2005－SLP－55　(2)
2005／2／4

（招待講演）特許から見た音声認識ビジネスの動向 ～平成１４年度特許庁特許出願技術動向調査分析に見る音声認識の特許動向～

庄境　誠

アブストラクト

長年，研究されてきた音声認識技術の実用化が遅々として進まないのは何故なのか？と叫ばれて久しい．筆者は１９８５年に音声認識に関する研究をスタートさせ，以来，研究開発に携わった後，１９９８年に音声認識ＬＳＩの製品化，２０００年に音声認識ミドルウェアの製品化を実現した．また，２００２年度（平成１４年度）に特許庁により実施された音声認識技術に関する特許出願技術動向調査分析にも関わった．本論文では，音声認識技術の実用化および音声認識技術の特許調査に関わった立場から，音声認識ビジネスの動向に対する私見を述べたい．

(Invited Paper) Business Trends of Speech Recognition from a Patent Point of View - Patent Trends of Speech Recognition in "Japan Patent Office Report of Investigation and Analysis on Trends of Patent Applications and Technologies in Fiscal 2002" -

Makoto Shozakai

Abstract

It has been pointed out for a long time that why commercialization of speech recognition technologies which have a long history of research is delayed. Author of this paper started a research on speech recognition technologies in 1985 and has been involved in research and development to lead to realization of speech recognition LSI product in 1998 and speech recognition middleware product in 2000. Furthermore, the author was involved in the project held by Japan Patent Office to investigate and analyze trends of patent applications and technologies on speech recognition in fiscal 2002. The author discusses speech recognition business trends from both points of view of commercialization and patent investigation of speech recognition technologies.

旭化成株式会社　情報技術研究所

Asahi Kasei Corporation, Information Technology Laboratory

１．はじめに

長年，研究されてきた音声認識技術の実用化が遅々として進まないのは何故なのか？と叫ばれて久しい．筆者は１９８５年に音声認識に関する研究をスタートさせ，以来，研究開発に携わった後，１９９８年に音声認識ＬＳＩの製品化，２０００年に音声認識ミドルウェアの製品化を実現した．また，２００２年度（平成１４年度）に特許庁により実施された音声認識技術に関する特許出願技術動向調査分析にも関わった．本論文では，まず，同調査分析の報告書からいくつかの要点を紹介した後，音声認識技術の実用化に携わる立場から，音声認識ビジネスの動向に関する考察を行う．

２．音声認識技術の特許出願動向に関する考察

本章では，特許庁により実施された，音声認識技術に関する，平成１４年度特許出願技術動向調査分析報告書[1]から必要な図と文章を抜粋する．本調査分析は，1968 年から 2002 年までの音声認識技術の特許動向を対象としている．尚，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（すなわち，公開特許＝登録特許である）ことに注意する．

