# CORPORATE ACTIVITIES IN SPEECH RECOGNITION AND NATURAL LANGUAGE: ANOTHER "NEW-SCIENCE" – BASED TECHNOLOGY


KONSTANTINOS KOUMPIS

*Dept. of Computer Science, University of Sheffield, 211 Portobello St*
*Sheffield, S14 DP, United Kingdom*
*email: K.Koumpis@dcs.shef.ac.uk*

KEITH PAVITT

*SPRU, Mantell Building, University of Sussex*
*Brighton, BN1 9RF, United Kingdom*
*email: K.Pavitt@sussex.ac.uk*



## Summary

We have used data on patents and publications, and from an Internet-based survey, to analyse corporate technological activities in Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) technologies. Two distinct clusters of firms exist: large firms mainly in telecommunications, desktop computing, and consumer electronics; and small firms specialising in speech technologies. The small specialised firms depend heavily on nearby universities and public research institutes, and to some extent on nearby large firms; their relations with the large firms are complementary as well as competitive. Similar patterns can be observed in other, recently emerging, "new-science" – based technologies.

Integration between ASR and NLP has so far been weak with the two research communities functioning more or less independently, with the former progressing more rapidly than the latter. Having built technological capabilities in ASR and NLP with a small proportion of their corporate technological resources, the large firms have two options depending on the rate of progress in these technologies (especially NLP) in the future. If it is high, substantial investments (including those in complementary technologies) could open up massive market opportunities. If it is low, modest investments will allow the exploitation of niche markets.



*Keywords:* automatic speech recognition, natural language processing, patents, publications, corporate strategies, spin-off companies.

## 1. Introduction

The computerisation of business and the automation of many office and home activities has increased the demand for interfaces based on spoken language. In the early stages, Automatic Speech Recognition (ASR) and Natural Language Processing (NLP) were approached by corporations through extensive exploration and experimentation, rather than through targeted product development; it has taken many years to reach commercialisation of related products and services. In mimicking the most basic means of human communication, ASR and NLP are potentially fundamental technologies. If they achieve radical improvements in performance per unit price, they could make a major and lasting impact on several markets and on the environment people live and work in (Freeman, 1984). Although it is a field dominated by U.S. firms[1], language diversity creates new opportunities for non-English applications world-wide[2].

ASR and NLP can be considered part of a wider category of "new science" – based technologies that have emerged through the combination of older sciences with the rapidly expanding capacities of information technology to store, manipulate and transfer vast quantities of digital information. Other examples of "new-science" – based technologies are computational chemistry, computational fluid dynamics, geographic information systems, remote sensing and neural networks (Mahdi & Pavitt, 1997). ASR and NLP are techniques that provide new functionalities beyond the scope of conventional computer or telephony interfaces. In this paper, we trace how firms approach a potentially radical new technology, from laboratory discovery in the 1950s, through two decades of uncertainty and experimentation in the 1970s and 1980s, to the beginnings of successful commercialisation in the late 1990s.

---

[1]When examining the US technological position the American Council on Competitiveness gave an A grade to speech recognition in computers. In the same class were database systems, biotechnology, jet propulsion, magnetic information storage, pollution reduction, software, vision in computers and computers generally (Stewart, 1991).

[2]The potential applications of ASR and NLP are numerous. Our research has been focused on those providing natural interactivity and accessibility of digital services through dialogues, understanding of messages and communicative acts, unconstrained language input and keyboard–less operation. Developments in fields such as speech synthesis, coding, speaker verification/identification or language generation/identification are very important but outside the scope of this study.

We do this through a combination of qualitative analysis of the evolution and technological trajectories in ASR and NLP before the recent market expansion (section 2), and quantitative analysis using three sources of data: patents, publications and survey results (sections 3 and 4). The mutually consistent results of these approaches confirm that ASR and NLP are at the early and exploratory stages of new product development, and show that most of the technological activities are performed by an often complementary combination of small specialised and large multi-technology firms. The framework of Mitchell & Hamilton (1988) - distinguishing knowledge building, strategic positioning and business investment – is particularly useful for understanding the activities of the large firms. The complementarities between the two sets of firms reflect to some extent the considerable uncertainties about the rate and direction of future developments in the field (Freeman, 1991). In the final section 5, we identify the strengths and weaknesses of our approach and possible avenues for future research.

# 2 Developments and Potential Applications in ASR and NLP

## 2.1 Some Recent History

The goal of ASR is to convert human speech into a string of text that represents what a person is saying. This is a very difficult task, partly because the field is motivated by the promise of human-like performance under realistic conditions. Solving such a real-world problem requires a thorough understanding of many heterogeneous disciplines, including digital signal processing, classification or pattern matching, and linguistics. ASR has so far gained some commercial success due to demonstrable increases in productivity by greatly assisting human operators or by replacing the human element altogether. A speech interface in users' own language is a very natural, flexible, efficient, and economical form of communication. The major areas of commercial application of ASR include information inquiry, dictation, personal computer interfaces, automated telephone services, and special purpose industrial systems. The miniaturisation of devices poses an obvious challenge for speech interfaces, while the provision of services at a distance offers opportunities for language-mediated interactions.

Although the roots of ASR can be found in the 19th century (Kurzweil, 1997), researchers began to use computers for ASR in the 1950s. The first systems built were brittle and primitive. Technology was rudimentary and there were no systems capable of understanding speech in near real time. Later on, recognizers got credible performance for restricted tasks (such as isolated digits spoken by a single speaker). The 1970s were notable for significant research efforts (Reddy, 1974). The Japanese scientists F. Itakura, H. Sakoe and S. Chiba introduced dynamic programming to compute optimal nonlinear time alignments, a technique that quickly became the standard. Itakura also developed an influential analysis of spectral-distance measures, a way to compute how similar two different sounds are. His system demonstrated an impressive 97.3% accuracy on two hundred Japanese words spoken over the telephone. Bell Labs also achieved significant success (97.1% accuracy) with telephony bandwidth speaker-independent systems - that is, systems that understand voices they have not heard before. IBM concentrated on the Markov modeling statistical technique and demonstrated systems that could recognize a large vocabulary.

Despite the optimism of researchers at the time, and the prototypes built, commercial speech-recognition products remained elusive. At the end of the 1970s available products could recognize small vocabularies spoken with pauses between words and they ranged from Heuristics' $259 H-2000 Speech Link, to $100,000 speaker-independent systems from Verbex and Nippon. Other companies, including Threshold, Scott, Centigram, and Interstate, offered systems at prices between $2,000 and $15,000. Speech recognition started to be widely used for collect calls. Bell-Northern Research - BNR (now Nortel) in the middle 1980s, launched one of the first telephone-based commercial systems by recognising customers' 'yes/no' answers during a telephone operator service. Another commercial turning point came up in 1992, when AT&T introduced a five-term speech recognition technology into its nation-wide long-distance network. The 1980s saw the commercial field of ASR split into two fairly distinct market segments. One group - which included Verbex, Voice Processing Corporation, and several others - pursued reliable speaker-independent recognition of small vocabularies for telephone transaction processing. The other group, which included IBM and two new companies - Kurzweil Applied Intelligence and Dragon Systems - began offering the first affordable commercial systems for creating written documents by voice on PCs.

The most important technological impetus came from the establishment of Hidden Markov Models (HMM) and Artificial Neural Networks (ANN) as the dominant technologies. HMM is currently the most successful technique in pattern matching for large vocabulary, speaker independent and continuous speech recognisers. Although ANNs themselves have not been shown to be effective for large scale continuous speech recognition systems, neural network techniques recently complemented traditional approaches to improve performance of state-of the-art systems (Bourland & Morgan, 1994). One of the most important external driving forces for the improvement of spoken language interfaces has been the cost reduction of processing power and memory capacity in computers. Processing power is critical, as it affects the speed of the analys and pattern-matching by a computer. Huge memory is necessary but clearly not sufficient without the solving of difficult algorithmic issues.

NLP belongs to the cognitive sciences and overlaps with the field of Artificial Intelligence. It can be used to augment speech recognition to provide speech understanding within specific application domains. This technology takes a string of words and parses out the vital elements, such that the computer can extract meaning in natural languages (semantics) such as English or Greek[3]. A growing number of research groups are discovering the potential of large-scale linguistic resources such as computer readable dictionaries, tagged recorded speech and bilingual texts. Developers of this technology deal with complex cognitive tasks such as information retrieval, machine-assisted translation, grammatical and stylistic analysis, natural language interfaces for databases, automatic localisation of software and its documentation. The special attraction of computational linguistics lies in the combination of methods and strategies from the humanities, natural sciences, and engineering. The problem of dialogue modelling remains a very difficult research field.

[3]The number of natural languages in current use is estimated to be around 6,500.

Current applications simulate a natural dialogue by careful expert programming of explicit grammars and dialogue flow. A number of skills such as programming, user interfaces design, and integration with the audio hardware is essential. In order to achieve accurate recognition, it is not only necessary to recognise sound units (phonemes, syllables, or words) but also to have an optimised language models. Furthermore, without restrictive grammars (which are not realistic for real speech), even the best speaker-independent systems currently give about 20% word error and about 80% sentence error (Bourland & Morgan, 1994). This is far from what one would require in a dictation system, for instance, it means that the fields of ASR and NLP should gradually converge. Therefore in the rest of this paper we shall refer to them using the term "ASR + NLP". The objective of ASR + NLP is properly speech understanding, not simply correct transcription of words in a spoken message. Systems, therefore, should be evaluated in terms of their ability to respond correctly to spoken messages about pragmatic problems. Since these systems combine many different bodies of knowledge, their development is complex and costly.

