVoice」
C) データ入力						アパシストメディアの画像診断読影レポート入力支援システム「Amivoice Medical」
D) 介護/福祉						旭テクノシステムの音声リコン「LIFETACT」
E) 教育						ベネッセの児童英語教育ソフト「BE-GO (ビゴ)」
F) コールセンター						全日空、JAL、航空、JR 東海などの電話予約システム
G) 音声ポータル						日本テレコム「Voizi」 NTT コミュニケーション「V ポータル」 KDDI (au)「ボイスエージェント」
H) 音声ブラウザ						Microsoft の IE (Internet Explorer) の SALT 対応
I) 索引付け						ScanSoft「Audio Mining」ソフト
J) 書き起こし						VoiceTreck (オジパス)
K) 放送						NHK の字幕スーパー
L) 音声翻訳						ATR の自動音声翻訳電話 日立の韓国での試験サービス NEC の成田空港での実験
M) 話者認識						アコム「VoicePassport」

第 3 節 音声認識技術の発展

音声認識技術の研究の歴史は 1952 年に米国のベル研究所でゼロ交差回数を用いた数字認識の研究が行なわれたことに始まる。これを拡張した単音節認識装置「音声タイプライター」が 1959 年に日本の京都大学において研究された。

その後、1971 年に発声時間の長さの変動を DP (Dynamic Programming : 動的計画法) を用いて非線形に正規化する DTW (Dynamic Time Warping : ダイナミックタイムワーピング) 法が日本とロシアの研究者により提案され、さらに、日本で連続数字を認識できる 2 段 DP マッチング法が提案された。1978 年にはこの 2 段 DP マッチング法を用いた連続単語認識装置が日本で製品化 (日本電気 DP-100) され、荷物の仕分けなど手がふさがっている (ハンズビジー) 状況下でのデータ入力に使われた。

米国では、1970 年代に統計確率的手法である HMM (Hidden Markov Model : 隠れマルコフモ

デル)を用いた音声認識の基礎的な研究が、CMU、IBM、続いて AT&T で行われた。1980 年代を通して、HMM は音声認識の標準的手法になっていった。

また、1980 年代から 1990 年代前半にかけて、人間の神経網の挙動の数学的モデルである NN(Neural Network : ニューラルネットワーク)による音声認識手法も活発に研究された。

1980 年代後半から 1990 年代前半に、米国 DARPA(Defense Advanced Research Project Agency:国防省高等研究計画局)で口述筆記(ディクテーション)プロジェクトが実施された。この中で、文章の言語モデルとして、Nグラムモデル(連続するN語間の統計確率)を用いる方式が提案され、大語彙連続音声認識が実現されるようになった。この成果をベースとして、パーソナルコンピュータ(PC)の中央演算処理装置(CPU)の大幅な性能向上を背景に、口述筆記用大語彙連続音声認識を行うパソコン用のソフトウェア(Dragon Naturally Speaking や IBM ViaVoice)が 1997 年に米国で販売され、日本でも日本語を対象とした口述筆記を行うためのパソコン用ソフトウェアの製品化(日本電気 SmartVoice など)が行なわれた。

一方、1990 年代前半からは、DARPA の手がけた諸プロジェクトで、音声対話システムの研究が行われ、音声対話を制御する対話処理技術の研究が盛んに行われた。その成果を基にして米国では 1996 年に電話による株式情報照会(Charles Schwab & Nuance)や、1998 年に電話による各種問い合わせ・予約サービス(コールセンター)の実用化が行われた。

日本では、カーナビゲーション用の音声認識 LSI が 1995 年に実用化されている。

現在の音声認識の技術レベルでは、利用者の音声の特徴を学習させておけば、明瞭な発声で読み上げた文章をほぼ完全に口述筆記することができる。また、自動車内の雑音環境でもカーナビゲーションの音声入力が可能になってきている。

第2章 特許動向分析

第1節 特許分析の考え方

1968年からの音声認識技術の特許動向を調査するために、全体動向、三極出願動向の分析には、一般的に使用される国際特許分類（IPC）ではなく、三極の各国・地域における独自分類を用いた。音声認識技術に相当するIPCは最新版である第7版は技術的に細分化されているものの、1999年以前の第6版までは分類が粗く、1968年まで遡っての技術区分別動向分析には不適と判断した。各国・地域の独自分類は、IPCよりも細分化され、かつ、過去に遡って付与されていることから、1970年代から現在までの長期間にわたる特許動向の調査に適している。

要約 2-1 表 特許全体調査で用いた特許分類

三極地域	使用した特許分類方式	特徴
日本	FI	IPC 6 版をベースに、日本特許庁で独自に細分化。新分類は過去に遡って付与されている。
米国	USPC(US Patent Class)	米国特許商標庁が策定した IPC とは無関係の独自分類。新分類は過去に遡って付与されている。
欧州	ECLA	欧州特許庁が IPC をさらに細分化したもの。IPC 7 版にも対応している。欧州の EP 特許以外に、欧州各国の特許にも多く付与されている。新分類は過去に遡って付与されている。

出典：各データベースの説明資料を基に作成。

音声認識技術を7つの技術区分に分け、上記各地域別の特許分類に従って要約 2-2 表に示す各技術区分へと分類した。音声合成や音声分析が混在する可能性のある分類は音声認識関連のキーワードで抽出した。一部対応分類がないものはIPCとキーワードを併用した。実用化技術は音声認識技術を用いた製品に関する特許も含むと考え、日本はFIでの対応分類を用い、米国は対応分類とともに音声認識関連のキーワードで抽出した。欧州は対応分類がないため、音声認識関連のキーワードで抽出した。

要約 2-2 表 音声認識技術の技術分類

技術区分分類	音響処理技術
	音響モデル作成・適応化技術
	マッチング・尤度演算技術
	言語モデル技術
	対話処理技術
	話者認識技術
	実用化技術

一方、音声認識応用製品の動向の分析には、実用化の中の「音声認識技術を用いた製品」を細分化し、その音声認識技術を用いた製品が出始めた1991年以降を調査した。応用製品の分類にはIPC第5版以降同一のIPCが付与されているためIPCを用いた。

詳細分析は上記全体分析で判明した重要出願人の特許動向をマクロ分析及び抄録、明細書を精査して分析するとともに、近年の音声認識技術を主導する重要技術に焦点をあて、その特許権利状況を調査した。なお、本調査の中で出願年とは優先権主張日の年を、また特許の出願人の国籍とは優先権を主張して出願した地域の国籍とした。

第2節 全体分析

1. 音声認識技術関連の特許出願件数・登録件数の推移

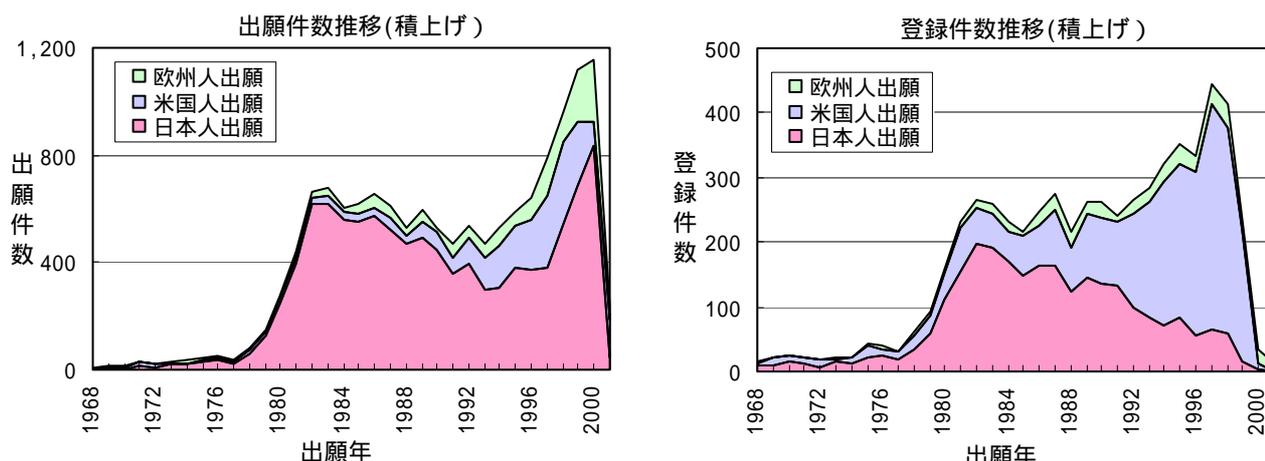
日米欧三極に出願された特許のうち、自国・地域の出願人が地域内に出願及び登録した特許件数(地域内出願)を抽出し、合計することで、重複なしに3極への出願件数、登録件数を数えた。1968年以降の出願件数は14,246件、登録件数は4,889件であった。

出願件数は1980年代に入り急増した後、いったん減少するが、1990年代以降再び増加している。1980年代は日本人の出願が大半を占めていたが、1990年代以降、米国人、欧州人の出願が増加してきている。

登録件数も1980年代初頭は日本人による出願が大部分を占めていた。しかし、日本人出願の登録件数は1982年をピークに減少し始めている。逆に、1990年代は米国人出願による特許の登録件数が増加してきている。こうした状況から、1980年代の音声認識技術の研究を日本が牽引し、1990年代は米国が牽引している様子を窺うことができる。

ただし、上記の分析については、日本では出願された特許は全件を公開しているが、米国では2000年までは登録された特許のみが公開されていたこと、また、日本では2001年の出願までは出願後7年間は審査請求の期間が認められているため、1990年代後半に出願され、登録される特許が今後増えてくる可能性を考慮しておく必要がある。

要約 2-1 図 音声認識技術の出願件数・登録件数推移



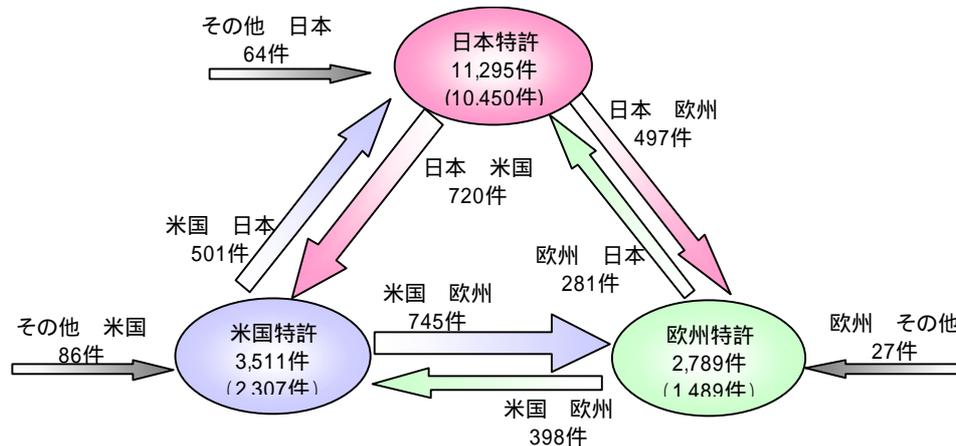
2. 三極別の特許出願構造

三極への出願構造をみると、日本への出願は大部分が日本人によるものである。米国への出願は日本人、欧州人の出願をあわせると3割程度となっている。欧州では、欧州人による出願は全体の50%強に過ぎない。日本に対して、欧米勢の出願が少ないのが目を引く。