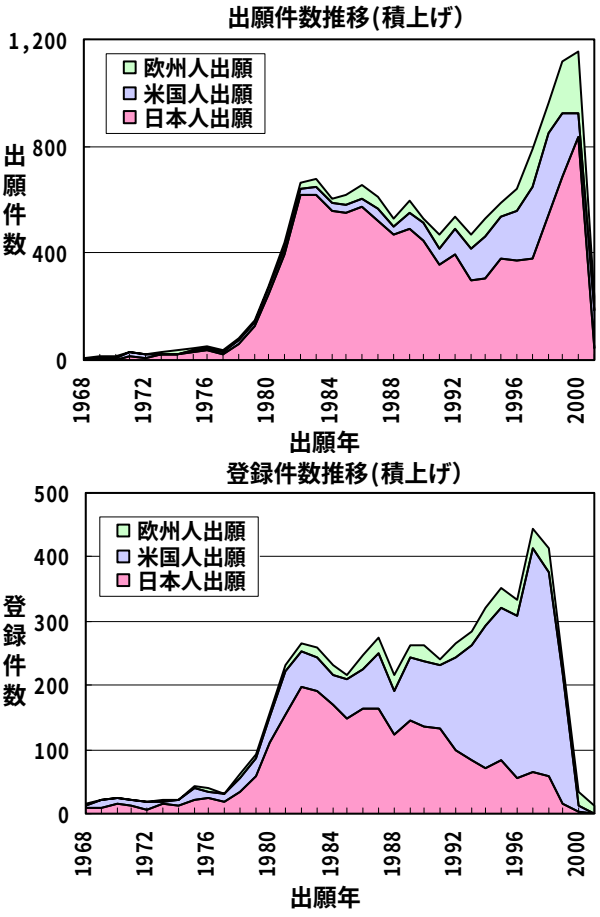

図１　日米欧自国・地域内への出願件数．登録件数の推移
（出展：要約 2-1 図[1]）

２．１　出願件数，登録件数傾向

図１に，日米欧三極の出願人の自国･地域内への出願件数，登録件数の推移を示す．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 年以降再び増加している．登録件数は，1980 年代初頭は日本人による出願が大部分を占めていたが，日本人出願の登録件数は 1982 年をピークに減少している．逆に，1990 年代は米国人出願による特許の登録件数が著しく増加している．実は，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，米国人の出願は 1990 年代の HMM に関するものが多い．日本の HMM に関する特許は，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た．これは，DTW 技術の時代は日本の研究開発が世界を牽引し，HMM 技術の時代は米国が牽引していることを示すものである．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．

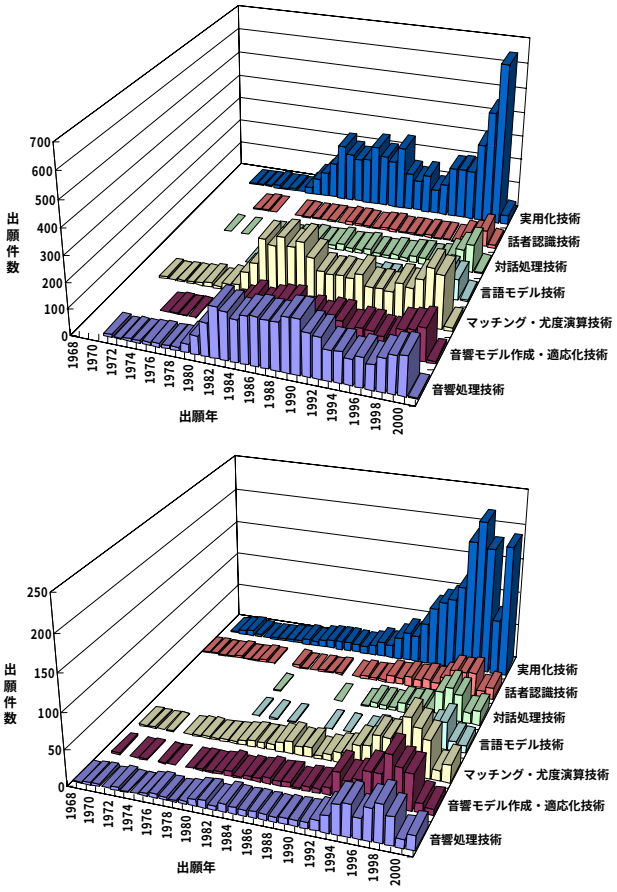

図２　技術区分別の特許件数の推移
（上段：日本人が日本に出願した特許，
下段：米国人が米国に出願し登録された特許）
（出展：要約 2-6 図[1]，要約 2-7 図[1]）

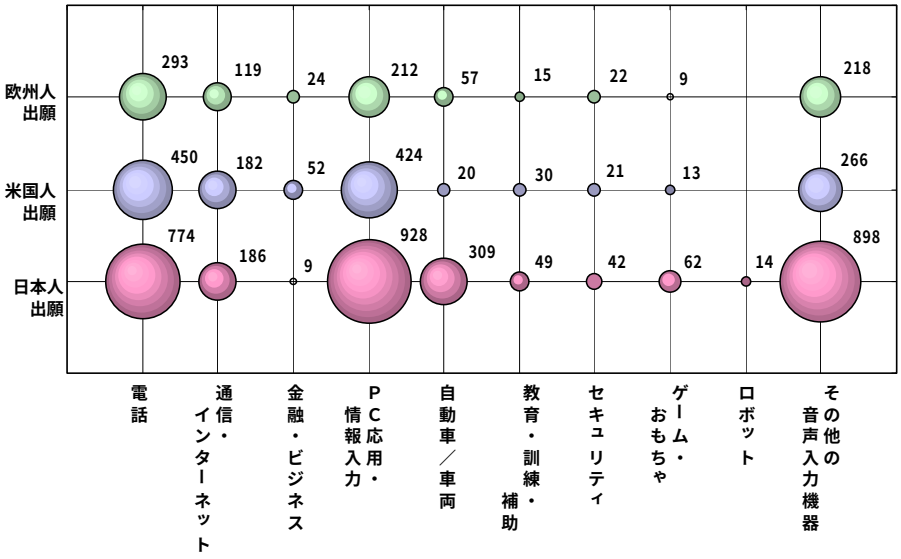

図３　実用化技術の三極出願人比較
（出展：要約 2-10 図[1]）

２．２　三極別の出願人の内訳

日本への出願では日本電気(1159 件)，松下グループ(1096 件)，富士通(799 件)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(305 件)と AT&T グループ(279 件)が抜きん出ている．その他，米国では専業メーカーである ScanSoft(91 件)が健闘している．欧州への出願では，Philips(259 件)が首位にあり，IBM(169 件)，Siemens(159 件)と続いている．尚，IBM，Philips，AT&T グループは日本にも 100 件程度を出願している．

