What do we know about innovation? A Conference in honour of Keith Pavitt

Patterns of Innovation in Spoken Language Engineering Industry

Konstantinos Koumpis

Vodera Research
Reading, UK
koumpis@vodera.com

Abstract

Spoken language engineering has the potential to become an enabling technology, powering advantages in many other areas. This paper examines this emerging industry from both a technology and market perspective. Spoken language engineering is a complex subject and companies need to possess competencies in several fields in order to be able to capitalise on its growth potential. The present study identifies six major technical fields and analyses US patenting activity in them over the last 25 years. The recent increase in number of patents reflects the numerous follow-through inventions of the rapid diffusion phase of innovation rather than the crucial original inventions and innovations. The paper also addresses the questions of how patterns of innovation change over time in this industry, their consequences for future developments and how the recent economic downturn has shaped them.

1. Introduction

Speech and natural language processing, together known as spoken language (SL) engineering, have grown from a multidisciplinary but highly specialised research topic to a core technology for generating, accessing and managing information content. After over five decades of basic research, much of it conducted in university and government laboratories, solutions based on SL engineering have progressed to the point where they are now viable for a broad range of applications, enabling humans to use computers in their own language. SL can be used either as input or as a knowledge source. Speech input addresses applications like dictation, navigation or transactional systems. Speech as a knowledge source covers applications such as meeting capture or spoken audio indexing, classification and retrieval.

SL engineering is important in terms of innovation, competitiveness and growth for the organisations that develop and commercialise related technologies but also for those companies and individuals who adopt them. Current growth in the SL engineering industry stems from consumer demands for simple, seamless interfaces on communication devices and from enterprise requirements for more efficient customer service applications. The advent of high-speed processors, embedded speech recognition systems, low-cost storage and the Internet have significantly boosted SL engineering capabilities and potential in the last decade. Koumpis and Pavitt (1999) classified SL[1] engineering as another 'new science'-based technology that has emerged through the combination of other sciences and figured corporate goals and technological positioning in that industry by comparing patenting and publishing performances. Their results confirmed that SL was at an early exploratory stage of new product development, and showed that most of the technological activities were performed by a complementary combination of small specialised and large multi-technology companies. The small specialized companies depended heavily on nearby universities and public research institutes, and to some extent on nearby large companies; their relations with the large companies were complementary as well as competitive.

[1] That study was limited to automatic speech recognition and natural language processing fields.

The goal of this paper is to examine the SL industry from both a technology and market perspective. Emphasis is given on the patterns of innovation, which mainly refer to how certain regularities in the diversity, types and locus of innovation change according to technical progress and market conditions. The rest of the paper is organised as follows. In the first part the definitions of the major technical fields are briefly presented. Corporate dynamics and nationality of the companies in SL engineering are then studied through analysis of US patenting activity from 1976 to 2002. The technological trajectories are reviewed and the market challenges are discussed by identifying the variables of the sectoral system that affect corporate performance. The paper concludes with strategy implications in a sectoral system perspective.

2. Major technical fields of spoken language engineering

As a consequence of its reliance on disciplines such as signal processing, acoustics, statistics, machine learning, computer science, phonology, linguistics and psychology, the technologies involved in SL engineering are diverse in nature and purpose. While brevity does not permit a full description of them, the analysis in this paper is based on the following taxonomy:

Voice coding: compresses speech signals for either storage or transmission without excessively distorting them. Coding is mainly used over wired and wireless networks.

Speech synthesis: converts arbitrary text into audible speech, with the goal of being able to provide textual information to people via voice messages.

Speech recognition: converts acoustic speech signals into a textual form.

Language understanding: extracts meaning from a string of recognised words and executes an action based on the abstract meaning. Spoken language understanding is essential in dialogue applications.

Language generation: constructs sentences whose structure is based on the relationships among the task attributes that need to be conveyed.

Speaker recognition: can be divided into speaker identification (determines which registered speaker provides a given utterance from amongst a set of known speakers) and speaker verification (accepts or rejects the identity claim of a speaker). Speaker recognition is used for secure access to information and services.

The above technical fields are complex and it takes time for companies to develop expertise in them. As SL systems become more sophisticated, requiring components from at least two distinct fields, innovations are increasingly dependent upon the interactions between these fields. Thus, either greater expenditures on R&D or increasing links with outside sources of knowledge will be necessary in order to exploit the synergies among the distinct fields.

3. Patent activity in the major fields

Patents are an important component of industrial intellectual property and are frequently used as measures of corporate technological activities as they co-evolve with technical change (Narin et al., 1987; Business Week, 1993). The requirement of novelty and the high costs of patent applications make them an indicator particularly appropriate for large companies. 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[2] and since the mid-1980s have begun to reflect advances in software technology[3]. In spite of their

---

[2] US Patent and Trademark Office, patents.uspto.gov

[3] Of the overall growth in patents granted by the US Patent and Trademark Office over 1992-99, ICT accounted for 31% and rose by almost 20% annually.

inevitable imperfections, e.g. the difficulty of assessing the quality of patents as distinct from their number, and the different patenting policies of companies in the same industry, patents serve as a proxy for innovation and can provide unique insights into the activities related to new technologies.