音声認識技術は、音響モデルや言語モデルは言語に依存する部分もあるが、音響処理技術

やマッチング・尤度演算技術、さらには言語モデルの作成などは根本的には言語への依存度が低い。出願の偏りは欧米の企業が日本語を認識させる製品、ミドルウェアを開発し、商品化したのは最近のことであることや、日本語を認識させる製品、ミドルウェアの市場規模も小さいことなどが理由と思われる。

要約 2-2 図 音声認識技術の三極出願特許収支



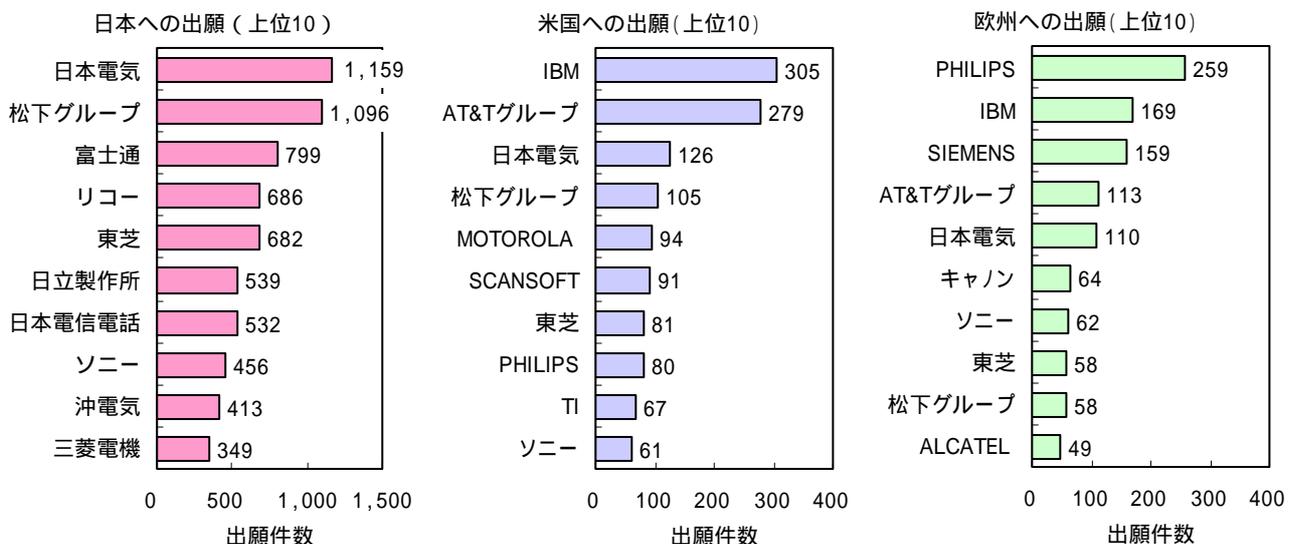
注) 1.対象は 1968 年以降に出願された特許。

2.日本特許、米国特許、欧州特許の下のかっこ内件数は各国、地域内からの出願件数。

3. 三極別の出願人の内訳

三極別の出願人の内訳をみると、日本への出願では日本電気、松下グループ、富士通などの大手電機・通信機器メーカーが上位を占めている。米国への出願は IBM と AT&T グループが抜きん出ている。そのほか、米国では専門メーカーである ScanSoft が健闘している。欧州への出願では Philips が首位にあり、IBM、Siemens と続いている。

要約 2-3 図 三極別上位出願人の内訳



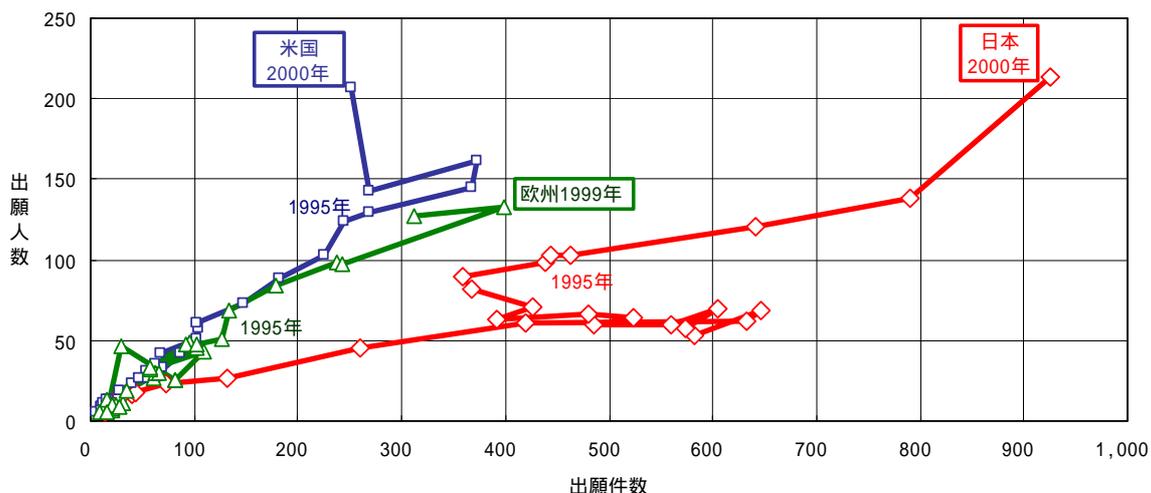
注) TIはTexas Instruments

4. 三極別の出願件数と出願人

三極別に特許出願件数と出願人数の関係をみると、三極ともに出願人数は100人を超えており、多様なメンバーにより研究開発が行われていることがわかる。また、三極ともにグラフが右肩上がりになっているのは、音声認識技術がまだ成長期にあることを示すものである。

日本への特許出願件数の推移をみると、1980年代後半に出願人数の大きな増減がないまま出願件数は大きく減少しており、一時的に研究開発が停滞した状況を読み取ることができる。この時期はマッチング・尤度演算技術がDTWからNN、さらにHMMへと移行する過渡期であり、旧来の技術が衰退し、新たな技術により再び研究開発が盛んになった時期である。

要約 2-4 図 音声認識技術の特許出願件数と特許出願人数



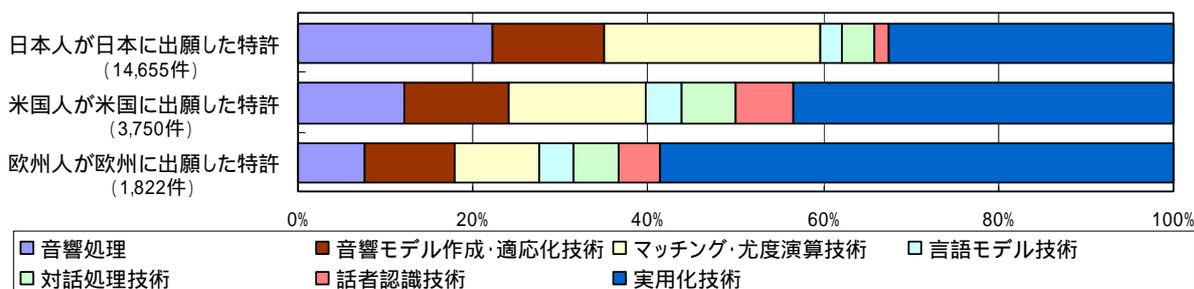
注) 対象期間は1971年から2000年。

5. 三極別の特許出願の内訳

三極の地域内特許出願の技術区分別動向を見ると、三極ともに実用化技術に分類される特許が多い。日本人が出願した特許は、欧米に比べて音響処理技術やマッチング・尤度演算技術が多く、言語モデル技術、対話処理技術、話者認識技術に関するものは少ない。

日本人が特許を多く出願していた1980年代にはDTWによるマッチング技術が多く研究され、言語モデル技術、対話処理技術はあまり研究されていなかったことが理由の一つと考えられる。