The U.S. Defense Department's Advanced Research Projects Agency (DARPA) provided the necessary initial impetus to expand ASR + NLP by starting funding of basic research in 1971 through its Speech Understanding Research (SUR) Programme. DARPA revisited speech technology as part of its Strategic Computing Initiative in the mid-1980s by funding research in five prestigious groups (at Carnegie Mellon University, SRI International, Bolt Beranek & Newman (BBN) Technologies, MIT Lincoln Labs and Systems Development Corp.). This initiative led to dramatic advances in systems that could understand speech in constrained environments (DARPA, 1997). The results of DARPA's current efforts are a range of speech understanding technologies that are being adopted in both commercial and military systems[4].

The DARPA programme has also featured an annual competitive evaluation of systems on common test corpora, which turned out to be highly successful in stimulating rapid algorithmic improvements by attracting significant international participation (Rudnicky et al., 1994). The purpose of this type of

[4]For example, a new telephone information access service 'VoiceBroker' is offered by the securities brokerage firm Charles Schwab. Schwab devoted more than 2 years to adapting DARPA technology to the brokerage industry.

evaluation effort differs slightly depending on who needs the evaluation: a researcher will want to know if his or her technology was improved and a funding agency if it meets the goals set for it. During the 1990s, the programme has concentrated on combining speech recognition with natural language understanding technologies, in order to create systems that are able to conduct interactive dialogues with users to complete transactions within specific application domains. DARPA has also funded research in NLP by sponsoring evaluations of information extraction systems known as Message Understanding Conferences (MUC).

## 2.2 Potential Applications

Figure 1 below illustrates the progress in this field during the last twenty years, in terms of speaking style and size of vocabulary. Communication systems using speech technologies can be classified (Kamm et al., 1997) in *human-computer communication* and *computer-mediated human-human communication* applications. Many of the first category applications are already available. These include personal calendars and files, stock market quotes, business inventory information, product catalogues, weather reports, restaurant information, movie schedules and reviews, and train or airline schedules. Applications in the second category include spoken language interfaces for voice calling, retrieving and sending email/voicemail, paging and faxing, and for real time translation of telephone conversation. Apart from supporting remote access that is both hands and eyes-free, language interfaces for these applications can also provide applications that are difficult or impossible with touch-tone inputs or other modalities.

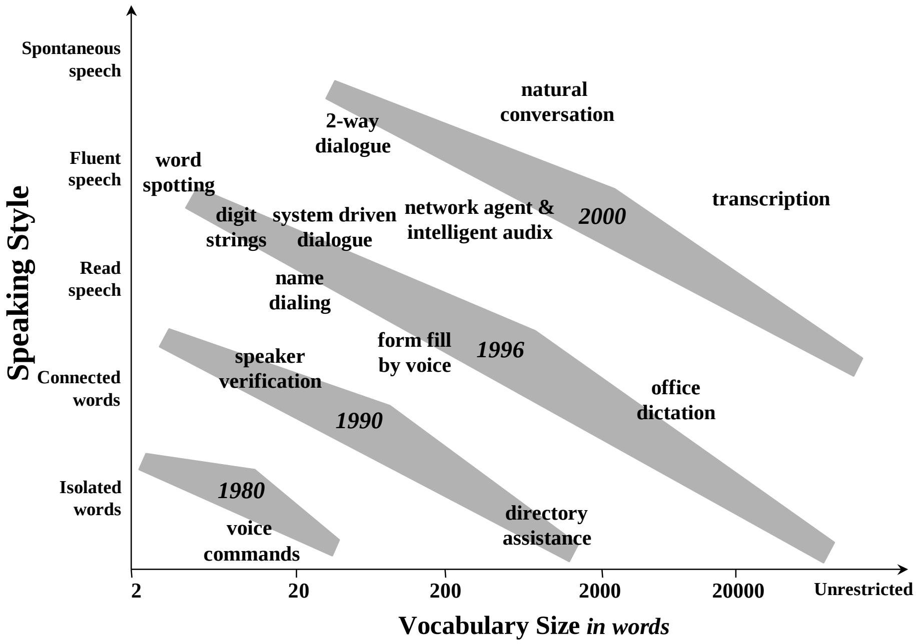


**Fig. 1 Progress in ASR + NLP Technologies,** Source: Jayant N. *Multimedia Communications*, Plenary Session, IEEE ICASSP 98.

More generally, ASR + NLP systems can be classified in *computer applications* and *embedded systems*. Table 1 refers to some key applications being deployed, piloted or developed in these categories. The flexibility of language technology makes it possible not only to develop mass market applications, such as dictation products, pagers, digital answering machines and automated living systems, but also complex high-end vertical market applications such as navigation systems for the automotive industry and clinical reporting systems for the medical market.

**Table 1. Potential Applications based on ASR + NLP**

| **Computer Applications** | **Embedded Systems** |
|---|---|
| Voice typewriter (dictation) | Voice dialling cellular or home phones |
| Education (e.g. language training) | Booking/purchase services on phone |
| Home entertainment | Personal assistants, organisers |
| Government, insurance, banking, finance, accounting | Encoding command and control of manufacturing process |
| Personal telephony (e.g. telecom assistant, data access) | Translation and foreign language issues |

As speech recognition products become more accessible and affordable, the range of user industries is growing rapidly. Currently, four vertical markets – desktop, finance, travel and telecommunications – are providing the best opportunities. Firms in these sectors are particularly interested, because these technologies are regarded as a means to reduce cost and extend the variety of provided services for market differentiation. While the contours of market demand are becoming clearer, the future of applied ASR + NLP will be determined by the growing need for user-friendly software. For instance, if the system's dialogue ambiguities confuse the users - or if the system does not recover gracefully from errors - the whole application can be a failure, even if recognition is satisfactory. It is necessary for spoken language technology to offer clear benefits to both service providers and users in order to be deployed successfully in the marketplace.

Thus, speech based technologies are passing from the invention phase to the innovation phase, but many uncertainties remain, whether technical (e.g. reliability, integration within bigger systems) or commercial (e.g. users' response to new products and services). Adapting the framework proposed by Mitchell & Hamilton (1988), we show in Table 2 the fields of ASR + NLP that in the late 1990s are at the stage of knowledge-building, strategic positioning and potential commercial applications.

**Table 2 Technological positioning in ASR + NLP in the late 1990s**

| **Area of Technology** | ***Knowledge Building*** ***(Exploratory Topics)*** | ***Strategic Positioning*** ***(Research Focus)*** | ***Business Investment*** ***(Potential Applications)*** |
|---|---|---|---|
| **ASR** | acoustic modeling<br>noise cancellation<br>front-end design<br>adaptation<br>rejection | APIs /tools to develop application<br>support of several languages<br>support of several platforms<br>detecting and recovering errors<br>portability | operating system enhancements<br>telecommunications applications<br>hands-free terminals<br>car applications<br>medical applications |
| | task/speaker variability<br>evaluating performance | | |
| **NLP** | language modeling<br>corpus design and collection<br>dictionary building<br>dialogue building | | |

## 3. Data Collection

Recent advances in computing power have made quantitative data on scientific and technological activities an increasingly powerful tool for analysis and action. (Freeman, 1987; van Raan, 1988; Patel & Pavitt, 1996). As in most other economic and social fields, no single quantitative indicator is perfectly satisfactory. But the combined use of a number of imperfect indicators can considerably improve our understanding. We shall combine here analysis of data on patents and on scientific publications, together with the results of an email survey.

### 3.1 Patents

Patents are frequently used as measures of corporate technological activities (Narin et al., 1987; Business Week, 1993). 65% of the participants in our Internet survey (see 3.3 below) said that their companies sought to patent. US patenting data have the particular advantage of reflecting rigorous examination procedures and the prospect of access to a large and technologically progressive market. They are also easily accessible and since the mid-1980s have begun to reflect advances in software technology. In spite of their inevitable imperfections[5], we shall see that they can provide unique insights into the patterns of innovating activities in radically new technologies. We shall assume that companies are competent in a given technical field when they are granted five or more patents between 1976 and 1998. Patents are often granted under the names of subsidiaries and divisions that are different from those of their parent companies. Consolidating patenting under the names of parent companies has been done manually and shown with the thicker borders in Table 3.

The patenting data were collected using the Patent Bibliographic Database[6], which is freely searchable and contains all U.S. patents issued from 1/1/1976 onwards. Especially for the ASR we have counted separately the patents before and after 1988, since Hidden Markov Models (HMMs) and Artificial Neural Networks (ANNs)[7] became dominant technologies after this year. Although every attempt to solve incompatibilities

---

[5]See OECD, 1993

[6]Provided by the US Patent and Trademark Office at http://patents.uspto.gov/

[7]The ANNs appeared in Applied Science and Technology Index for the first time in 1987 (Howard, 1987) while the theory and applications of HMMs in ASR became widely known after the work of Lee (1989).

with assignees' names has been made, there may still be minor errors in the total number of patents recorded for each company.

| Patent Data Collection Method |
| --- |
| The classification of patents may be a source of problems[8] especially in a fast-evolving and multi-technology field. Therefore, the identification of companies patenting in ASR and NLP started by making the following set of queries on the Patent Advanced Search Page of U.S. Patent Office:<br><br>• abst/"speech recognition" OR "voice recognition"<br>• abst/"language processing" OR "language understanding" AND NOT ("speech recognition" OR "voice recognition")<br><br>By this technique we managed to find all assignees even if they have been granted only one patent.<br>In order to count the total number of patents of each firm the query is:<br>• ISD/1/1/76->12/31/98 AND AN/"Company Name" |

## 3.2 Scientific Publications

Scientific publications also provide important clues about the development of scientific and technological fields in public research institutions and corporations[9]. There are many journals where research achievements in ASR or NLP can be published world-wide. Relevant publications in the *SCI* database provided by the *Institute for Scientific Information* were identified though BIDS using the procedure described in the box below.