２．３　技術区分別の特許出願件数の推移

図２に，技術区分別の日本の出願件数および米国の登録件数の推移を示す．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登録特許が増加している．米国では，近年は実用化技術に関する登録特許の伸びが著しい．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．尚，対話処理技術，言語モデル技術が 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．

２．４　三極の実用化技術特許出願内容の違い

図３に，実用化技術に関する三極出願人の比較を示す．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での音声認識応用製品の開発が進んでいることを示す．

２．５　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

（１）高まる実用化技術の開発への取り組みの必要性

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．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．

（２）ユビキタスコンピューティングの時代への適合

日本の産業上の競合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／組み込み連携型の音声認識技術の研究開発を目指すべきである．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音声認識利用の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．

（４）高齢者や障碍者支援のための技術の開発

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．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．

（５）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．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．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．

（６）産業構造に関する課題

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，様々な機能の協調が有効である．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，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．

（７）企業の研究開発投資意欲低下の問題

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，ＡＴＲと共にＮＴＴ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と，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．

（８）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（現 ScanSoft 社）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．

（９）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，ＡＴＲ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．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．

（１０）音声認識技術の開発基盤整備の必要性

音声コーパスやテキストコーパス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，著作権が問題となり利用できない場合もある．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．

３．音声認識技術のビジネス動向に関する考察

音声認識技術のビジネス動向について考察する場合に留意しなければならない点は，ビジネスの優劣は必ずしも技術の優劣と合致しないことである．ビジネスは，技術戦略，事業戦略，知的財産戦略が三位一体となって初めて成功の確率を高めることが出来る．音声認識技術の実用化が進まない原因を分析するためには，技術，事業，知的財産のいずれかに解決されていない課題があると考えるべきである．ここでは，技術，事業，知的財産の３つの観点から音声認識ビジネスの在るべき姿について様々な軸に沿って考察してみたい．

３．１　技術

音声認識が，技術として完成されたということに同意する人はほとんどいないであろう．音声認識技術が進展したといっても，今実現されているのは，本来，音声認識技術に期待されている実環境性能と比べたら限定的であるのは間違いない．ここでは，実環境における技術が未完成であるとの立場から，現状を分析する．但し，筆者も長らく音声認識技術の研究に携わっている．絶対的１００％性能への道がどれほど困難であるかはよく承知している．しかしながら，音声認識機能を利用する人が，体感として１００％性能と感じてくれることが音声認識普及の鍵であるとの信念から，敢えて「体感１００％性能」という言葉を以下で用いることにする．まず，指摘すべきは，音声認識のニーズに関するものである．

◎ニーズ軸

○ハンズフリーニーズ：何らかの理由で手による入力が困難な状況での新たな入力インターフェイスとしてのニーズである．後述の環境要因，人間要因の困難さから小語彙に限定される場合が多い．

○大語彙情報入力ニーズ：画面が備わっていて，手による入力が可能であるが，体感１００％性能の音声入力の方が容易な場合のニーズである．大語彙の情報入力が必要な場合であるが，人間要因の複雑さから低雑音環境に限られる．