要約 2-5 図 音声認識技術出願特許の技術区分別件数



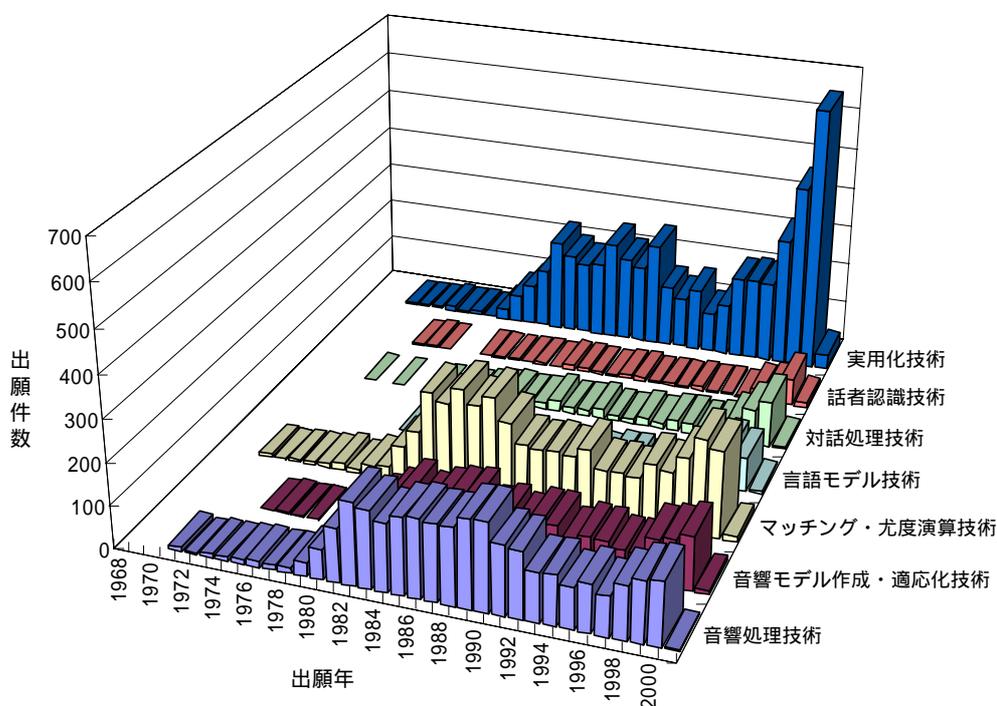
注) 対象は1968年以降に出願された特許。

6. 三極別にみた技術区分別の特許出願件数の推移

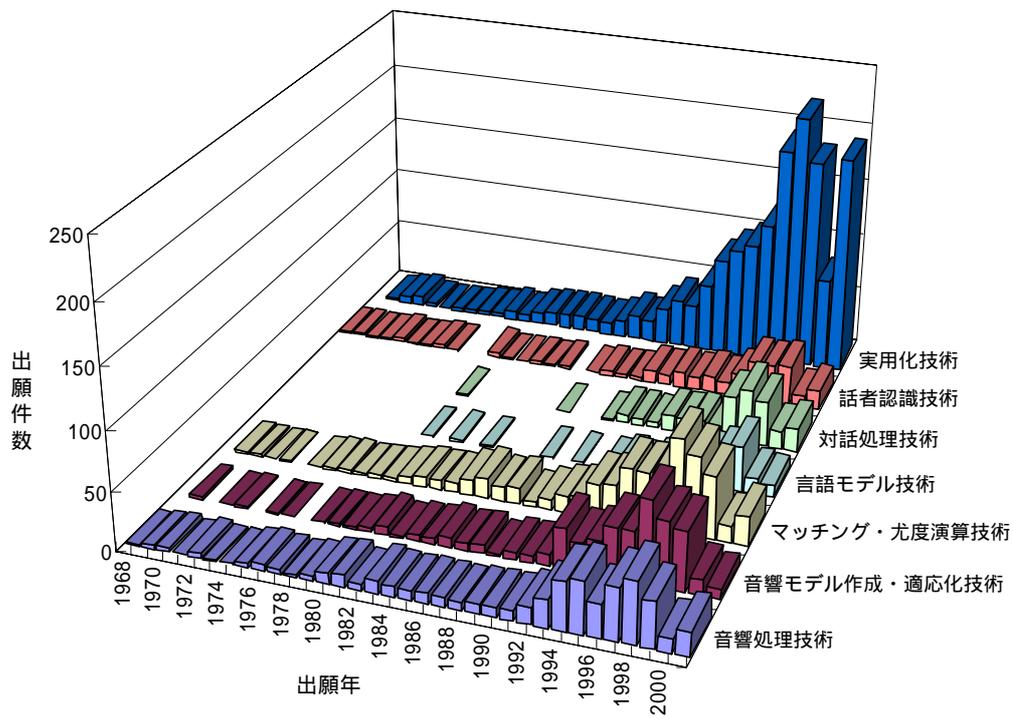
日米欧三極の出願人がそれぞれの地域に出願した地域内出願に注目し、技術区分別の出願件数の年次推移を見ると、1980年代は日本人による音響処理技術、マッチング・尤度演算技術に関する出願が多い。米国人は1990年代に音響処理技術、音響モデル作成・適応化技術、マッチング・尤度演算技術に関する特許出願が増加している。

話者認識技術は1990年代に米国人による出願件数が増えており、米国では電話による話者認識のニーズが高いことを示している。また、近年は実用化技術に関する出願の伸びが著しい。これは、音声認識技術が実用的なレベルまで達し、応用製品の研究開発が進められてきたことを示すものである。なお、対話処理技術、言語モデル技術は1990年代に特許が出願され始めるのは、同技術の研究が1990年代にようやく本格化したためであろう。

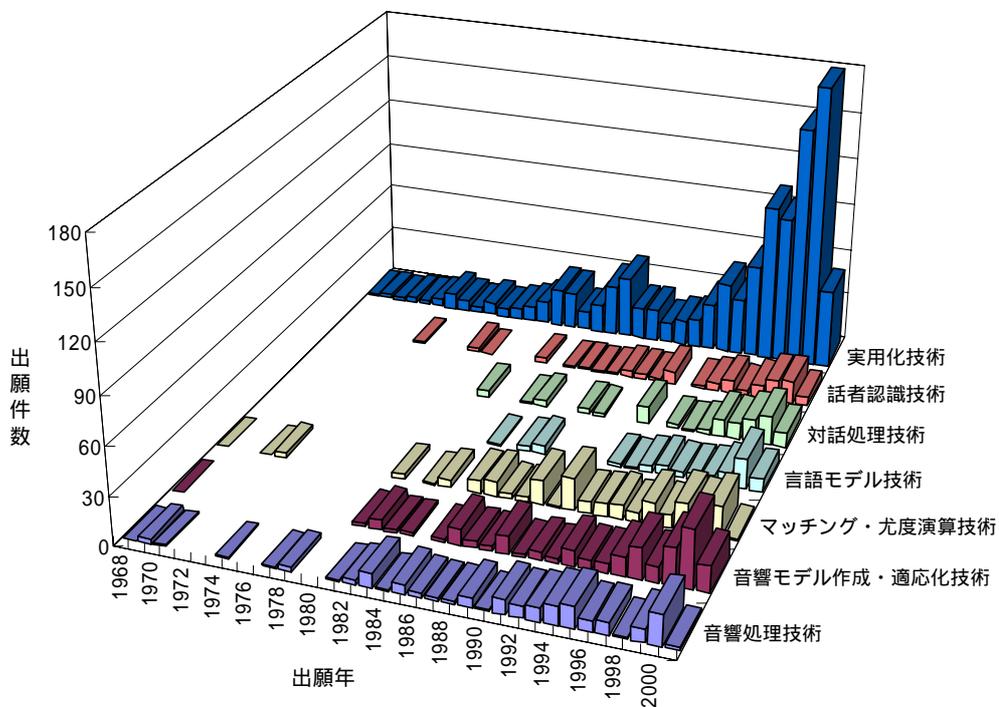
要約 2-6 図 音声認識技術出願特許の技術区分別件数推移（日本人が日本に出願した特許）



要約 2-7 図 音声認識技術出願特許の技術区分別件数推移（米国人が米国に出願した特許）



要約 2-8 図 音声認識技術出願特許の技術区分別件数推移（欧州人が欧州に出願した特許）

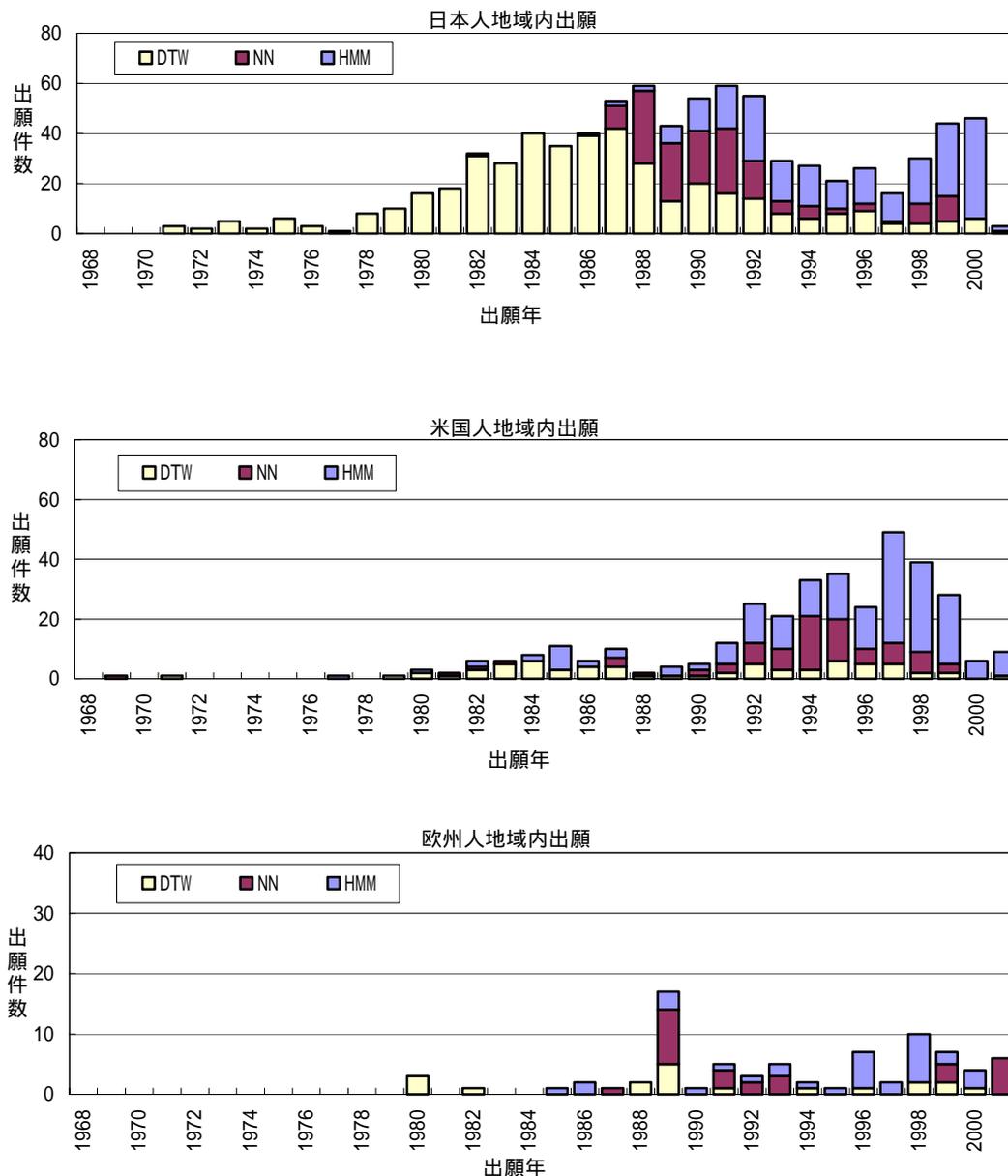


7. 三極別にみた特許出願における主要技術の変遷

要約 2-9 図はマッチング・尤度演算技術の中で DTW、NN、HMM に関する特許の出願件数の推移を示したものである。

日本人の出願は 1980 年代の DTW から NN、HMM へと内容が変わっているのに対して、米国人の出願は 1990 年代の HMM に関するものが多い。HMM に関する特許は、日本では 1990 年代前半から出願されているものの、しばらく件数の著しい増加はなく、米国で 1997 年に立ち上がった後に追従して 1999 年頃から立ち上がってきている。これは、DTW 技術の時代は日本の研究開発が世界をリードし、HMM 技術の時代は米国がリードしていることを示すものである。