A few of those journals can claim to be devoted exclusively to these technologies. As the most important references, the *IEEE Transactions on Speech and Audio Processing*, *Speech Communication*[10] and *Journal*

[8]Most of the patents in ASR and NLP are found in CLASS 704 (Data Processing: Speech Signal Processing, Linguistics, Language Translation and Audio Compression/Decompression). Patents can also be found in CLASS 381, (Electrical Audio Signal Processing Systems and Devices), CLASS 340 (Communications: Electrical) CLASS 455 (Telecommunications), CLASS 379 (Telephonic Communications) CLASS 707 (Data Processing: Database and File Management, Data Structures, or Document Processing) CLASS 364 (Electrical Computers and Data Processing Systems) CLASS 706 (Data Processing: Artificial Intelligence) etc.

[9]See Nelson (1959); Rosenberg (1990) and Hicks (1995).

[10]Instead of BIDS, data have been collected from the site of ELSEVIER's Speech Communication Online, for the period 1994-98 at http://www.elsevier.nl/locate/specom

*of the Acoustical Society of America (JASA)* were examined for ASR; and the *Computational Linguistics*, *Communications of the ACM* and *Journal of Artificial Intelligence Research (JAIR)*[11] for NLP.

In order to facilitate co-operation, particularly between companies and researchers, conferences are held to bring together software developers, representatives of government agencies and researchers to exchange views on speech technologies and to discuss ways of future collaboration. The most important of these are the *IEEE ICASSP (International Conference on Acoustics, Speech and Signal Processing),* the *Eurospeech (European Conference on Speech Communication and Technology),* and the *ICSLP (International Conference on Speech and Language Processing).* The ICASSP is held annually, whereas the other two conferences are held bi-annually. In fact, some groups and the majority of corporate researchers publish only in ICASSP.

**Publications Data Collection Method**

The data were collected using the advanced search tool of the BIDS ISI Service (www.bids.ac.uk), BIDS is the best known and most used bibliographic service for the academic community in the UK, providing access to key databases covering subjects from science, engineering and medicine to economics, politics, education and the arts. Three multi-disciplinary citation indexes and an index of conference proceedings are provided. The data are supplied and owned by the Institute for Scientific Information Inc.
In order to find authors' corporate affiliation or address, the Science Citation Index (SCI) was used. The respective queries are:

- speech & recognition , voice & recognition (in the title or abstract)
- language & processing, language & understanding (in the title or abstract)

Whenever two or more authors either had a joint paper or they were from different companies/universities the publication entry was "shared" among all authors.
Notes:
Collaborative papers within the same organisation are not counted separately.
For multinationals only one country or state is mentioned.

For the ICASSP Proceedings the Index to Scientific & Technical Proceedings was used, which allows search only in titles while providing name and affiliation for the first authors only. This limitation of the search engine restricts the number of results. Data for 1984, 1989 and 1995 are not available by BIDS.

## 3.3 Internet-Based Survey

A specially designed questionnaire was put on the world wide web and could be accessed via Internet browsers. This allowed respondents to fill it in and have their answers delivered to us by email. We informed the industry experts by sending an introductory note about the scope of research and the confidentiality issues. We offered participants access to the aggregate results of the research. All major companies were

[11]Instead of BIDS, data have been collected from the site of Journal of Artificial Intelligence Research for the period 1993-97 at http://www.jair.org/

informed about the survey via email, while several contacts were also made directly during conferences and scientific meetings. Over a three-month period the survey was repeatedly announced in the most relevant electronic newsgroups[12]. This type of data collection reflects the opinions of only those Internet users who have chosen to participate, and over a relatively short period. Although the results cannot be assumed to represent the opinions of speech industry as a whole, they offer very interesting indications that complement the other data sources.

There were 53 industrial participants in the survey from 51 different firms. Our responses are from researchers in commercial sites, so the data under-represents research done in the universities. Our respondents have medium-high professional status: 55% were senior researchers or developers, 30% presidents, directors or VPs of R&D and 15% product/marketing managers. Given the these proportions, the data are a good representation of those responsible for product development. The high percentage of such researchers and developers should not be assumed as a drawback for strategic planning analysis. Nonaka & Takeuchi (1995) point out that middle managers are the knowledge engineers of the *knowledge-creating* company, converting tacit images and perspectives into explicit concepts.

## *4. What the Evidence shows*

### 4.1 A New Science-Based Field

Taken together, the data collected confirm that technological opportunities in ASR + NLP have developed only recently, more rapidly in the former than the latter, and have been based on earlier scientific activities. In Figure 2, we show the patenting trends in both technologies. Data are available since 1976, but it took assignees almost two decades to increase from 4 patents in 1976 to 142 patents in 1998, with major increases beginning in the late 1980s. Patenting in ASR is far higher than NLP.

[12]These newsgroups are comp.speech and comp.dsp (Speech & Digital Signal Processing), comp.ai.nat-lang and sci.lang (Artificial Intelligence/Natural Language).

Figure 3 below shows the number of papers on ASR published by ICASSP during the period 1982-98. By comparison with Figure 2, it is clear that ASR was an active field of research well before its results began to appear in patenting data. This reflects a long process of evolution instead of a revolution. For HMMs - the dominant technology for finding the invariant information in the speech signal - neither the theory nor their applications to ASR are new. Markov defined the basic structure to describe the stochastic process in

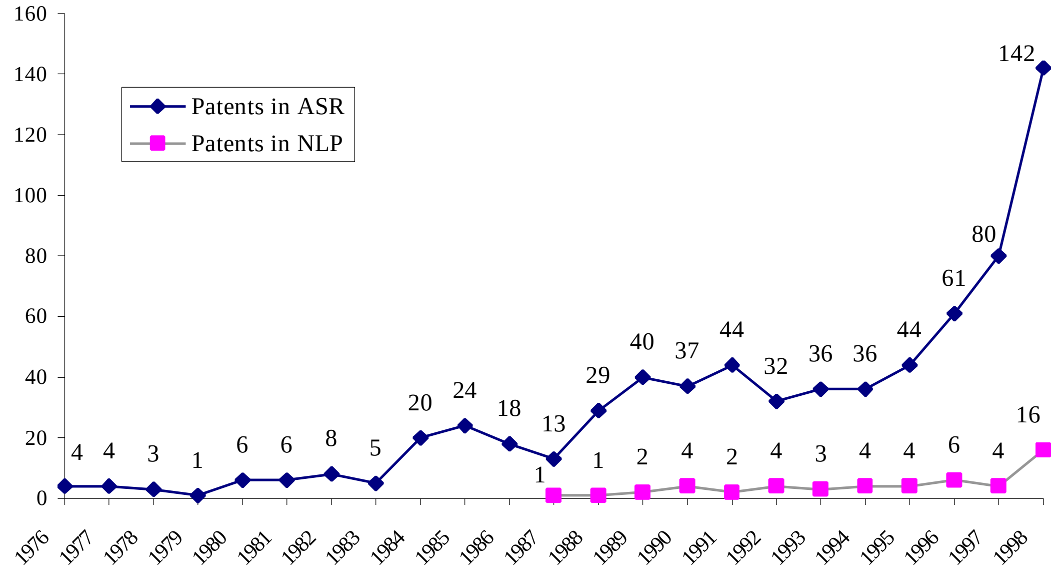


**Fig. 2 Trends in Patenting in ASR and NLP**

1913. L. E. Baum and his colleagues published the basic theory in the late 1960s. J. Baker (Carnegie Mellon University, co-founder of Dragon Systems in 1982) and IBM's F. Jelinek implemented HMMs in the 1970s. Widespread understanding and application of their theory occurred some decades after their initial investigation, since the basic theory was published in mathematical journals, which were not generally read by engineers. And the original applications of the theory of speech processing did not provide sufficient tutorial material for the majority of readers to understand the theory and to be able to apply it to their own research (Rabiner, 1989).

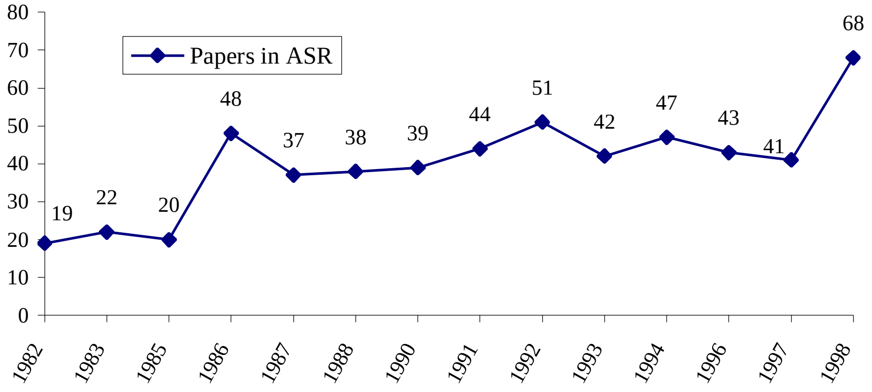


**Fig. 3 ICASSP Papers in ASR per year**

## 4.2 A Multi-technology and Still Experimental Field

According to our survey, a speech technology department in a large company or a small company has on average 27 computer engineers, 6 linguistics scientists and 2 psychologists/cognitive scientists. Only one company had more linguistic scientists than computer engineers, while in another company there were more psychologists than linguists. Although two thirds of companies develop systems in English language, they have started transferring their applications into other languages seeking new markets. Reaching global markets will require applications customised for many languages.