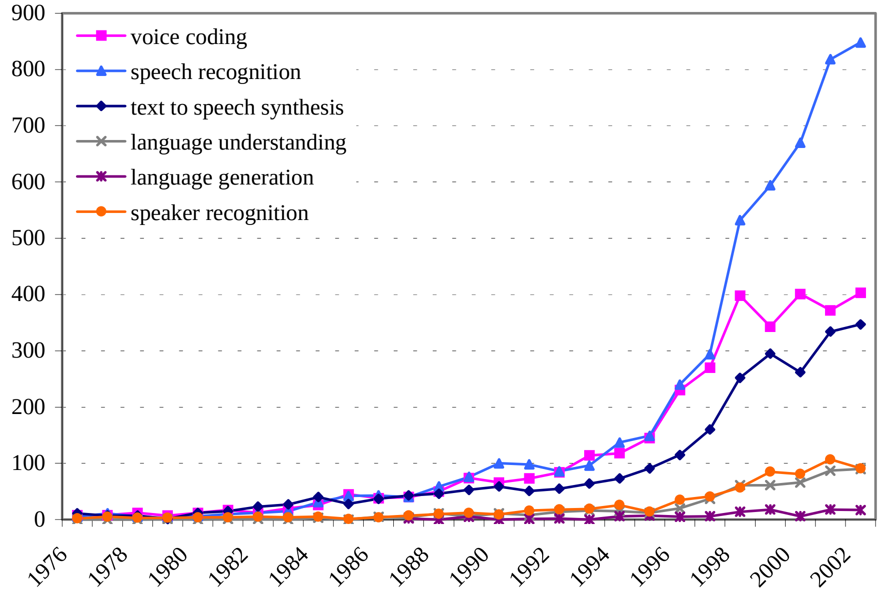

Figure 1 Annual evolution of US patent activity in the major SL technical fields since 1976.

Figure 1 depicts the number of US patents issued annually since 1976 for inventions related to each of the major SL technical fields described above. The differences among the patenting patterns indicate that these fields have attracted unequal amounts of research effort, while it is also true that the fundamental difficulties across fields vary significantly. Until the mid-1980s the number of patents issued, i.e. successful patent applications, for all six technologies was low (in line with the definition of a 'new science'-based technology).

During the following decade there was a consistent increase in the number of patents issued for inventions in speech recognition, voice coding, and speech synthesis. The number of patents for speaker recognition followed a modest growth while the number of patents on the two remaining fields (language generation and understanding) was relatively low. The patent activity in most fields increased rapidly since mid-1990s though. A three-fold growth over a period of five years characterises the annual patent activity in the speech recognition field. This can be partly explained by the fact that speech recognition became feasible for an increasing number of mobile applications, including cellular phones, in-vehicle navigation/information systems, and personal handheld computers, in which speech is easier and safer to use than keyboards or stylus-based interfaces. Patents for voice coding follow a similar pattern to those in speech recognition until 1998, with signs of saturation since. The latter could be due to the availability of ever larger and cheaper amounts of storage and channel capacity that diminishes the need for extra compression. Patent activity in speech synthesis is characterised by a consistent increase and given the trend in voice coding, the number of patents issued for speech synthesis inventions is expected to exceed those related to voice coding during the next few years.

Patent activity in speaker recognition and language generation fields has increased modestly since the mid-1990s, with the number of patents issued for speaker recognition related inventions being slightly higher than those granted for language generation. The need for greater security, particularly related to fraud and terrorist activities, is expected to add to the demand for new developments in voice biometrics and subsequently in solutions based on speaker recognition within the next years. Finally, the patent activity in language understanding is far lower than all the above technical fields. Understanding language means knowing what concept words stand for and knowing how to link those concepts together in a meaningful way. The several types of ambiguity present in natural language, e.g. word-sense and structural ambiguity, make the task of creating software to attach meaningful links between words (often containing transcription errors) very hard.

4. Top patent recipients

Table 1 shows the top 20 patent recipients in the six major SL technical fields covering the period 1976-2002. The majority of companies in this table are also listed in the Top 100 R&D spenders (IEEE Spectrum, 2003), suggesting that SL engineering has been identified as a key area among several large multi-technology companies[4]. The table reveals that two multi-technology companies (AT&T/Lucent Technologies and IBM) have built significant patent portfolios in all six major technical fields. In particular, AT&T/Lucent Technologies hold the first place in speech synthesis and language generation, the second place in speech and speaker recognition and the third place in voice coding and language understanding. IBM has the first place in speech recognition and language understanding fields, the third place in speech synthesis and speaker recognition and the fourth place in voice coding and language generation.