要約 2-9 図 マッチング・尤度演算技術の細分の出願・登録件数推移

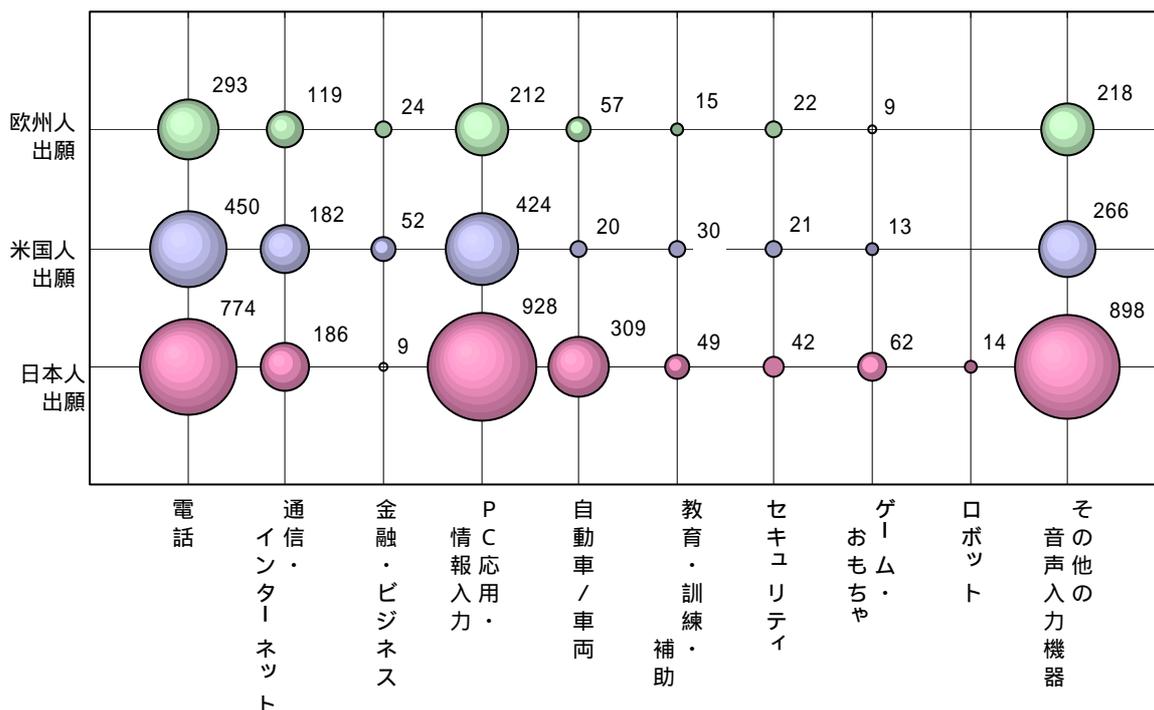


8. 実用化技術の細分化による三極の特許出願内容の違い

実用化技術を細分してみると、三極を通じて電話・通信に関する特許及びパソコン・情報入力に関する特許が多い。米国人の出願はこの2つが大半を占めている。

一方、日本人による出願をみると、自動車関連の特許やゲーム・おもちゃに関する特許出願が他地域に比べ多くなっている。これは、カーナビゲーションやテレビゲーム、ハイテクおもちゃなど日本が強い製品分野においての音声認識を応用した製品の開発が進んでいることを示すものである。

要約 2-10 図 実用化技術細分出願件数の三極出願人比較



第3節 出願人別動向分析

最近の10年間について、日米欧三極で特許を出願した件数が最も多い日本電気、IBM、Philips に関して出願・登録の件数推移と出願特許の技術区分別の構造を分析した。また、日米欧三極で出願件数が二番目に多い企業（松下グループ、AT&T、Siemens）までを対象として、出願された特許の内容についての技術区分ごとの分析を行った。

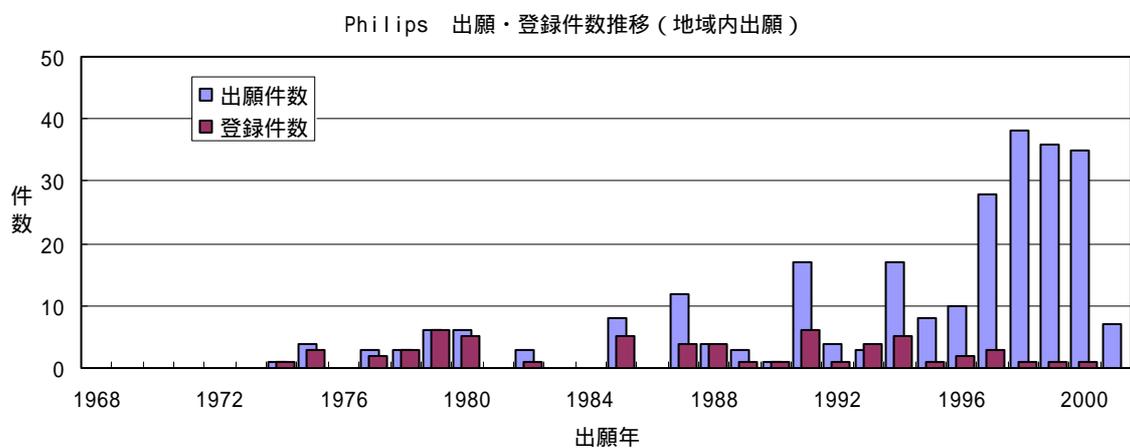
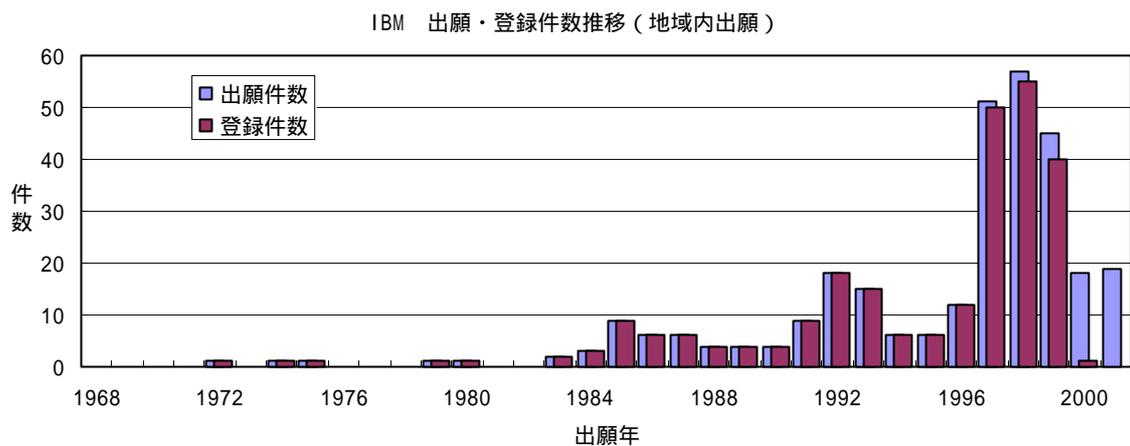
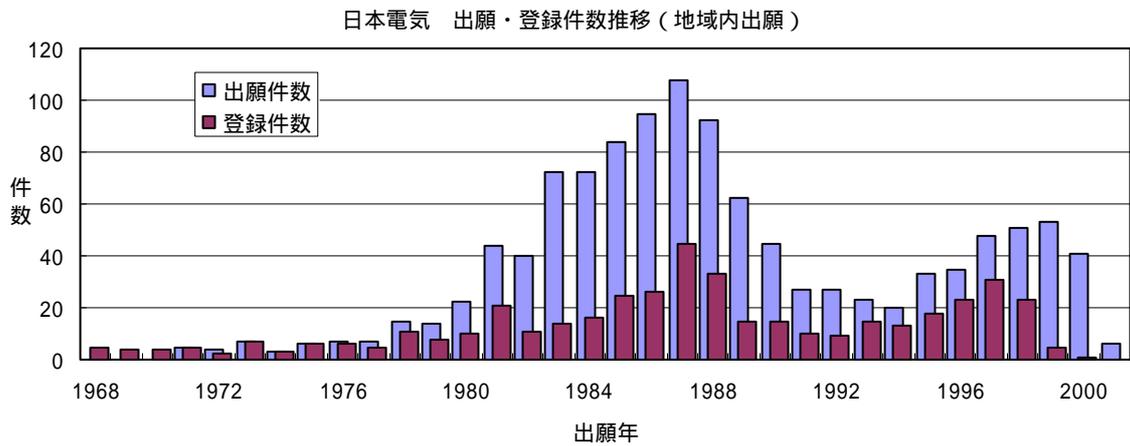
日本電気は1987年を中心に多数の出願を行っていたが、1990年代の初頭に出願件数が大きく減少している。これは、DTWでの技術開発に注力していた研究開発がトーンダウンしたことを反映するものである。同社の出願内容をみると、近年の特許件数の増加はHMMをベースにしたものに変ってきている。

一方、IBMは1980年代の出願件数は少なく、1992年と1997年以降の出願件数の増加が目立っている。IBMの出願件数が大幅に増加した年はDARPAの音声認識に対する諸プロジェクトが終了した年と相関がある。DARPAの諸プロジェクトで得られた研究開発の成果を特許として出願していると推察することができる。また、1997年はViaVoiceが発売された年であり、

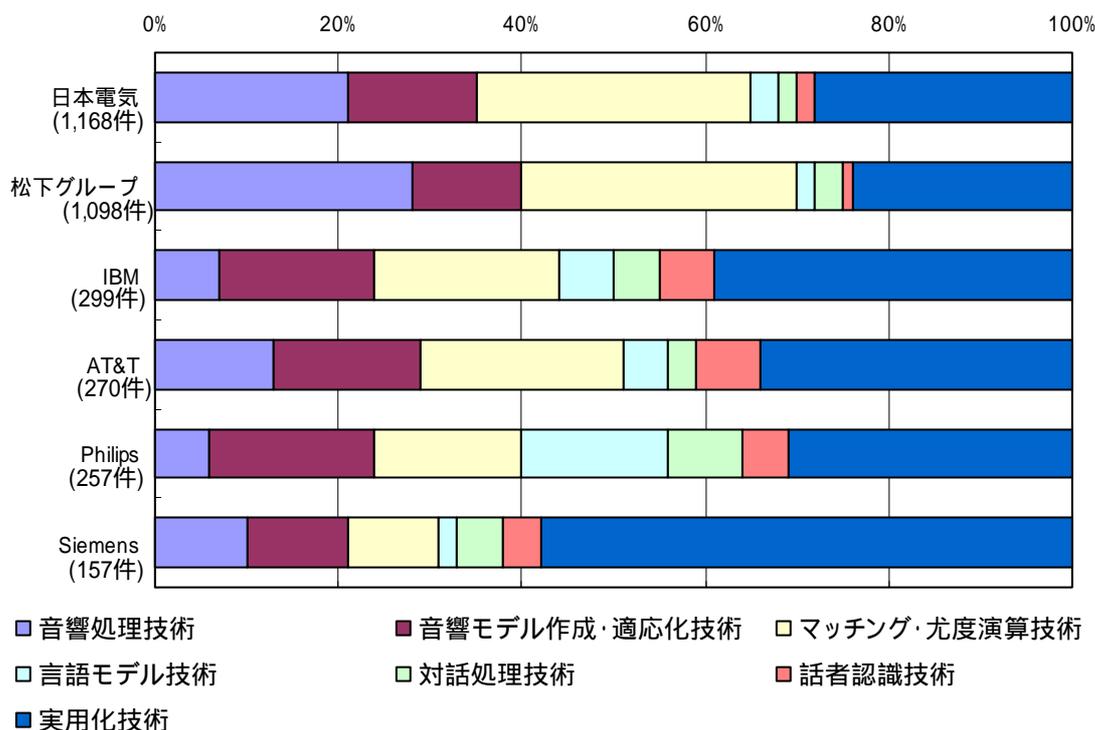
この製品との相関もあると思われる。

Philips は 1970 年代の早い時期にも特許を出願しているが、出願件数が増加してきたのは 1997 年からである。出願件数の増加が目立つ 1991 年や 1994 年は欧州の音声認識プロジェクト ESPRIT との相関が見られる。