The survey also showed that few companies have hired large numbers of scientists to implement full scale product development. Only 10% of the companies have speech departments with over 100 researchers and 12% have 50-99 researchers. 40% of the responses indicate very small companies or exploratory activity. 65% of the companies managed to deliver their first complete product to market during the past five years, while 18% of them are currently conducting research without having any product yet in mind. Several respondents said that they expect a product to be ready in the near future, while others could not even estimate when their effort will be ready to reach the market. As ASR + NLP systems generally require an extended training session during which the computer system becomes accustomed to real-world data, companies that managed to get early in the market have a competitive advantage of availability of large amounts of data through their existing customers. Very often recording of users' responses during automatic telephony services is a part of the agreement between a technology and a service provider.

### 4.3 Sources of New Entrants

One of the most notable features of US patenting data is the increasing share of small firms. In response to our survey, 14% of companies began their lives as university spin-offs and 20% began as specialised in ASR and NLP firms. Founders of the small companies employed a variety of means and sources to start them. They commonly began with an individual or small group of people convinced of the potential of a new business based on these particular technologies. Some clues about their origin emerge from the regional breakdown of patenting in the two fields in the USA, as shown in Figure 4.

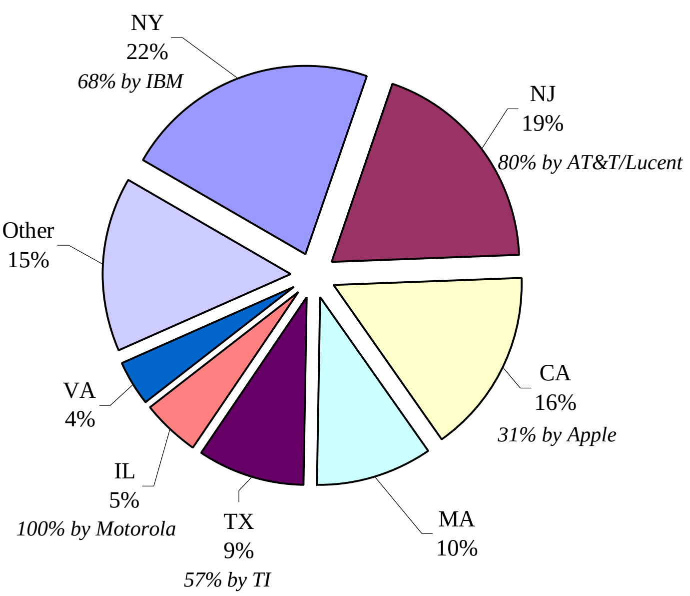


**Fig. 4 Regional patenting patterns within the USA and the respective contribution of large firms**

This shows that New York, New Jersey, California, Massachusetts and Texas are the states with the highest patenting rates in ASR + NLP. In some, high performance is explained by the location of large IT firms:

68% of patents in New York State belong to IBM, 80% in New Jersey to AT&T, 57% in Texas to TI and 100% in Illinois to Motorola. In California and Massachusetts, however, the pattern is different. The cases of Silicon Valley (CA) and Route 128 (MA) have been examined by many analysts, but especially for ASR + NLP, the DARPA funding gave a boost to academic partnership in these two regions. Most of the top ranking small firms in ASR + NLP patenting are established there (CA: Dialogue Systems and Emerson & Stem, SRI/Nuance, MA: Dragon Systems and Kurzweil Applied Intelligence). The emergence of successful specialised firms has depended heavily on the strength of local universities and public research institutes in the particular sciences, and less on local large firms.

### 4.4 The Co-existence of Very Large and Very Small Firms

Table 3 shows that ASR + NLP business environment is a combination of large traditional companies and a host of smaller companies and start-ups (written in Italics). The small firms patenting in ASR + NLP arevery specialised in their technological portfolio; comparisons with market newsletters[13] show that not all of them patent. The active large companies include eight of the ten top firms patenting in the USA between 1990 and 1998[14]. Although taking only a very small share of their total technological resources, these companies seem to have recognised that this field may have a long-term strategic significance. They are therefore trying to build up a competence or undertake exploratory research in order to provide strategic flexibility to respond quickly when opportunities arise.

---

[13]Three monthly newsletters describe the latest developments in speech recognition applications, new products, R&D, marketing and investments in this emerging industry. ASRNews, by Voice Information Associates (VIA), Speech Recognition Update, by TMA Associates, and VoiceNews, by Stoneridge Technical Services.

[14]These are Mitsubishi, IBM, Canon, Hitachi, Toshiba, Motorola, NEC and Matsushita. The two missing firms are General Electric and Eastman Kodak.

**Table 3: Top Recipients of U.S. Patents in ASR and NLP 1976-1998**

| Company Name | ASR 1976-87 | ASR 1988-98 | NLP 1987-98 | Total Patents of Firm in ASR&NLP | | Total Patents of Firm (all sectors) | | % Firm's Share in all ASR & NLP patents[15] | | % Share of ASR & NLP in Total Firm's Patents[16] | | Country / US State |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AT&T Bell Laboratories | 5 | 12 | | 17 | 56 | 4088 | 10242 | 2.16 | 7.10 | 0.42 | 0.54 | NJ |
| AT&T Corp. | | 9 | | 9 | | 1340 | | 1.14 | | 0.67 | | NJ |
| Lucent Technologies Inc. | | 26 | 1 | 27 | | 2068 | | 3.43 | | 1.30 | | NJ |
| Bell Telephone Labs Inc. | 3 | | | 3 | | 2746 | | 0.38 | | 0.11 | | NJ |
| Inter. Business Machines | 1 | 48 | 6 | 55 | | 19054 | | 6.99 | | 0.29 | | NY |
| NEC Corp. | 1 | 19 | 6 | 26 | 36 | 9227 | 9404 | 3.30 | 4.57 | 0.28 | 0.38 | Japan |
| Nippon Electric Company Ltd. | 10 | | | 10 | | 177 | | 1.27 | | 5.65 | | Japan |
| Kabushiki Kaisha Toshiba | | 22 | | 22 | 28 | 14335 | 17251 | 2.79 | 3.55 | 0.15 | 0.16 | Japan |
| Tokyo Shibaura Denki K. K. | 6 | | | 6 | | 2916 | | 0.76 | | 0.21 | | Japan |
| Matsushita Electric Industrial | 8 | 14 | 1 | 23 | 26 | 11682 | 11947 | 2.92 | 3.30 | 0.20 | 0.22 | Japan |
| Panasonic Technologies, Inc. | | 3 | | 3 | | 85 | | 0.38 | | 3.53 | | NJ |
| Hitachi Ltd. | 7 | 9 | 8 | 24 | | 20248 | | 3.05 | | 0.12 | | Japan |
| Motorola | | 19 | | 19 | | 12154 | | 2.41 | | 0.16 | | IL |
| Texas Instruments Inc. | | 18 | 1 | 19 | | 7029 | | 2.41 | | 0.27 | | TX |
| *Dragon Systems, Inc.* | | *18* | | *18* | | *24* | | *2.29* | | *75* | | *MA* |
| Apple Computer, Inc. | | 17 | 1 | 18 | | 1103 | | 2.29 | | 1.63 | | CA |
| *Kurzweil Applied Intelligence*[17] | | *16* | | *16* | | *16* | | *2.38* | | *100* | | *MA* |
| Ricoh Company Ltd. | | 16 | | 16 | | 4482 | | 2.03 | | 0.36 | | Japan |
| Fujitsu Ltd. | | 11 | 3 | 14 | | 8158 | | 1.78 | | 0.17 | | Japan |
| Sharp Kabushiki Kaisha | 3 | 7 | 2 | 12 | | 5837 | | 1.52 | | 0.21 | | Japan |
| *Voice Control Systems Inc.* | | *12* | | *12* | | *15* | | *1.90* | | *80* | | *TX* |
| Canon Kabushiki Kaisha | 2 | 8 | 1 | 11 | | 17362 | | 1.40 | | 0.06 | | Japan |
| Mitsubishi Denki Kab.Kaisha | 2 | 6 | 1 | 9 | 11 | 12450 | 12470 | 1.14 | 1.39 | 0.07 | 0.08 | Japan |
| Mitsubishi Elect. Res. Labs | | 2 | | 2 | | 20 | | 0.25 | | 10 | | MA |
| Nissan Motor Company Ltd. | 11 | | | 11 | | 2817 | | 1.40 | | 0.39 | | Japan |
| Sony Corp. | | 10 | 1 | 11 | | 9637 | | 1.40 | | 0.12 | | Japan |
| British Telecommunications | | 7 | 1 | 8 | | 675 | | 1.01 | | 1.19 | | U.K. |
| Exxon Corporation | 3 | | | 3 | 8 | 3 | 2059 | 0.38 | 1.01 | 100 | 0.39 | NY |
| Exxon Res. and Eng. Co. | 2 | 3 | | 5 | | 2056 | | 0.63 | | 0.24 | | NJ |
| ITT Corp. | 1 | 5 | | 6 | 8 | 871 | 904 | 0.76 | 1.01 | 0.69 | 0.89 | NY |
| ITT Defense Communications | | 2 | | 2 | | 33 | | 0.25 | | 6.06 | | NJ |
| Seiko Epson | | 7 | | 7 | | 5389 | | 0.89 | | 0.13 | | Japan |
| U.S. Philips Corp. | 2 | 5 | | 7 | | 13200 | | 0.89 | | 0.05 | | NY |
| Alcatel | | 6 | | 6 | | 2026 | | 0.76 | | 0.30 | | France |
| ATR Labs | | 5 | | 5 | | 47 | | 0.63 | | 10.64 | | Japan |
| Casio Computer Co. Ltd. | 2 | 3 | | 5 | | 1470 | | 0.63 | | 0.34 | | Japan |
| SRI International | | 5 | | 5 | | 351 | | 0.63 | | 1.42 | | CA |
| *Threshold Technology Inc.* | *5* | | | *5* | | *9* | | *0.63* | | *55.55* | | *NJ* |