次に，挙げる軸は，研究に取り組む姿勢である．

◎研究指向軸

○人工知能指向　体感１００％性能は達成できないことを前提に人工知能的アプローチによりシステム的に解決しようとする指向である．

○センサー指向　音声認識をセンサー技術と捉え，あくまで体感１００％性能のセンサー実現を目指す指向である．

音声認識技術は音声入力を可能とするインターフェイス技術である．残念ながら，音声認識技術単独で最終システム製品が構成されることはない．最終システム製品にとって，音声認識技術は一構成要素でしかない．音声認識機能を最終システム製品に採用したいと希望するシステムデザイナーは体感１００％性能のインターフェイス技術を期待するのが常である．しかし，現状の音声認識技術はどう誤認識するかさえ予測できない技術レベルにある．体感１００％性能から乖離すればするほど，システムデザイナーにかかる負担が重くなる．そこまでして音声認識機能を採用することに疑問を呈するようになり，やがて音声認識技術に対する失望へと変わる．音声認識技術は一体どれほど多くの失望を人々に与えてきたことであろうか？最早，失望の歴史を繰り返してはいけない．音声認識技術に対する市場からの要求仕様は，単純に体感１００％性能の自立したインターフェイス技術である．結局，研究開発の本質は体感１００％性能を目指すということではないか？音声認識の研究開発に携わる人は初心に立ち返り，自立すべき時期に来ている．音声認識を必要としている最終システム製品の多くが実環境で使用されるということを考えれば，なぜ実環境における音声認識率は体感１００％性能にならないのかという課題に切り込んで行くべきである．実環境性能の研究は，泥臭い課題解決の連続である．科学的研究が成立しにくい局面も多い．しかし，そのことを避けている限り，音声認識の置かれている現状を打破することは出来ない．体感１００％性能を阻む要因としては以下の２つが考えられる．

◎要因軸

○環境要因　音声認識機能を搭載した最終システム製品が使用される音環境は多くの場合，時不変ではなく，絶えず変化する．個々の自動車，住宅または屋外には，それぞれ固有の音環境がある．それらの音環境の把握技術や制御技術を確立しない限り，体感１００％性能を実現することはできない．

○人間要因　音声認識機能を搭載した最終システム製品は多くの場合，様々な人間が使用する．使用する人間が増えれば増えるほど，固有の声道特性，喋り方，方言，感情表現の変動が増える．それらの人間要因の把握技術や制御技術を確立しない限り，体感１００％性能を実現することはできない．

次の重要な軸は，音声認識機能の利用者の意識（コンシャスネス）の軸である．

◎コンシャスネス軸

○コンシャス型　利用者が何を喋るかを明確に意識しており，目的的に音声入力する場合で，以下のように分類される．いずれも，リアルタイム性が要求される．

・コマンド入力（例：ハンズフリー電話の音声操作）：小語彙の離散単語認識や連続数字認識が代表的である．語彙外発声の問題は回避し易い．

・情報入力（例：カーナビゲーションシステムへの目的地登録）：大語彙の連続単語認識が必要であり，地名や施設名の認識が代表的である．話し言葉に起因する語彙外発声の問題を如何に回避するかがポイントとなる．

・テキスト入力（例：ディクテーション）：最も難しいタスクであり，語彙外発声の問題が深刻である．

○アンコンシャス型　利用者が喋る内容は目的的だが，認識されることは想定していない場合である．リアルタイム性は要求されない．

・アノテーション（例：ＴＶ番組の書き起こし）：語彙外発声の問題が起こらないように，大規模な言語モデルを用意する必要がある．

音声認識機能を持つ最終システム製品において，利用者が何を喋って良いか分からないという状況が発生するとすれば，その製品デザインは失敗と言わなければならない．音声認識機能の利用は利用者とシステム（間接的にはデザイナー）との間で，語彙に関して容易に合意が形成できるようなタスクに止めるべきであり，その合意が形成できる範囲で音声認識機能は実装されるべきである．

図４には，性能および実環境度と難易度等高線の関係を描いた．普及型の最終システム製品は低価格でなければならず，実環境度の高い環境で利用される傾向が強い．性能と実現コストは，単調増加の関係が成り立つことが多いが，コストパフォーマンスを大きく改善するアルゴリズムの発見は大きなブレークスルーであり，事業的価値が高い．普及型の最終システム製品の音声インターフェイスを実現することが研究開発の目的で有れば，トレンド１を指向すべきである．高機能の音声認識技術の研究の前に，低機能であっても体感１００％性能を実現するための研究開発を加速すべきである．例えば，音声認識機能を必要とする最終システム製品では，ハンズフリーでの位取り表現を含めた連続数字入力のニーズがかなり高い．しかしながら，実環境での連続数字認識は，体感１００％性能に遠く及ばない．誰の声でも連続数字を入力できる機能こそまず実現されるべきであろう．一方，高機能，低雑音の音声認識機能を実現することが研究開発の目的で有るならば，トレンド２を指向することになろう．