要約 2-11 図 三極主要出願人の特許出願・登録件数推移



要約 2-12 図 三極主要出願人の地域内に出願した特許の技術区分別構造



注) 対象は 1968 年から 2001 年に出願した特許。

IBM や Philips は音響処理技術よりは音響モデル作成・適応化技術、マッチング・尤度演算技術に注力し、言語モデル技術の割合も高い。一方、日本電気や松下グループは、音響処理技術が音響モデル作成・適応化技術、マッチング・尤度演算技術と同程度の割合を占めている。AT&T は話者認識技術に関する特許が多く、Siemens や IBM は実用化技術に関する特許も多い。

日本企業に音響処理技術やマッチング・尤度演算技術に関する特許が多いのは、1980 年代にこの技術が日本企業によって広く研究され、多数の特許が出願されていたためである。言語モデル技術や対話処理技術の特許が少ないのは、日本企業が特許を多数出願していた 1980 年代にはこれらの技術の研究が進んでいなかった面もある。また、日本企業は実用化の特許が比較的少ないが、欧米の企業は原理原則的な技術の他に実用化の技術の開発にも注力していると見ることもできる。

現在、米国で音声認識ビジネスのリーダー的存在となっている Nuance Communication (以下、Nuance)の特許出願件数は 5 件、SpeechWorks International (以下、SpeechWorks および Nuance)の特許出願件数は 8 件あり、内容は対話処理技術に関するものや実用化技術に関するものが多い。

音声認識ビジネスを牽引しているベンチャー企業は、音声認識の技術的な部分を経営者の出身母体である大学・研究所やその他の特許保有企業からのライセンス供与や、必要な場合は特許の購入、企業買収による特許の権利取得によってまかなっている。これらの企業が取得する技術の多くはビジネスに直結する分野である。

第3章 音声認識技術の研究開発動向

第1節 歴史的経緯

本節では、音声認識技術の歴史的発展経緯を代表的技術によって振り返っておく。

1．DTW法（1970年代 - 1980年代）

DTW(Dynamic Time Warping: ダイナミックタイムワーピング)法は、特徴パラメータの時系列パターンより音響モデルを作成し、DP(Dynamic Programming: 動的計画法)の最適性原理に基づいて、入力音声の特徴パラメータの時系列パターンと標準モデルの特徴パラメータの時系列パターン(しばしばテンプレートと呼ばれる)を非線形に直接比較して照合する方法である。日本では、DP マッチングと呼ばれることも多い。主に、特定話者音声認識の実現に使用された。

日本では日本電気の迫江、千葉が1971年にDTWによる音声認識を提案(特登986698)して以来、多くの企業が研究開発を重ね、staggered array DP、SPLIT、連続DP、2段DPなどの多くの手法を発明し、特許の出願、登録が行なわれた。しかし、不特定話者音声認識を実現するには不特定話者間の発声変動を表現するために1つの単語当たり複数のテンプレートを用意する必要があり、実現コスト(メモリサイズ、プロセッサ価格)がかさむため、HMMに主役の座を取って代わられた。

2．HMM法（1980年代 - 現在）

米国では、1970年代に統計確率的手法であるHMM(Hidden Markov Model: 隠れマルコフモデル)の研究がCMU、IBMに続いて、AT&Tで進められた。1982年には、AT&TのRabinerらにより、HMMを利用した音声認識の特許が出願された(優先権主張1982年10月15日:US4587670)。不特定話者の多量の音声データから求められた特徴パラメータの時系列パターンから最尤推定の原理(EMアルゴリズム)に基づいて、HMMによる音響モデルを作成し、ビタビアルゴリズム(通信分野への適用について、1990年に特許出願・登録:US5193094)により、入力音声の特徴パラメータの時系列パターンの尤度を計算する。

HMM法は不特定話者音声認識の実現が容易なため、導入から25年たった現在でも、主流の研究手法となっている。また、HMM法は音響モデルの適応化も実現が容易なため、話者適応や環境適応の研究開発も近年盛んに行われている。

3．NN法（1980年代 - 1990年代前半）

人間の神経細胞における情報処理の数学モデルであるNN(Neural Network:ニューラルネットワーク)によるパターン認識が1990年代に盛んに研究された。音声認識にも適用され、ヘルシンキ工科大のT. KohonenによるSOM(Self-Organization Map:自己組織化マップ)、LVQ(Learning Vector Quantization:学習ベクトル量子化)、ATRのA. WaibelによるTDNN(Time Delay Neural Network:時間遅延ニューラルネットワーク)(いずれも特許出願なし)などが盛んに研究された。

一時は、HMMによる認識性能を越えるのではと熱い期待を浴びたが、長年の研究開発の結果、HMMを超える性能は確認されず、現在では研究開発が沈静化している。実用化の点でも、NN法を標榜している企業は数社に過ぎない。

第2節 研究開発動向分析

音声認識を含む音声技術の学会・カンファレンスとしては、ICASSP(The International Conference on Acoustics, Speech and Signal Processing)が世界的に有名である。

最近8年間のICASSPの発表の傾向を分析した結果として以下のような特徴が見られる。

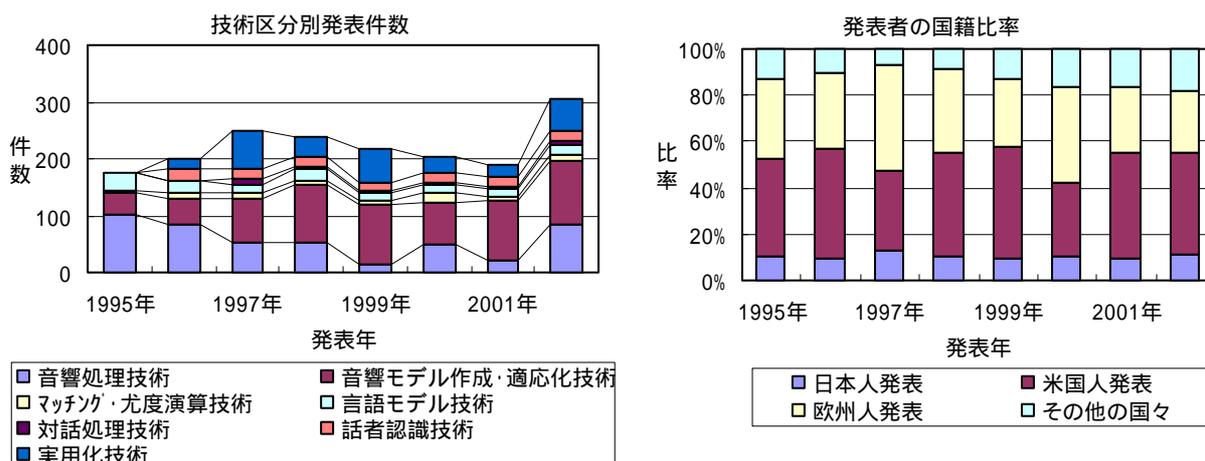
音響モデル作成・適応化技術、音響処理技術、実用化技術の順で研究発表が多くなっており、研究の関心が実環境での雑音に対する頑健性の向上や話者適応に向かってきている。

一方、マッチング・尤度演算技術、言語モデル技術、対話処理技術、話者認識技術の研究発表は相対的に少ない。

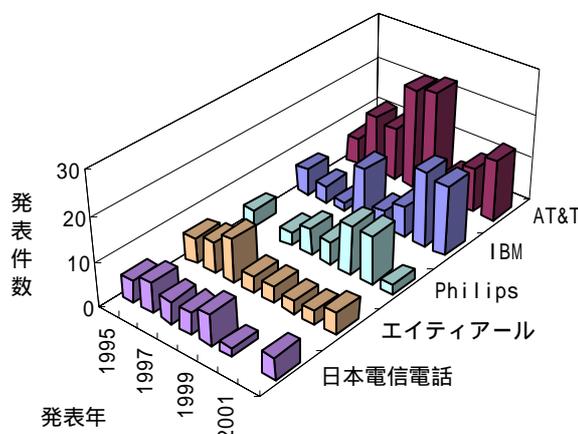
欧米の発表者の比率が多く、日本の研究者の発表は概ね10%程度で推移している。

特許出願の多いIBMやAT&Tの発表件数が多く、日本企業では日本電信電話、エイティアール以外の発表件数は少ない。

要約3-1図 ICASSP 発表件数推移



主要企業の発表件数推移



出典：IEEE Proceedings of ICASSP, (1995-2002) を基に作成。

第3節 国家的プロジェクトと研究開発リーダー

1. 日本

日本では文部省（現文部科学省）の科学研究費補助金、情報処理振興事業協会、日本学術振興会などの助成金が大学の音声認識の研究開発を支援してきた。東京大学、京都大学、名古屋大学、早稲田大学、慶応大学、東京工業大学、奈良先端科学技術大学院大学、豊橋技術科学大学院大学などが学会を主導してきた。

産業界では、日本電信電話、旧 KDD などの通信会社、日本電気、日立、東芝、三菱電機、富士通、松下、リコー、ソニーなどの電気メーカーの研究所が精力的に研究開発を行ってきた。特に、1980年代は板倉（日本電信電話）らの線形予測分析の研究や迫江（日本電気）らの DTW に基づく音声認識技術に触発され、活発かつ理論的な研究が進められ、特許出願・登録件数も欧米を圧倒し、世界をリードしていた。1986年以降は、エイティアール(ATR：国際電気通信基礎技術研究所)が HMM に基づく音声認識の研究開発の世界的拠点として中心的役割を果たしてきた。過去数年間は、企業による学会発表の件数が大幅に減少する一方で、カーナビゲーションやロボットなどへの音声認識機能搭載において世界的に先陣を切っている。

音声認識は研究所で研究を進める段階から、個別製品に組み込み、広く利用を進める実用化の段階へ移行してきている。その中で、日本は今後とも実用化技術の研究開発において、世界を主導し続けることになるだろう。一方、話者認識に関しては日本電信電話を中心として数社程度しか参入企業がない状況である。

2. 米国

米国は 1980年代後半から 1990年代にかけて、DARPA の諸プロジェクトが音声認識の研究を支援してきた。DARPA の諸プロジェクトは当初、音声認識の基礎的研究から始まり、制約のある発話の認識など徐々に難易度を高めていった。電話音声などの自然会話、英語以外の言語を認識する研究を進める一方、ここ数年は組み込み型の音声認識、雑音に頑健なコマンド制御の分野も研究されている。同プロジェクトの特徴としては、強力なリーダーシップの下での各研究機関の研究テーマの組織化、共通の評価タスクの設定、評価用の音声コーパス（音声データベース）の公開に基づいたコンペティションの開催による研究機関同士の競争促進などを指摘することができる。