The above data are consistent with the results of our survey, where the companies most frequently recognised as competitors were[18]: IBM (14), Kurzweil Applied Intelligence (11), Dragon Systems (10), AT&T/Lucent (8), Philips (8), SpeechWorks (7), Nuance (7), VCS (5), Purespeech (4), Microsoft (3), Sony (3), Apple (2), NEC (2) and over 15 other firms that were referred to only once. Interestingly, Japanese companies were not recognised by survey participants as major competitors in the field – a point to which we return in section 4.7 below. Otherwise, the recently established SpeechWorks and Nuance have exploited

[15]Total number of patents of all firms in ASR and NLP is 788, which is constituted from 145 patents in ASR for the period 1976-87, 592 patents in ASR during the period 1988-97 and 51 patents in NLP during the period 1976-98.
[16]Final figures estimated on a parent company basis.
[17]Kurzweil Applied Intelligence was bought by Lernout & Hauspie in 1994.
[18]In parentheses the frequency in answers.

the research of their parent organisations: MIT and SRI respectively; their achievements are reflected in publications in Tables 4 and 5 below.

**Table 4 Top Publishers of ICASSP Papers in ASR 1982-97**

| Company , University or Research Lab Name | 1982-98 | % Papers Share | Country/State |
|---|---|---|---|
| AT&T Bell Labs | 50 | 8.95 | NJ |
| Carnegie Mellon University | 40 | 7.16 | PA |
| IBM Corp. | 31 | 5.55 | NY |
| ATR Research Labs | 28 | 5.00 | Japan |
| BBN Labs | 17 | 4.64 | MA |
| MIT | 13 | 2.34 | MA |
| NTT | 13 | 2.34 | Japan |
| University of Cambridge | 13 | 2.34 | U.K. |
| Matsushita / Panasonic Technologies | 11 | 1.97 | CA |
| Philips Research Labs | 11 | 1.97 | Germany |
| University of California Berkeley / ICSI | 9 | 1.25 | CA |
| SRI International | 8 | 1.43 | CA |
| NEC | 7 | 1.25 | Japan |
| Royal Signals and Radar | 7 | 1.25 | U.K. |
| CSELT | 6 | 1.07 | Italy |
| University of Edinburgh | 6 | 1.07 | U.K. |
| Brown University | 5 | 0.89 | RI |
| Dragon Systems | 5 | 0.89 | MA |
| Georgia Institute of Technology | 5 | 0.89 | GA |
| Inst. of National Research in Information and Automation | 5 | 0.89 | France |
| Rutgers State University | 5 | 0.89 | NJ |
| University of Karlsruhe | 5 | 0.89 | Germany |
| Other | 261 | 46.72 | - |
| *TOTAL* | *559* | | |

Table 4 shows the top firms and institutions publishing ICASSP papers in ASR during the period 1982-97; and Table 5 lists firms and universities which have published more than five journal papers in ASR + NLP combined in the period 1982-98. The small firms or spin-offs of listed universities are generally the same ones often mentioned by newsletters and business newspapers as successful innovators. Several large firms are also listed in the top patenting positions in Table 2: there is therefore a consistency in patents and publication performance for

**Table 5 Top Journal Papers Publishers in ASR + NLP 1982-98**

| | ASR in JASA / IEEE Transactions on Speech and Signal Processing / Speech Communication | | | | NLP in Computational Linguistics / Communications of the ACM / Journal of Artificial Intelligence Research (JAIR) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Company, University or Research Lab | 1982-98 | Collab. | Total | % | 1982-98 | Collab. | Total | % | Total Papers | Country / US State |
| AT&T Bell Labs / Lucent Technologies | 21 | 8 | 29 | 4.44 | 9 | 5 | 14 | 7.77 | 43 | NJ |
| MIT | 11 | 7 | 18 | 2.75 | 3 | 2 | 5 | 2.77 | 23 | MA |
| IBM | 11 | 5 | 16 | 2.45 | 2 | 4 | 6 | 3.33 | 22 | NY |
| University of California Berkeley (ICSI) | 6 | 7 | 13 | 1.99 | 2 | 6 | 8 | 4.44 | 21 | CA |
| University of Cambridge | 9 | 5 | 14 | 2.14 | 1 | 3 | 4 | 2.22 | 18 | U.K. |
| ATR Research Labs | 8 | 9 | 17 | 2.60 | | | | | 17 | Japan |
| Carnegie Mellon University | 9 | 7 | 16 | 2.45 | | | | | 16 | PA |
| NTT | 7 | 5 | 12 | 1.84 | 3 | 1 | 4 | 2.22 | 16 | Japan |
| Duke University | 6 | 3 | 9 | 1.38 | 3 | 2 | 5 | 2.77 | 14 | NC |
| CNRS, LIMSI | 7 | 3 | 10 | 1.53 | 2 | 1 | 3 | 4.99 | 13 | France |
| University of Minnesota | 7 | 2 | 9 | 1.38 | 2 | 1 | 3 | 1.66 | 12 | MN |
| University of Texas | 6 | 2 | 8 | 1.22 | 1 | 2 | 3 | 1.66 | 11 | TX |
| Johns Hopkins University | 4 | 3 | 7 | 1.08 | 1 | 2 | 3 | 1.66 | 10 | MD |
| Oregon Graduate Inst. Science and Technology | 4 | 6 | 10 | 1.53 | | | | | 10 | OR |
| SRI / Nuance Communications | 4 | 6 | 10 | 1.53 | | | | | 10 | CA |
| Arizona State University | 3 | 6 | 9 | 1.38 | | | | | 9 | AZ |
| Indiana University | 5 | 4 | 9 | 1.38 | | | | | 9 | IN |
| Technion Israel Institute of Technology | | | | | 4 | 5 | 9 | 4.99 | 9 | Israel |
| Tel Aviv University | 2 | | 2 | 0.31 | 4 | 3 | 7 | 3.88 | 9 | Israel |
| University of California Los Angeles | 4 | 5 | 9 | 1.38 | | | | | 9 | CA |
| University of Waterloo | 5 | 4 | 9 | 1.38 | | | | | 9 | Canada |
| CUNY, Center Research Speech and Hearing | 2 | 4 | 6 | 0.92 | | 2 | 2 | 1.11 | 8 | NY |
| Boston University | 3 | 5 | 8 | 1.22 | | | | | 8 | MA |
| INRIA Lorraine | 5 | 2 | 7 | 1.01 | | 1 | 1 | 0.55 | 8 | France |
| Medical University of S. Carolina | 4 | 4 | 8 | 1.22 | | | | | 8 | SC |
| Philips Research Labs | 5 | 3 | 8 | 1.22 | | | | | 8 | Netherlands |
| University of Iowa | 3 | 4 | 7 | 1.01 | | | | | 7 | IA |
| Free University Amsterdam Hospital | 5 | 1 | 6 | 0.92 | | | | | 6 | Netherlands |
| Harvard University | 4 | 2 | 6 | 0.92 | | | | | 6 | MA |
| House Ear Institute | 3 | 4 | 7 | 1.01 | | | | | 6 | CA |
| Lernout & Hauspie (L&H) | 1 | 3 | 4 | 0.61 | | 2 | 2 | 1.11 | 6 | Belgium |
| Korea Advanced Inst. Science and Technology | 4 | 2 | 6 | 0.92 | | | | | 6 | S. Korea |
| McGill University | 2 | 4 | 6 | 0.92 | | | | | 6 | Canada |
| Northeastern University | 3 | 3 | 6 | 0.92 | | | | | 6 | MA |
| University of Melbourne | 4 | 2 | 6 | 0.92 | | | | | 6 | Australia |
| Universitat Politecnica De Catalunya | 2 | 3 | 5 | 0.76 | | | | | 5 | Spain |
| University of Hong Kong | 2 | 3 | 5 | 0.76 | | | | | 5 | China |
| University of London | | 1 | 1 | 0.15 | 2 | 2 | 4 | 2.22 | 5 | U.K. |
| University of Sheffield | 3 | 2 | 5 | 0.76 | | | | | 5 | U.K. |
| University of Utrecht | 4 | 2 | 6 | 0.92 | | | | | 5 | Netherlands |
| University of Washington | 3 | 2 | 5 | 0.76 | | | | | 5 | WA |
| USA, Walter Reed Army Medical Center | 4 | 1 | 5 | 0.76 | | | | | 5 | DC |
| Individual Authors | 1 | 1 | 2 | 0.31 | 9 | | 9 | 4.99 | 11 | |
| OTHERS | 136 | 128 | 264 | 40.39 | 37 | 51 | 88 | 48.84 | 352 | |
| **TOTAL** | 395 | 284 | 653 | 100 | 85 | 95 | 180 | 100 | | |

the large firms. However, small firms such as Dragon Systems and universities such as Carnegie Mellon, MIT, Cambridge etc. are also found in the top rankings in publications. Interestingly, over 40% of the published papers in Table 5 are collaborative. We shall examine one of the reasons why in the next section.