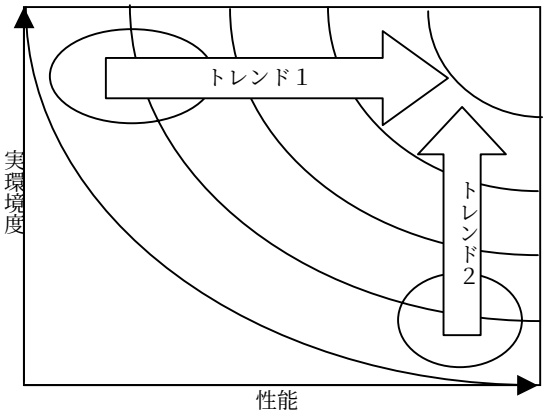

図４　難易度等高線

３．２　事業

ビジネスの成功は，技術ではなく，ビジネスモデルに依るとの指摘がある．音声認識の研究開発を行う場合も，ビジネスモデルに整合したミッションの明確化という軸がある．

◎ミッション軸

○専属型　自社グループの既存ビジネスに貢献するための研究開発の場合である．

○独立型　自社グループの枠にとらわれず，研究開発した音声認識技術そのものを事業化する立場である．音声認識ビジネス自体の採算性を問われる．

大企業の場合，一般的に社内で良好な評価を受けるのは，最低でも年間売り上げ１００億円である．この規模は，一人当たりの年間売り上げを２千万円とすると，５００人が従事することを意味する．残念ながら，体感１００%性能を実現していない音声認識技術は，単独で１００億円の年間売り上げを市場から得られていない．従って，ほとんどの企業において実施されているのは専属型であり，独立型のミッションを掲げている会社は僅かである．次に，考察する軸は，商材軸と収入軸である．

◎商材軸

○エンジン指向　音声認識アルゴリズムを実装したエンジンソフトウェアおよび関連ソフトウェア（通常，バイナリコード）を主たる商材とする．

○コンテンツ指向　アプリケーション特化の音響モデルや言語モデルなどのコンテンツを主たる商材とする．

◎収入軸

○開発費主体　音声認識を最終システム製品に組み込む場合のソリューション提供に対する対価（開発費）を主たる収入源とする．最終システム製品の開発段階での収入源である．

○ロイヤリティ主体　音声認識を最終システム製品に組み込む場合の継続的なライセンス費を主たる収入源とする．最終システム製品の量産段階での収入源である．

３．３　知的財産

音声認識は，様々なアルゴリズムの組み合わせで実現されており，物質や装置に関する特許とは異なり，他者の登録特許を回避する手段が多彩である．また，ソフトウェアのバイナリコードの形で具現化され，流通されるため，侵害の検証は非常に困難である．一般に，３年程度とされる最終システム製品の製品寿命は，ソフトウェアの権利の侵害の検証に必要な歳月に対して十分に長いとはいえず，アルゴリズムに関する登録特許を積極的に行使する場面は多いとは言えない．また，特許出願から登録まで５年程度の歳月を要する一方で，技術革新が毎年進むという面もある．知的財産権の出願目的についてよく検討しておく必要がある．アルゴリズムそのものよりも，侵害の検証が容易な製品形態やビジネスモデルに係る知的財産権の方が事業的価値は高いといえる．

◎出願目的軸

○権利取得目的　権利の行使の如何に関わらず，権利を取得することが目的の場合である．将来のクロスライセンスの根拠が目的である場合も多い．

○公知化目的　権利取得が目的というよりは，他者に権利取得されないことが目的の場合である．

企業においては，年間ノルマをこなすための出願や学会発表のための手続きとしての出願である場合が多い．このようなケースでは，先行特許調査をほとんどしない．従って，公知技術の存在を知らずに出願している場合も散見される．一体過去にどのような内容の特許出願がされ，何が登録され，何が公知なのか，把握されていない現状がある．これでは，知的財産戦略など組立てられない．

４．おわりに

ビジネスの優位確率は，技術，ビジネスモデル，知的財産のそれぞれの優位確率の掛け算である．技術，事業，知的財産の総てにおいて優位確率が高くなければ，ビジネスの成功は危うい．音声認識の実用化が進まないのは，三位一体で考えなかった過去にも原因がある．音声認識技術の研究開発者は，ビジネスモデルや知的財産についても検討すべきであるし，その道の専門家の力を借りるべきである．技術の優位性は，認識性能が高いということではなく，事業や知的財産の専門家から見ても魅力的なものでなければならない．音声認識の実用化に携わる人は，これらの観点を再確認すべきではなかろうか．その上で，信頼される音声認識技術を実現し，世の中で言われる「死の谷」を越えようではないか．

５．参考文献

[1] http://www.jpo.go.jp/indexj.htm より「音声認識」で検索