企業では AT&T のベル研究所、IBM などが、大学では MIT(マサチューセッツ工科大学)、CMU(カーネギーメロン大学)、OGI(オレゴン科学技術大学院大学)などが、研究機関では SRI(スタンフォード研究所)などが DARPA の諸プロジェクトに参画しながら音声認識技術の牽引役として過去 30年間大きな役割を果たしてきた。最近数年は MIT や SRI からスピナウトしたベンチャー企業である SpeechWorks、Nuance が電話系の音声認識ビジネスを主導している。

一方、話者認識に関しては、最近では米国 NIST(国立標準技術研究所)が評価タスクを設定し、コンペティションが定期的に行われている。

3. 欧州

欧州では 1984年からの 5期にわたる Esprit プロジェクトの中で、各国の電話会社系の研究所や大学が中心となり、各国語の音声認識の研究が精力的に進められた。このプロジェク

トで取り上げられた主なテーマには電話回線を用いた音声対話システムの研究や多言語音声認識、翻訳などがある。

企業としては、Siemens、旧 L&H (2002 年 3 月に L&H の破産に伴い、Speech Technology 部門を ScanSoft が買収)、Philips(2002 年に音声認識技術部門の一部を Scansoft が買収と発表)、Temic などが欧州において音声認識技術の研究開発を積極的に行ってきた企業である。

4．総括

現在の音声認識で使われる主要技術の多くは 1970-1980 年代に遡ることができる。1980 年代中頃から、日米欧三極で国家的プロジェクトが盛んになってきた。これは半導体技術の飛躍的な進歩による計算機技術の進展に後押しされ、様々な発見、発明が加速度的に成されてきたことと深い関係がある。特に、HMM に基づく音声認識、話者認識技術は大規模な音声コーパス(研究機関等にて一般利用可能な音声データベース)やテキストコーパス(研究機関等にて一般利用可能な文章・記事などのデータベース)の蓄積と相まって、過去 15 年で長足の進歩を遂げ、不特定話者、大語彙連続音声認識を可能としている。

第 4 節 今後、解決されるべき技術課題

1．実環境(雑音、千差万別な発話様式、話し言葉など)における音声認識性能の向上

解決されるべき最優先の課題は実環境における音声認識、話者認識性能の向上である。

半導体技術の進歩は年々費用的に優れたプラットフォームをより安価に提供している。しかし、費用的に優れたプラットフォームが提供されたとしても、実環境における音声認識性能が根本的に向上するわけではない。雑音に対する堅牢性、千差万別な発話様式、話し言葉の多様性などの複雑な要因に対する有効な手法は未だ提案されていない。

音声認識利用の必然性やニーズが高いアプリケーションに焦点を当て、それがおかれている実環境及び実装プラットフォームを考慮しながら、音声認識、話者認識性能を阻害する要因を 1 つずつ解析し、しかるべき対策を講じる必要がある。

2．多言語対応

日本で音声認識技術を組み込んだ製品を製造、発売している企業の多くは海外への輸出も手がけている。音声認識、話者認識技術が実用レベルに近づきつつあるとの認識が浸透するにつれて、これら企業の製品に対する多言語対応への要求も高まっている。

これまで、日本では多言語対応の研究開発の重要性が看過されてきた。しかし、多言語対応も今後注目が集まるテーマになっていくであろう。

3．音声認識と話者認識の統合

音声認識と話者認識は技術的な親和性が高いものの、これまでは独立に研究開発が進められてきた。しかし、今後は、安全性を求める要望が高まってくることが予測されるため、音声認識と話者認識機能との統合、すなわち、音声話者同時認識により、誰が何を喋ったかを認識する機能のニーズが高まってくると考えられる。

4. マルチモーダルインターフェイス

今後は、狭義の音声認識機能に加えて、音声合成機能、対話管理機能を統合した音声インターフェイス機能の実用化のニーズが高まってくるであろう。また、ベイジアンネットワークなどの統計的な枠組みを利用した、画像認識、センサー情報認識との統合による状況認識や画像合成などの出力機能を統合したマルチモーダルインターフェイス機能の研究開発も今後さらに活発化するであろう。

第4章 音声認識技術の市場概況

第1節 概況及び市場予測

現在、音声認識技術が使われている市場は要約4-1表「音声認識技術の使われている市場」のように大別することができる。

要約4-1表 音声認識技術の使われている市場

分野	用途	商品・サービスの例
a) 電話・通信・インターネット	電話応答、内線番号案内、ボイスポータル 音声検索	電話音声応答システム（CTI）製品 Vポータル（NTTコミュニケーションズ） Voizi（日本テレコム） ボイスエージェント（au, KDDI）
b) PC	ディクテーション	ViaVoice（IBM）、Dragon Speech（ScanSoft）
c) その他音声操作・入力機器	車載機器 ゲーム・おもちゃ・ホビー ロボット 言語教育・言語訓練・身障者補助 話者認証によるセキュリティー その他様々な装置の音声操作	カーナビゲーション、ボイスダイヤル パウリンガル、シーマン アイボ、アシモ 英語発音教材、音声リモコン 音声認証機器

現在、これらの市場はコンピュータによるCTI（Computer Telephony Integration：コンピュータによる電話応答システム）を利用した受付案内などの電話用途及びパソコンによる口述筆記（ディクテーション）用ソフトウェアで占められている。

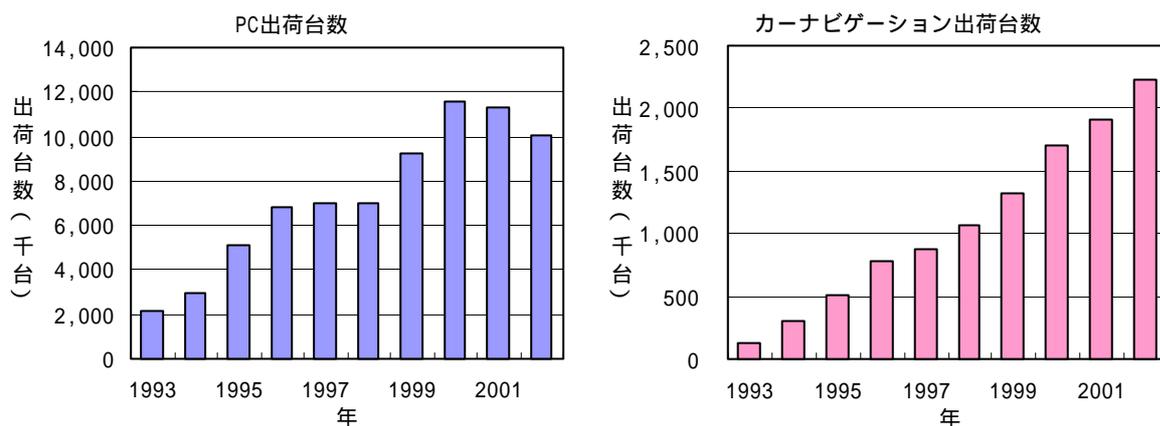
口述筆記用ソフトウェアを搭載するパーソナルコンピュータ（PC）の市場は最近伸び悩んでいる。今後音声認識機能に標準で対応するOSであるWindowsXPの普及などで音声認識機能の搭載率は高まると考えられるが、PCでの音声認識市場の先行きは不透明である。

一方、カーナビゲーションは順調に出荷台数を伸ばしている。2002年の出荷台数は222万台と200万台の大台を突破してきており、2007年には460万台に達するとの見通しもある。また、道路交通法などで車載機器にハンズフリー機能が求められるために音声認識機能付きのカーナビゲーションの市場はさらに伸びるとの見方もある。（要約4-1図参照）

NTTコミュニケーションズのVポータルや、au（KDDI）のボイスエージェントなどに代表される音声ポータル、音声検索のサービスが近年開始され、有料コンテンツとの連携も拡充されつつある。この分野の市場も今後の伸びが期待されている。

その他、今後音声認識市場としての発展が予想される分野は、家庭内機器の音声操作、業務用端末の音声入力、ロボットへの音声指示などである。

要約 4-1 図 日本国内のパソコン（PC）及びカーナビゲーションの出荷台数

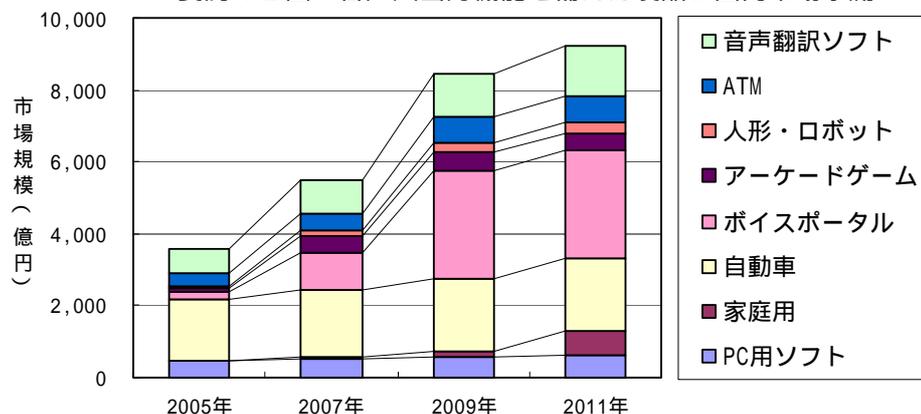


出典：社団法人電子情報技術産業協会「コンピュータおよび関連装置等出荷統計」、「民生用電子機器国内出荷統計」を基に作成。

日本の音声認識・合成ソフトウェアの市場について、日本経済新聞社・三菱総合研究所¹では2000年に188億円だったものが2010年には1,200億円規模になると予測している。

音声入出力機能を備えた製品・モジュール・ソフトウェアの国内市場の規模を既存データから予測すると、2011年には9,000億円に達する見通しとなった。（要約 4-2 図参照）これは音声入出力機能を備えた製品群として音声認識と音声合成機能を備えた製品、ソフトウェア、サービス又はモジュールを抽出したものである。この中で音声認識技術に直接対応する金額は定かでないが、一般には製品の中でミドルウェアのライセンス価格は1%から3%とされている。仮に、3%とするとミドルウェアの市場規模は2010年で300億円規模になる。

要約 4-2 図 音声入出力機能を備えた製品の国内市場予測



出典：旭リサーチセンター推計。

海外市場の見通しでは、Frost & Sullivan が電話向け音声認識市場は2002年の1億ドルから2006年には20億ドルに、非電話向け米国音声認識市場は2002年の9.5億ドルから2007年には23.4億ドルになるとの予測を発表している。同社の予測では、話者認識技術に関してはバイオメトリクス認証の一つとして期待感があるものの、市場規模は2006年に35百万ドルと小規模にとどまると見ている。

¹ 大予測 21世紀の技術と産業 p203 日本経済新聞社・三菱総合研究所編

要約 4-2 表 米国音声認識技術の市場規模予測

	2002 年	2006 年
電話用途音声認識	1 億ドル	20 億ドル
非電話音声認識	9.5 億ドル	23.4 億ドル

出典：Frost&Sullivan「North American Telephony Based Speech Technology Software Market #6329-62, 2002」
「US Non-Telephony Speech Recognition Market #5093-11, 2001」を基に作成。