## 4.5 Different but Complementary Strategies

Table 3 shows that ASR + NLP have very different positions in corporate technology strategy, depending on firm size. Applying the classification method proposed in Granstrand et al. (1997), technological competencies of the firms result to the pattern shown in the Figure 5. Along the Y-axis is the percentage share of ASR + NLP in the total patenting of the firm, reflecting the relative importance of the field in the firm's technological portfolio. Along the X-axis is the firm's revealed technology advantage index, which is defined as the firm's share in ASR + NLP patenting, divided by the firm's share of total patenting in all fields, reflecting the firm's degree of specialisation in ASR + NLP compared to the corporate average.

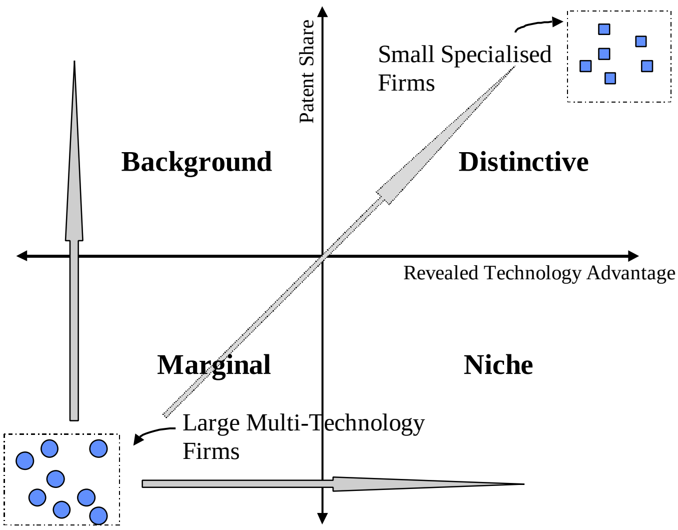


**Fig. 5 Technological profiles of large and small firms in ASR + NLP**

For large multi-technology firms in the *marginal* quadrant, ASR + NLP consumes only a small proportion of corporate technological resources, at least so far. For the future, they will have two options, depending on the rate of progress in these technologies, especially NLP. If it is high, substantial investments (including those in complementary technologies) could open up massive market opportunities. ASR + NLP in this case will move into the *background* quadrant, and firms will need to be able to benefit from technical opportunities in the field, by combining them with their own established competencies. If progress in the technologies is low, modest investments will allow them to exploit niche markets. It is unrealistic for large firms to move from marginal to destinctive quadrants as that prerequisites high share of corporate technological resourses and a strong revealed technology avantage compared to the competition.

Small, new entrant companies such as Dragon Systems, L&H (with Kurzweil Applied Intelligence), Threshold Technology, VCS Systems, Dialogue Systems and Emerson & Stem are classified in the *distinctive* quadrant. These are companies who are investing a high percentage (from 56% up to 100%) of their research portfolio in ASR + NLP. Their rates of growth will depend mainly on the future rate of progress of their distinctive technologies.

The two clusters of firms can also be recognised in Figure 6 where they have been classified according to their share of U.S. patenting in ASR + NLP and the importance of these technologies for the firm. Interestingly, all firms are within a very narrow band (7% or less of total patenting), and none of them dominates the field. Figure 6 also shows the split between the large, multi-technology firms, for whom speech represents 0.5% or less of their patents, and small specialised firms for whom speech represents more than 50%.

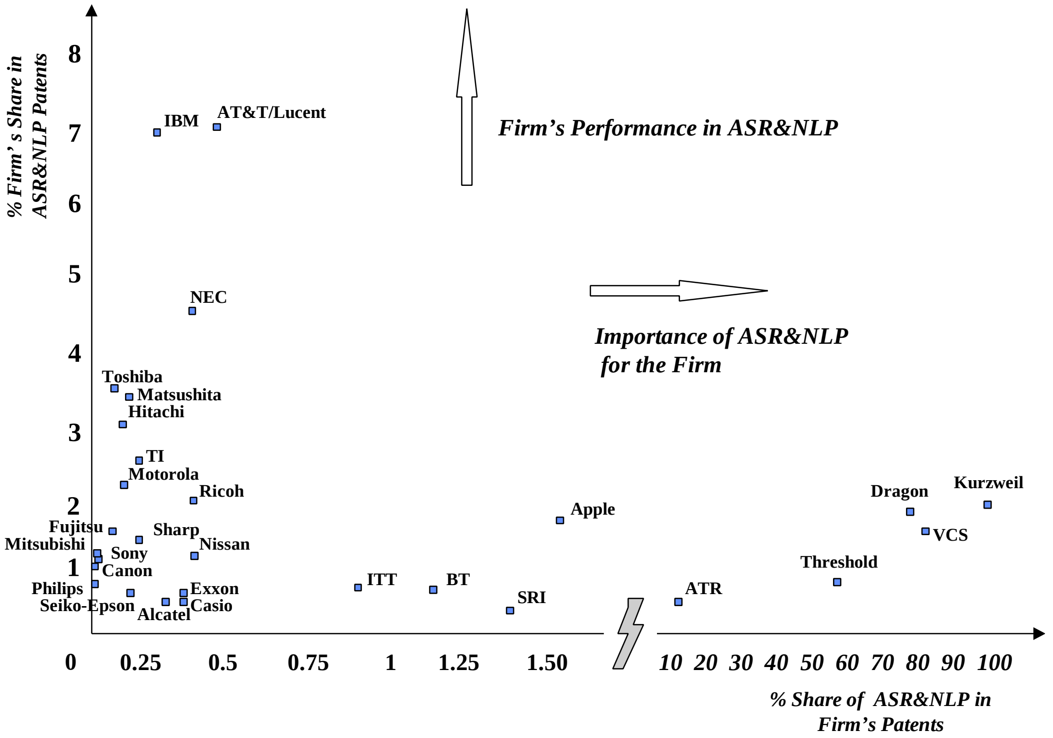


**Fig. 6 The coexistence of very large and small firms in ASR + NLP**

Within the framework of Mitchell & Hamilton (1988), we conclude that many large corporations are currently committing relatively modest R&D expenditures in ASR + NLP to seek opportunities for profitable investments in the future. Many companies have established programmes for the *knowledge building* stage, which requires relatively inexpensive research for developing and maintaining expertise in fields of potential future advantage. Fewer companies have moved to the *strategic positioning* stage, which involves applied R&D and feasibility demonstration, and is concerned with reducing technical uncertainties, identifying potential markets and building in-house competencies. Finally, even much fewer companies have moved to the stage of *business investment*, where they are able to develop and distribute new and better products and have positive revenues from their R&D expenditures.

Although the strategies of these large and small firms are clearly very different, they are interdependent and complementary, reflecting the considerable technical and commercial uncertainties surrounding fields of rapid and potentially major changes. Commercial opportunities emerging from major scientific and technological advances are not always clear. Product features valued by users are not always obvious at the outset; and once identified they can often be easily imitated by competitors (Tidd et al., 1997). The need to acquire external knowledge and competencies from other firms and universities increases with the number of component technologies. In our survey, we examined whether companies develop speech applications independently or in collaboration with others. 30% of the responses indicated basically independently developments, while 56% had external partnerships with OEMs (Original Equipment Manufacturers) licensees, VARs (value added resellers), ISVs (Independent Software Vendors), channel and strategic partners. Universities and related research laboratories seem to be preferred partners in this exploratory phase of development, rather than other companies. Most companies in our survey rely mainly on internally generated knowledge, claiming that their in-house activities are the major source of technical know-how. Nevertheless, they all enjoy a variety of often strong external technological links with universities, research institutes and other industrial companies. The majority of participants also referred to joint projects, which shows that the complexity of problems faced in these technologies need inter-firm co-operation, and that only a few companies have the ability to cope with complicated projects on their own.

Decisions on acquisition and collaboration are an important component in firms' strategies, in dealing with technological opportunities and threats. Based on data on mergers, acquisitions and partnerships published by Lewis (1998), we can identify three patterns. *First*, firms with a record in ASR + NLP (e.g. IBM) try to extend their research portfolio in more specific areas with selected partnerships. *Second*, new entrants (e.g. Microsoft) try to build a background through massive investments in several specialised companies. Interestingly, large firms also invest simultaneously in different small firms, which compete against each other and have similar types of competencies and resources. For example, the seed money Intel invested in Nuance and SpeechWorks gives the opportunity to influence these specialised companies to develop applications compatible with Intel's own platforms. *Third*, small companies often agree to share their resources and know-how in order to survive the competition. This pattern is fully consistent with Freeman's (1991) conclusion that the rapid development and diffusion of new technologies, especially in IT, is the main reason for growth of strategic alliances.

## 4.6 Classification of Corporate Profiles

Our data show that many companies are active in ASR and NLP. Large firms are making long term investments to build technological capabilities by capitalising economies of scope, enabling them to explore and experiment with these technologies for possible deployment. Some of the firms have been early starters, but have failed to sustain their competencies. Few late comers have caught up, through various organisational routines, such as alliances, mergers or buy-outs. Technological competencies and advantage have been obtained either through internal R&D, joint venturing, subcontracting, exchange of researchers, and most commonly by employing a combination of these. In none of the large firms has ASR become a core technology, while the figures for NLP are much lower. A more refined classification follows, of corporate profiles of active firms in ASR + NLP according to their major business.

*Telecommunications Driven Firms*. For many years, telecommunications service providers traditionally focused on voice traffic by primarily providing a voice connection from one location to another. This perception has changed in recent years, as markets have become increasingly liberalised, deregulated and competitive. In this new environment service providers have begun to offer sophisticated products and services to differentiate themselves. With all its current and potential applications, speech technology is rapidly becoming an essential component of how organisations operate. Management teams are recognising that the integration of speech technology is a key factor to increased productivity, enhanced customer service and cost savings to businesses of all types. As a result, telecommunications systems are moving from their former voice-only emphasis to an emphasis on voice and data. For example, AT&T, Nortel, Bell Atlantic - GTE, and BT have developed their own speech processing technology, refined it in their own labs and used it in their products/services. This shows that large firms have identified the importance of spoken language interfaces for their traditional markets. R&D is normally conducted through a combination of specialised laboratories as well as corporate divisions that operate flexibly and independently. Additional R&D efforts are conducted through partnerships with universities and other companies. The integration of ASR + NLP with the Voice over Internet Protocol (VoIP) - an emerging industry trend in computer telephony - is expected to offer substantial revenue opportunities for telecommunication and Internet Service Providers (ISPs).

*Desktop Computing Firms*. In this category the major business is computer operating systems, desktop applications and peripherals. ASR + NLP are complementary technologies, knowledge of which is essential in product design, especially for additional product features. The major companies - which include IBM, Dragon Systems Lernout & Hauspie (L&H) and Philips - are locked in a battle to dominate the PC space. Penetration of speech as an interface will hinge upon how quickly applications and operating systems embrace it. Moreover, it will be critical to develop new user interfaces by re-inventing applications and operating system controls around speech, rather than simply voice-enabling pull-down menus. Unlike the above companies, Microsoft, which dominates the PC operating systems market, has yet to turn its speech recognition software into a product. With its investment in L&H, it tried to rush a first generation programme to market, in order to have something to put up against more advanced competitors. However, Microsoft's strong links with Carnegie Mellon University, which is a leader in academic ASR research, may offer the necessary know-how to be able to soon develop competitive products which could be easily promoted by taking advantage of the large installed base of company operating systems. Apple itself may enter the speech recognition market, as shown from the patenting data. While it already has PlainTalk, which allows voice navigation through menus, the company has promised dynamic new features when it releases its new operating system.

*Firms specialised in Speech Technology*. Firms in this category are entirely focused on ASR + NLP and constitute core technologies, dominating corporate thinking. This often enables them to build better tools, applications and services than large multi-technology companies that treat this market as peripheral to their main business. Normally, they have a flexible enough organisation to respond adequately to the numerous quickly changing markets. The number of these firms is quite high, although many of them do not appear in the patenting or publications data, because they have been recently established or have a secrecy policy. Laboratory culture is strong in these companies, even amongst senior managers, and the R&D directors are mainly from speech related fields. Two important characteristics of this group of firms are common to small firms in other fast-moving technologies - the exceptional skills and experience of the researchers, and the combination of technical and non-technical personnel in the management teams. If the large companies are overly cautious, some of these energetic firms may capture a large share of the market and – if speech technology fulfils its promise - become superstars.

Very often these companies are university spin-offs such as SpeechWorks (MIT), Nuance (SRI), or Entropic (Cambridge U.K. and Washington DC). Spin-offs tend to cluster around their respective incubator organisations, public or private, forming regional networks of expertise. In our case, patenting data show that most new companies have not been established close to the giants such as AT&T or IBM in New Jersey or New York, but in California and Massachusetts, where the DARPA funding recipients are located, and where the proportion of patenting by individuals is also very high. This indicates that academic researchers decide to start high tech companies more than those already working in industry, taking advantage of their established links and of the availability of capital. In the USA, the role of defence industry investment, liberal tax regimes and venture capital explains this trend. Our data are consistent with the pattern that small firms are even more reliant than large ones on the national or regional systems of innovation in which they are embedded.

*Consumer Electronics Firms*. Kodama (1992) points out that high-tech companies now have to face ‘invisible competitors’, not knowing which sector these competitors come from. Although it is difficult to say how seriously any of the consumer electronics companies (e.g. NEC, Toshiba, Matsushita, Canon and Ricoh) have become involved with speech processing technologies, their presence should not be underestimated. Although it is not their main business, they develop their own speech technology and incorporate it in their future products or services. Their activities will allow natural language in automobiles and other eyes-busy environments. Ricoh, for example, has recently announced a voice-activated camera, while more cars are expected to have Global Positioning Systems (GPS) that will help drivers reach their destinations by conversing with the system about possible routes. Also, in November 1998 Canon launched the first fax machine with voice dial feature containing speech recognition technology.

## 4.7 National Differences in Language

Our survey shows that language interface systems are under development in North America, Europe and Asia. These are focused on a small set of common languages including English, Chinese, Spanish, Japanese, German and French as well as on their respective regional dialects. English[19] may be the dominant language for developing applications, but two thirds of companies have started transferring their applications into other languages seeking for new markets, where competition will be weaker. Crucially, reaching global markets will require companies to customise applications for many languages. The linguistic fragmentation of the world software market represents an opportunity for country-specific solutions, since a large part of development and marketing costs depend on the number of language versions of a product, while the size of language market first addressed constitutes an important competitive advantage.

The majority of the (mainly US) participants (74%) recognise their national environment as particularly favourable. Some others (mainly from Japan) said that it is favourable for the availability of researchers, while facing problems dealing with the language structure and representation. Some participants from France, the U.K. and Italy accepted their national environment as second best to the US. Finally, a number of participants from small language groups did not consider the national environment as favourable for corporate activities, but expressed interest in this field, which soffers opportunities arising for small local companies.

[19]English language in ASR and NLP covers mainly US English, but also British, Canadian, Australian and Irish English.

The patenting data in ASR + NLP clearly confirmed the US dominance of the field, having been granted 46% of all patents. Japanese firms have the 32.5%, while the third country (U.K.) has only 2.2%. Germany has 1.11%, Canada 0.73%, Netherlands 0.48%, France and Sweden 0.42%, and all the remaining countries count 16.56%. Although the performance of Japanese giants (Hitachi, Toshiba, NEC, etc.) in patenting is very strong, market penetration is not recognised either by the survey participants or in the industry newsletters. This is because complicated writing schemes in Japan and China make keyboard use problematic. The Chinese language uses between 4,000 and 6,000 characters, while Japanese uses 2,000 but also relies on three other alphabets in parallel. Both languages are also riddled with homonyms (words that sound the same but have different meanings). Moreover, words are strung together without gaps, making parsing even more difficult. Moffet (1998, p. 56) reports as follows about that:

*"What the West takes for granted - the ability to put language into digital form - is a vexing problem for cultures whose writing is based on thousands of ideograms, the symbols used in Chinese, Japanese and other Asian languages."*

From the publications data we find in Japan a growing focus on cognitive science research and an appreciation of its importance for human-computer interaction, but it is not as strong as in the USA. There have been few commercial ASR telecommunications applications in Japan since the ANSER system was introduced for public use (Nitta, 1994; Sugamura et al., 1994)[20]. However, the long-term research of several non-computing companies (SONY, Nippondenso, etc.) has led to products such as car navigation systems, controlled by voice. Several microprocessor companies develop hardware based ASR + NLP. For instance Sharp and OKI have recently announced voice control software with a microcontroller chip and multilingual text-to-speech (TTS) chip respectively.

[20]Some PC manufacturers (NEC, IBM-Japan, Toshiba, etc.) have developed their own ASR software to run on their own platforms. Non PC manufacturers (NTT, NTT-DATA, etc.) have also developed their own ASR to run on PCs. Significant work has also taken place in recognition of broadcast news speech. Interested readers in Japanese ASR products can refer to a comprehensive description made by Kitai et al. (1997). One of the most important programmes covering most state-of-the-art issues in ASR and NLP began in 1993 at the ATR Interpreting Telecommunications Research Laboratories using massively parallel machines for speech processing and speech/language databases.

Non-Asian firms also target the market of quality speech recognition products in the Asian Pacific region. Motorola, Lexicus division, for instance, recently announced an agreement to license its Chinese speech technology to L&H, while most of the American and European companies in the survey claimed to have developed applications for Japanese and Chinese. For this reason, alliances of local with American companies may proved to be very important for market success.

American firms are also active in European market[21], which is characterised by a language diversity[22]. However, several continental companies are involved in speech technologies. European public telephone operators have recognised that a single company in each country cannot achieve the required uniformity of speech understanding services. EURESCOM[23] has therefore been established with the objective to explore how general multilingual services could be provided in the participating countries: France, U.K., Italy, Germany, Spain and Portugal (Johnston et al., 1997). In addition, Europe's leadership in digital telephony through large R&D expenditures and continuous innovation in component and interface technologies (Davies, 1997) provides a platform for countless applications towards the computer-telephony integration. With the exception of Philips, most non-telecommunications European companies involved in ASR + NLP are small ones.

## *5. Conclusions and Unanswered Questions*

Our analysis confirms ASR + NLP as emerging "new-science" – based technologies, the progress of which has been heavily conditioned by the rapid advances in IT, and the early exploitation of which involves both very large, multi-technology firms exploring future options, and very small firms exploiting opportunities emerging from publicly funded research in universities. It also confirms the usefulness of the framework of

---

[21] For instance, the IBM's European Speech Research Teams are situated in Winchester (U.K.), Paris (France), Heidelberg (Germany), Seville (Spain), Rome (Italy) and Cairo (Egypt), working on six languages: U.K.-English, French, German, Spanish, Italian and Arabic.