第 2 節 主要企業の動向と市場シェア

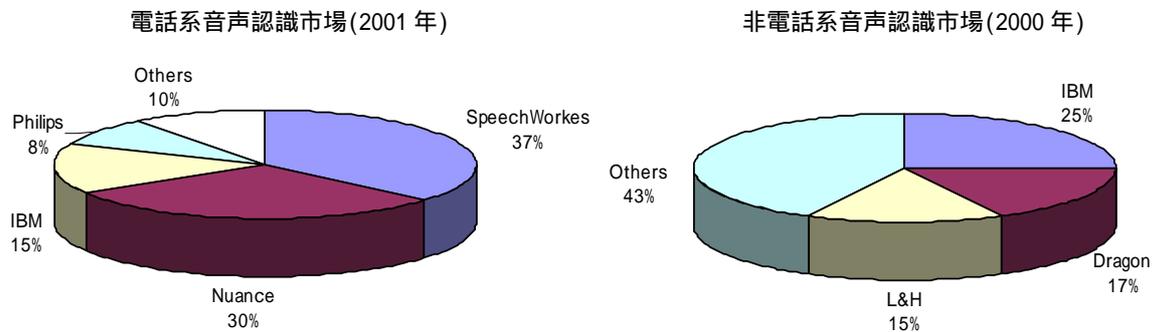
世界の音声認識市場を俯瞰してみると、

- ・ Nuance、SpeechWorks に代表される電話音声認識市場
- ・ IBM、Dragon Systems (現 ScanSoft) に代表されるディクテーション市場
- ・ L&H (現 ScanSoft) やベンチャー企業が中心の組み込み向け音声認識市場

に大別できる。

米国の音声認識市場のシェアをみると、電話用途音声認識は SpeechWorks と Nuance が、非電話用途音声認識は IBM と Dragon Speech の割合が高い。非電話用途の市場は現時点では口述筆記が主になっており、組み込み用途市場はまだ大きくはない。

要約 4-3 図 米国の音声認識技術市場におけるシェア



注) 調査レポートの発刊年が異なるために、2つのグラフの年次が異なる。

出典：Frost&Sullivan「North American Telephony Based Speech Technology Software Market #6329-62, 2002」
「US Non-Telephony Speech Recognition Market #5093-11, 2001」を基に作成。

日本における電話用途音声認識、口述筆記等非電話用途の音声認識の市場シェアは定かではないものの、音声認識市場への主な参入企業には IBM、日本電気、日立製作所、東芝、日本電信電話、旭化成、アドバンスト・メディアといった企業がある。

近年海外では ScanSoft が L&H や Philips の組み込み音声認識部門を買収するなど、合従連衡の動きが盛んである。これは、今後拡大してくると見られている音声認識市場に対して、合従連衡によって、競合他社より優位に立とうとする動きと考えられる。また、音声認識技術の企業が話者認識技術や音声合成技術を持つ企業と連携する例 (SpeechWorks が T-Netix の話者認識技術を買収) も多く、音声認識ベンダーでありながら音声合成ではライセンスを受ける例もある。音声認識のみでなく、総合的な音声ソリューションを市場が求め始めている

と推察される。

近年、従来電話型の音声認識や PC の口述筆記に目を向けていた音声認識ベンダーが組み込み用途向けの分野に進出してきている。IBM は従来の PC 向け ViaVoice に加え Embedded ViaVoice という製品を開発し、日本のカーナビゲーションや人型ロボットに採用されている。SpeechWorks も Speech2Go という組み込み向け音声認識製品を発表している。

一方、日本電気をはじめとする日本企業は従来から自社のマイクロコンピュータで動作する音声認識ミドルウェアを商品化していたが、他社のマイクロコンピュータへの対応を図ったり、Intel 製のチップを採用した Pocket PC や PDA で動作するように調整して市場展開を図るなどのクロスプラットフォーム戦略を取り始めている。また、電話、携帯電話によるボイスポータルに加え、携帯電話や PDA を用いた翻訳サービスなど、最近普及してきた IT インフラと音声認識技術のコラボレーションによる市場創出の動きも激しくなっている。

第 5 章 政策及び標準化動向

第 1 節 音声認識に関わる政策

1. 日本の改正道路交通法

音声認識技術に関係する政策としてはまず、車載機器の操作に関するものがある。

日本では 1999 年 11 月に施行された改正道路交通法において、運転中の携帯電話の使用や、カーナビゲーションの画面の注視が禁じられた。欧州では同様の法令を施行している国が多く、米国でも 2001 年 11 月にニューヨーク州で運転中の携帯電話使用が禁じられ、各州及び連邦政府でも議論がなされている。これにより、携帯電話のボイスダイアリング機能のニーズが高まっているため、カーナビゲーションの音声認識搭載率も高まると予想される。

2. 米国リハビリテーション法第 508 条

米国リハビリテーション法第 508 条の影響も見逃すことができない。同法は、米国連邦政府各省や機関が電子情報技術を開発・調達・保守・利用する際に、

- ・ 障害を持つ連邦政府職員の電子情報技術の利用が、障害を持たない職員と同等にできるようにする。
- ・ 障害を持つ一般の人が、連邦政府の各省や機関が提供する情報・サービスを、障害を持たない一般の人と同等に利用できるようにする。

ことを義務付けており、罰則など強制力を持つものである。

同法が影響するのは、連邦政府及び政府からの援助を受けている州政府機関などが調達する製品群である。ただ、調達時のメーカー選定基準として、インターネットのホームページの利用を始め障害者への対応姿勢や同法対応製品の品揃えが問われることも考えられるため、米国企業だけでなく、複写機、ファクシミリなどを製造、輸出販売している日本企業にも同法への対応を検討する動きがでてきている。

音声認識による音声操作や情報入力第 508 条への対応はこれまでマイクロソフト、IBM などによる PC の音声操作に関する取り組みが中心であった。しかし、複写機、ファクシミリ

などを視力障害の人が使うための有力な操作手段として必要と考えられるため、メーカーの関心が高まっている。また、インターネットのホームページなどを利用する手段としても今後音声認識が検討されていくと思われる。

3. 米国リハビリテーション法第 508 条に類似する日本の動き

米国リハビリテーション法第508条に類似する動きとして、日本でも通商産業省が「障害者・高齢者等情報処理機器アクセシビリティ指針」(2000年6月)を、総務省が「障害者等電気通信設備アクセシビリティガイドライン」(2000年7月)などを公開している。これらの中では情報処理機器や電話通信機器に対しての音声操作、音声入力に関しても言及されている。また、1999年にはW3C(World Wide Web Consortium、<http://www.w3.org/TR/patent-policy/>)も障害者のインターネットのホームページ利用を可能にするために、ホームページ制作者向けにウェブアクセシビリティの指針を勧告している。

この他、日本でも1999年5月、郵政省と厚生省により共同開催された「情報バリアフリー環境の整備の在り方に関する研究会」報告書において、「インターネットにおけるアクセシブルなウェブコンテンツの作成方法に関する指針」及びその解説がとりまとめられるなど、インターネットのホームページでの音声読み上げ、音声操作の必要性が高まっている。

今後も、情報バリアフリーのニーズの高まりとともに、その実現手段としての音声認識技術が求められるであろう。

第 2 節 標準化及び権利活用状況

音声認識に関する主な標準化規格としては、電話から音声によってインターネットを利用するための言語規格である VoiceXML、PC や PDA などの情報機器での音声認識を含むマルチモーダルインターフェイスの規格としての Speech Application Language Tags(SALT)、そして欧州携帯電話のネットワークでのクライアント・サーバ型音声認識の規格である GSM DSR(Distributed Speech Recognition)が知られている。

VoiceXML を含む VoiceBrowser の関連特許は、各社申告の特許が W3C のホームページに明示されている。W3C として、標準規格に関する特許に対して課金権利を認めるかどうかの方針が揺れ動いており、現在は特許使用料を無料とする方針で各社と最終調整中である。

SALT に関しては、基本的に特許使用料を無料とすると SALT フォーラムが宣言している。

DSR 規格に関しては、モトローラが特許を保有し、また規格策定元の ETSI (European Telecommunications Standards Institute: 欧州電気通信標準化機構)は妥当な範囲の課金を認める方針である。また、Qualcomm や Philips などにもクライアント・サーバ型の音声認識システムに関する特許があり注意を要する。

日本では(社)電子情報技術産業協会(JEITA)を中心とする音声認識インターフェイスの標準化の検討がようやく始まった。日本語の音声表記等から標準化の検討が進められている。

パテントプール等を使って日本及び海外の音声認識に関する流通特許を調査した結果、流通特許は日本、海外ともに 100 件以上検索された。1990 年以前に出願されたものが多く、内容としては実用化技術に関するものが多い。

第6章 今後日本が目指すべき研究開発、技術開発の方向性と取り組むべき課題

音声認識技術において、今後日本が目指すべき研究開発、技術開発の方向性と課題を提示する前に、一連の調査を通じて、浮き彫りになってきた点を簡単に整理しておく。

1970年代から1980年代にかけて、DTWベースの音声認識で日本は世界をリードしていたが、標準技術がHMMにシフトしてきた1990年代以降、日本はHMMの開発で米国に立ち遅れた。これらが、1990年代の特許出願件数減少の要因になった。

国家的プロジェクトであるDARPAプロジェクトが牽引した米国の事情とは異なり、日本ではHMMによる音声認識技術の製品化の努力は各企業の自助努力に委ねられた。日本の特許出願の傾向を分析すると、審査請求をしないで終わる場合や事業戦略と整合しないで出願だけにとどめる場合も多い。また、米国に比べて、実用化技術の出願・登録が少ない。

計算機の大幅な進歩、HMMによる音声認識シミュレーションツールの登場、音声コーパス（音声データベース）の蓄積、音声認識を必要とするアプリケーションの登場が、近年、音声認識市場が形成されてきた要因となっている。

第1節 研究・技術開発の方向性と取り組むべき課題

1. 高まる実用化技術の開発への取り組みの必要性

今回調査によると、日本は1995年以降、実用化技術の出願、登録が増加し始めた欧米に比べて、実用化技術の出願件数増加の時期が遅れている。これについて、有識者に尋ねたところ、「従来、日本企業は音声認識のコア技術の特許出願に熱心であった。しかし、ビジネスの観点から見れば、実用化技術の特許出願のほうが重要である。」との意見があった。

現在、音声認識の市場が形成されだしている。日本でも今後は実用化につながる音声認識技術の研究がより求められてくるのではないかと考えられる。

2. ユビキタスコンピューティングの時代に適合する音声認識技術の開発

近年の音声認識技術の進展はコンピュータ技術の発展なしには考えられない。このコンピュータ技術のさらなる発展はその存在を意識させずに生活環境の中に溶け込むユビキタスコンピューティング時代の到来を想像させるまでになっている。こうしたコンピュータ技術の今後の発展に適合した音声認識技術の実現形態は、大別して、