[22] In the EU alone 13 different official languages are spoken.

[23] EURESCOM is a private company composed of 23 European telecommunication operators.

Mitchell and Hamilton (1988) for analysing the technological positioning of large firms in exploring new technologies. But it inevitably leaves some questions unanswered, and raises new ones.

1. Both the U.S. patenting and the conference and publications data may have an English language bias, the extent of which for the moment can only be guessed at.

2. Our data in patenting and scientific publications shows very weak integration of ASR and NLP technologies in the same organisation[24]. Simply providing higher volume of resources (e.g. computing power in ASR) cannot solve the cognition process and other fundamental problems. Consequently, a very interesting question is raised: can system providers who are not experts in language processing and semantics, integrate this technology or will they lose their market share to those who are?

3. The emergence of small-specialised firms has depended heavily on the strength of universities and public research institutes in the particular sciences, and rather than on large firms found in the same region. The role of initial funding from the US DARPA programme has been very influential for the development and diffusion of these sciences. The overwhelming lead of the US firms - especially in the field of ASR - underpins an even more overwhelming lead in related university-based science. A similar pattern can be observed in biotechnology, where publicly funded National Institutes of Health in the USA has been instrumental in advancing academic research in molecular biology, and where such research has become the basis of many "new-science" – based firms. Does this mean that national (and European) policies to emulate the US performance in science-based entrepreneurship are missing the point by giving high priority to providing risk capital and science parks, and provoking so-called 'cultural' changes. In the light of the above evidence, should not more emphasis be given to improving the quality of European science, especially in new and emerging fields? If so, are Eurpean institutions and procedures capable under existing contraints of being as effective in the future as the US DARPA has been in the past??

[24] As exceptions in patenting one could refer to Hitachi, IBM, and AT&T/Lucent who have many publications in NLP as well.

4. Finally, we can speculate that our analysis of the "new-science" – based technology of ASR + NLP, reflects a more general trend, namely the more direct involvement in the initial *development* of fundamental and pervasive technologies of university-based research and related small-firm spin-offs. For earlier pervasive technologies – production machinery, process technologies, measurement and control instrumentation, computer aided design – the locus of development was the large business firm and the small firms of which they were the origin (Mowery & Rosenberg, 1989). Now, the origin of such spin-off firms is increasingly the universities. The glib explanation is that, as we move from "material production" to "knowledge production", the university naturally replaces the factory as the locus of change. But this misses the point. University-based knowledge has always been an important input to radical innovations, but mainly in the form of *research* and background understanding, and not of *development* and specific applicable techniques. The distinguishing difference today could be that recent advances in both the science of molecular biology and the technology of information processing have reduced the *costs* of search and experimentation for specific technical problems and their solutions, and have thereby made it easier for universities to contribute effectively in certain fields to development as well as to research activities. Our research does not give decisive proof of such a trend, but it is not inconsistent with it.

**Acknowledgements**

This paper is based on Koumpis K. *Corporate Technological Positioning in Automatic Speech Recognition and Natural Language Processing*, M.Sc. Dissertation, SPRU, University of Sussex, 1998. A heavily reduced version appeared in the *Speech Technology Magazine*, Feb/Mar 1999. We have appreciated the valuable discussions with Steve Renals, which have significantly improved the quality of the paper, as have the constructive comments of two anonymous referees. We also thank Miguel Carreira-Perpinan, Nick Fikouras and Valia Sdralia for their helpful comments.

## References


Bourland H. and Morgan N. (1994) *Connectionist Speech Recognition – A Hybrid Approach,* Kluwer Academic Publishers.

Business Week (1993) *The Global Patent Race picks up Speed,* August 9, pp. 49-54.

DARPA (1997) *Defence Advanced Research Projects Agency Technology Transition,* available online from http://www.arpa.mil

Davies A. (1997) Competitive Complex Product Systems: The Case of Mobile Communications, *Institute for Prospective Technological Studies (IPTS) Report*, No 19, Seville, Spain.

Freeman C. (1984) Prometheus Unbound, *Futures,* Vol. 16, No 5, pp. 495-507.

Freeman, C (ed) (1987) *Output Measurement in Science and Technology*, Amsterdam: North-Holland.

Freeman C. (1991) Networks of Innovators: a Synthesis of Research Issues, *Research Policy*, Vol. 20, pp. 499-514.

Granstrand O., Patel P. and Pavitt K. (1997) Multi-technology corporations: Why they have "distributed" rather than "distinctive core" competencies, *California Management Review*, Vol.39, No.4, pp. 8-25.

Hicks D. (1995) Published Papers, Tacit Competencies and Corporate Management of the Public Private Character of Knowledge, *Industrial and Corporate Change*, Vol. 4, No. 2, pp. 401-424.

Howard J. M. (1987) *Applied Science and Technology Index*, New York, The H.W. Wilson Company.

Johnston D., Sorin, C., Gagnoulet C., Charpentier F., Canavesio F., Lochschmidt B., Alvarez J., Cortazar I., Tapias D., Crespo C., Azevedo J. and Chaves R. (1997) Current Experiment Applications of Speech Technologies for Telecom Services in Europe, *Speech Communication* Vol. 23, pp. 5-16.

Kamm C., Walker M., and Rabiner L. (1997) *The Role of Speech Processing in Human-Computer Intelligent Communication*, NSF Workshop on Human-Centered Systems Information, Interactivity and Intelligence, February 17-19, Arlington, VA, USA.

Kitai M., Hakoda K., Sagayama S., Yamada T., Tsukada H. Takahashi S., Noda Y., Takahashi J., Yoshida Y., Arai K., Imoto T, and Hirokawa T. (1997) ASR and TTS Telecommunications Applications in Japan, *Speech Communication* Vol. 23, pp. 17-30.

Kodama F. (1992) Technology Fusion and the R&D, *Harvard Business Review*, Jul.-Aug., pp. 70-88.

Kurzweil R. (1997) When will HAL understand What we are saying? Computer Speech Recognition and Understanding in D. G. Stork (Editor), *Hal's Legacy: 2001's Computer as Dream and Reality*, MIT Press.

Lee K-F. (1989) *Automatic Speech Recognition: The Development of the SPHINX System*, Boston, Kluwer.

Lewis B. (1998) Mergers, Alliances and Acquisitions: Mergers Shape Speech Industry, *Speech Technology Magazine,* Jun-Jul Issue.

Mahdi S. and Pavitt K. (1997) Key National Factors in the Emergence of Computational Chemistry Firms, *International Journal of Innovation Management, Vol. 1, No. 4 pp. 355-386.*

Mitchell G.R. and Hamilton W.F. (1988) Managing R&D as a Strategic Option, *Research Technology Management*, Vol. 31, No. 3, May-Jun. pp. 15-22.

Moffet S. (1998) Kissing Kanjui Goodbye? *Business Week,* Special Report, Feb. 23. p. 56.

Mowery D. and Rosenberg, N. (1989) *Technology and the Pursuit of Economic Growth,* Cambridge University Press.

Narin F. and Davidson (1988) *The Growth of Japanese Science and Technology: Science and Technology Indicators show a Dramatic Strengthening*, CHI Research, Washington, George Washington University.

Narin F., Noma E. and Perry R. (1987) Patents as Indicators of Corporate Technological Strength, *Research Policy,* Vol. 16, pp 143-155.

Nelson R. (1959) The Shape Economics of Basic Scientific Research, *Journal of Political Economy,* pp. 297-306.

Nitta T. (1994) Speech Recognition Applications in Japan, *Proc. ICSLP '94,* Vol. 2, pp. 671-674.

Nonaka I. and Takeuchi H. (1995), *The Knowledge-creating Company*, Oxford University Press.

OECD (1993) *The Measurement of Scientific and Technological Activities Using Patent Data as Science and Technology Indicators, Patent Manual,* Paris.

Patel, P and K. Pavitt (1996) "Patterns of Technological Activity: their Measurement and Performance", in P. Stoneman (ed), *Handbook of the Economics of Innovation,* Blackwell.

Price P. (1996) *A Decade of Speech Recognition; The Past as Springboard to the Future,* Proceedings ARPA 1996 Speech Recognition Workshop, Morgan Kaufmann Publishers.

Rabiner L. (1989) A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition, *Proceedings of the IEEE,* Vol. 7, No. 2, pp. 257-286.

Reddy D. R. (ed) (1974) *IEEE Symposium on Speech Recognition*, Carnegie-Mellon University, New York Academic Press.

Rosenberg N. (1990) Why do Firms do Basic Research with their own money, *Research Policy*, Vol.19, No.2, pp. 165-174.

Rudnicky A.I., Hauptmann A.G. and Lee K-F. (1994) Survey of Current Speech Technology, *Communications of the ACM,* Vol. 37, No. 3, pp. 52-57.

Stewart T.A. (1991) Where we Stand, in The New American Century? *Fortune*, Special Issue, p. 17.

Sugamura N., Hirokawa T., Sagayama S., and Furui S., (1994) Speech Processing Technologies and Telecommunications Applications at NTT, *Proc. IVTTA 94,* pp. 37-42.

Tidd J., Bessant, J. R. and Pavitt K. (1997) *Managing Innovation: Integrating Technological, Market and Organizational Change,* John Wiley & Sons.

van Raan, A (1988) (ed.), *Handbook of Quantitative Studies of Science and Technology*, Elsevier.