- ・サーバ型：全ての音声認識処理をサーバで実行
- ・組み込み型：全ての音声認識処理を組み込み機器で実行
- ・サーバ/組み込み連携型1：
音響処理を組み込み機器内で実行し、マッチング処理をサーバで実行
- ・サーバ/組み込み連携型2：
組み込み機器内での音声認識を基本とするが、音声認識できない場合、音声データ又は音響処理結果をサーバに送り、サーバでマッチングする

の4つが考えられる。日本の産業上の競争優位性は組み込み機器の設計、製造、輸出メーカーが米欧に比べて圧倒的に多いことにある。今後は組み込み機器メーカーのニーズに親和性の

高い実現形態である組み込み型、サーバ/組み込み連携型 1 及びサーバ/組み込み連携型 2 の音声認識技術、話者認識技術の研究開発、技術開発を目指すべきである。

そのためには以下の研究開発テーマが特に重要である。

組み込み機器に搭載可能なエンジンソフトウェア（コンパクトな音声認識、話者認識アルゴリズムと、デジタル信号処理技術、組み込みプロセッサ向けソフトウェア技術との融合技術）

人（声質、喋り方）、環境（場所、時間）、装置への平等性を実現する音響処理及び音響モデル

多言語音声認識の実現に必要な音響モデル、発音辞書、言語モデル

Microsoft の SAPI に代わる、組み込み機器向け小型軽量の日本発 API

サーバ/組み込み連携型 2 における、サーバ側音声認識と組み込み型音声認識の互換性さらに、ユビキタスコンピューティングに接続された組み込み機器をプラットフォームとしたサービスビジネスの開拓も必要である。

3．デジタルデバイド対応の技術の開発

情報通信技術の急速な普及に伴い、日本でもデジタルデバイド（情報弱者）と目される人口が増加している。今後は、デジタルデバイド人口の解消、電子政府の実現に役立つ音声認識技術の研究開発、技術開発を目指すべきである。

そのためには、以下の研究開発テーマが特に重要になってくる。

- ・人工知能、エージェント、ロボット、知識獲得などの他の IT と連携した、「人に優しい」マルチモーダルインターフェイス
- ・IT リテラシー（能力）の低い労働人口が利用可能で、知的生産活動が可能となる音声認識技術

4．高齢者や障害者支援のための技術の開発

国立社会保障・人口問題研究所の「日本の将来推計人口（平成 14 年 1 月推計）」の中位推計では、日本の 2050 年の全人口に対する 65 歳以上の老年人口の比率は 35.7% にまで上昇する見通しである。このように高齢化が進展し、労働力人口が不足してくる中では、高齢者への生活支援や、高齢者の介助業務への支援のニーズが増大してくる。音声認識技術はこれらのニーズに応える手段を実現するための必要技術の一つと考えられている。

この他、今回調査の結果を見ると、障害者・高齢者の電子情報技術の利用を支援する政策的な動きが三極に見られる。米国リハビリテーション法第 508 条や日本の通商産業省の「障害者・高齢者等情報処理機器アクセシビリティ指針」などはその一例である。音声認識技術は、この電子情報技術の利用を支援するために有効であると考えられている。

今後はこうした高齢者や障害者の支援のための技術として音声認識技術の研究がますます求められるようになってくるであろう。具体例としては、障害者が手を使わずに情報端末を使用するために音声による機器の操作を可能にする技術や、手がふさがった状態で音声によって介護機器の操作することを可能とし、障害者・高齢者を介助する家族やヘルパーの介助業務を直接支援する技術などが考えられる。

これらの研究開発は世界に先駆けて日本が取り組むべき課題であり、応用産業で製品化された製品を日本より遅れて高齢化社会を迎える諸外国に輸出することが可能になる。

第2節 その他の音声認識技術を巡る課題

1．特許権の戦略的活用の必要性

日本企業は、将来、特許紛争が発生する可能性も考慮して、

- ・登録特許のライセンス収入ビジネスモデルの確立
- ・クロスライセンスに適用可能な登録特許の取得数を増やす努力
- ・他社との戦略的アライアンスによる特許係争の回避

なども視野に入れ、特許権を戦略的に活用していく必要がある。今回調査では、音声認識に関する特許係争の事例はあまり見受けられなかった。しかし、今後、市場が拡大し、事業を成長させる企業が増えてくると、特許係争が発生してくる可能性を否定できないからである。

日本企業は、特許権を有効に活用する方策を検討したほうがよいのではないかと。今回調査によると、1980年代から1990年代前半に日本に出願された音声認識関連の特許のうち約半分は審査が未請求であった。1995年以降では60%以上が審査請求されていない。有識者へのヒアリング調査によると、「日本企業の中には、学会発表の際に特許出願をルール付けているものや、特許として出願することで技術を公知にして他社の権利化を防ぐものがある。こうしたことが審査請求の少ない理由の一つではないか」とのことであった。単に出願するのみではなく、権利化を行い将来の予想される状況に備えておくことも必要であろう。

2．産業構造に関する課題

近年、音声認識市場成長の条件が整いつつある。これを好機として、音声認識技術を利用した産業が育っていくには、それに相応しい産業構造が必要である。望ましい産業構造が形成されていくためには、以下に示すような様々な機能の協調が有効である。

- ・音声認識アルゴリズム研究機関
- ・音声認識エンジンソフトウェアベンダー
- ・音声認識エンジンコンテンツ（音響モデル、発音辞書、言語モデル）プロバイダー
- ・音声認識技術と市場ニーズを理解し、両者をつなぐ解決策を提案するコンサルタント
- ・音声認識を利用した製品／サービスのシステムインテグレーター
- ・音声認識を利用した製品／サービスプロバイダー
- ・音声認識を利用した製品／サービスの評価機関

音声認識を取り巻く市場や技術の環境は今後めまぐるしく変化していくであろう。従来のような垂直統合的な動きでは、時宜を得た対応が難しくなる可能性がある。水平分業的な事業構造へとシフトして、上記の機能の内、どの部分に焦点を当て、どの部分を他社との提携で補うといった戦略を構築することが日本企業には求められている。

3．企業の研究開発投資意欲低下の問題

本調査の結果、明らかになったように、日本における1980年代から1990年代前半の音声認識の研究、技術開発は大学、エイティアルと共に日本電信電話や大手電機メーカーなどの研究所が担っていた。しかし、最近では景気低迷の影響もあって、企業の研究開発のアクティビティは低下している。研究開発の場を失った企業の研究者が大学に転出して研究開発を継続するケースや、企業に止まるも音声認識の研究開発を継続できないケースが増えている。

こうした状況は最近の登録件数の減少にも現れている。現状を放置しておく、日本の研究開発力、技術開発力が将来、欧米に比べて、極めて脆弱になる恐れもでてくる。

4．ベンチャー企業の育成

米国の音声認識市場において主導的役割をはたしているNuanceやSpeechWorksの1社当たり売上高は数十億円程度である。数十億円という規模は、日本では、大企業が事業化するよりも、ベンチャー企業が手がけるのに適した規模である。音声認識市場の将来的市場規模は大きいものの、現在はまだ市場が立ち上がりかけた段階であり、音声認識市場に参入するには「小さく生んで、大きく育てる」経営方針で臨むことが適しているといえよう。

5．産学官連携国家プロジェクトに関する課題

日本では文部科学省科学研究費、IPA、エイティアルが日本の音声認識技術の育成に大きな役割を果たしてきた。しかし、米国で産官学が連携したDARPAプロジェクトと比べた場合に、音声認識市場の創成に果たした役割は相対的に小さいと言わざるを得ない。

米国の産官学が連携したDARPAプロジェクトに習い、十分な音声認識技術の研究開発経験を持つ人材をリーダーとして、日本企業がより高い産業競争力を有する市場分野に焦点を当て、ビジネス的な必要性が高く、かつ、限定された時間の中で達成可能な研究開発テーマを選択的に設定し、進捗管理、事後評価を徹底しながら産学官連携国家プロジェクトを運営する仕組みを考えてはどうか。複数の研究機関やベンチャー企業も含めた多くの企業が参加し、全体としてまとまった成果を創出する産官学連携の国家プロジェクトが今求められている。

日本企業にも従来の自前主義に拘らず、こうした試みに参加して、研究開発を加速する姿勢や「日本語口述筆記用基本ソフトウェア(Julius)¹」、「擬人化音声対話エージェント基本ソフトウェア」などの国家的プロジェクトの成果を積極的に活用する姿勢が求められている。

6．音声認識技術の開発基盤整備の必要性

今回の調査結果で明らかになったように、欧米と比べ、日本は音声認識技術の開発基盤となるインフラが不足している。

例えば、音声コーパス（音声データベース）やテキストコーパス（文章データベース）は音響モデルや言語モデルを研究開発するための基となる重要なものである。しかし、これを民間の一企業が自前で作成することはコストがかかり過ぎるため難しい。日本にも助成金で収集された音声コーパスやテキストコーパスが存在するものの、その利用は研究目的に限定されており、民間企業の商業目的の利用は制限されている場合が多い。また、これらの利用

¹ 情報処理振興事業協会（IPA）の援助の下に大学を中心に行われた（実施中の）音声認識及び音声対話技術の開発プロジェクト。Juliusの実施時期は1997年～2000年、擬人化音声対話エージェントの実施時期は2000年～2004年（予定）。

に際して著作権が問題となり利用できない場合もある。一方、欧米では、国が関与して音声コーパスやテキストコーパスの有償利用が可能な制度（LDC、ERLA）が整備され、機能している。今後、日本でも外国語対応のニーズが高まると予想されるため、多言語の音声コーパス、テキストコーパスの整備が音声認識技術の研究開発を進めていく上でますます重要になってくる。こうした課題を解決するために、音声コーパスやテキストコーパスの利用範囲を拡大するような方策を検討し、音声認識技術の開発基盤整備を進める必要がある。

また、音声認識市場の育成も重要である。今回調査の結果によると、「ディクテーション、カーナビゲーションや情報コンテンツアクセスなどのアプリケーション分野は民間企業の活力によって音声認識市場が成長していくことが期待できる。しかし、高齢者や障害者のための福祉、介護の分野、デジタルデバイドの解消などの分野は、民間企業に任せただけでは、スムーズな市場の立ち上がりは難しいかもしれない」との声が有識者より聞かれた。

来るべきユビキタスコンピューティング時代のヒューマンインターフェイスとしての音声認識技術の役割に根ざした新産業の育成とそれに関連する雇用の創出のために様々な施策の必要性を指摘する有識者も多数あった。こうした状況に対して、官民一体となって取り組んでいく必要がある。

【お問い合わせ先】特許庁 総務部 技術調査課 技術動向班

TEL : 03-3581-1101 (内 2155) FAX : 03-3580-5741

E-mail : PA0930@jpo.go.jp