Motorola, Texas Instruments, Sony and Microsoft follow in terms of diverse patent portfolios in SL technologies. Motorola holds the first place in voice coding, the fourth in speech recognition and the fifth in speaker recognition. Texas Instruments, whose core competence is in digital signal processing and analog technologies, holds the first place in speaker recognition, the second in speech synthesis, the third in speech recognition, the eighth in voice coding and the nineteenth in language understanding. Sony shares the first place in language generation with AT&T/Lucent Technologies and further holds the sixth place in voice coding, the eighth in language understanding, the tenth in speech synthesis and the fourteenth in speech recognition. Microsoft holds the second place in language understanding, the eleventh in speech synthesis and the twelfth in speech recognition. Medium-sized but diverse patent portfolios with respect to the above technologies are also held by companies such as Toshiba, Philips, Canon, Matsushita, NEC.

Table 1 does not cover the small technology-based companies, which usually do lots of ‘informal’ and non-measured R&D. Pavitt (2002) suggested that directly useful knowledge and inventions related to SL engineering are also emerging from universities and spin-off companies. This opens a wider range of technological opportunities for established companies, but also increases the technical and commercial uncertainties surrounding each potential development. The patent data revealed that the university patent activity is on the increase. In speech recognition field alone, 45 patents have been assigned to some 25 universities.

[4] The firms shown in bold (Table 1) are those *not* listed in the Top 100 R&D spenders.

rank	Voice coding		Speech recognition		Speaker recognition		Speech synthesis		Language understanding		Language generation	
1	Motorola (IL)	244	IBM (NY)	303	Texas Instr. (TX)	56	AT&T/Lucent (NJ)	166	IBM (NY)	42	AT&T/Lucent (NJ)	11
2	Ericsson (Sweden)	239	AT&T/Lucent (NJ)	275	AT&T/Lucent (NJ)	53	Texas Instr. (TX)	136	Microsoft (WA)	34	Sony (Japan)	11
3	AT&T/Lucent (NJ)	225	Texas Instr. (TX)	138	IBM (NY)	52	IBM (NY)	121	AT&T/Lucent (NJ)	33	Sharp (Japan)	8
4	IBM (NY)	116	Motorola (IL)	106	Xerox (CT)	22	Canon (Japan)	50	Xerox (CT)	28	IBM (NY)	4
5	NEC (Japan)	106	Canon (Japan)	96	Motorola (IL)	17	Matsushita (Japan)	47	Hitachi (Japan)	20	Toshiba (Japan)	4
6	Sony (Japan)	90	Matsushita (Japan)	62	NEC (Japan)	13	**WL Gore (DE)**	43	Toshiba (Japan)	18	**ATR (Japan)**	3
7	Matsushita (Japan)[5]	90	NEC (Japan)	59	Philips (Nether.)	13	Motorola (IL)	43	Matsushita (Japan)	16	HP (CA)	3
8	Texas Instr. (TX)	89	Toshiba (Japan)	58	**Verizon (VA)**	12	**LTS Lohmann (Germany)**	38	Sony (Japan)	16	Hitachi (Japan)	3
9	Nokia (Finland)	81	**Verizon (VA)**[6]	56	**ITT (NY)**	11	HP (CA)[7]	32	**ScanSoft (MA)**[8]	12	**Nanotronics (OR)**	3
10	**Qualcomm (CA)**	69	Nortel (Canada)	55	**VCS (TX)**	11	Toshiba (Japan)	31	Fuji Xerox (Japan)	11	**Stentor (Canada)**	3
11	Nortel (Canada)	64	Hitachi (Japan)	53	Matsushita (Japan)	10	Sony (Japan)	30	Canon (Japan)	10	**VOIS (CA)**	3
12	Hitachi (Japan)	57	Microsoft (WA)	48	Nortel (Canada)	9	Microsoft (WA)	28	HP (CA)	10	**AOL (VA)**	2
13	Philips (Nether.)	52	Philips (Nether.)	47	**BT (UK)**	8	**Bell South (GA)**	25	NEC (Japan)	10	**Centigram (CA)**	2
14	Fujitsu (Japan)	48	Sony (Japan)	44	Walker Digital(CT)	8	**Apple Comp. (CA)**	24	Philips (Nether.)	9	Intel (CA)	2
15	Toshiba (Japan)	45	Genesys (CA)	42	**MCI/Worldcom (DC)**	5	**BT (UK)**	21	Fujitsu (Japan)	8	**Interval Res. (CA)**	2
16	Siemens (Germany)	43	**MCI/Worldcom (DC)**	40	Hitachi (Japan)	5	Fujitsu (Japan)	19	**Educ. Train. Serv. (NJ)**	7	**MIT (MA)**	2
17	AMD (CA)	37	Ricoh (Japan)	38	Siemens (Germany)	5	Hitachi (Japan)	17	Ricoh (Japan)	5	Nortel (Canada)	2
18	Mitsubishi (Japan)	35	Intel (CA)	37	Intel (CA)	4	NEC (Japan)	16	**Exxon Research (NJ)**	5	**Patriot Scient. (CA)**	2
19	**Conexant (CA)**	32	**Apple Comp. (CA)**	35	Canon (Japan)	4	Verizon (VA)	16	General Electric (NY)	4	**Speech Plus (CA)**	2
20	Canon (Japan)	28	Fujitsu (Japan)	31	**Bell South (GA)**	4	Sharp (Japan)	12	Texas Instr. (TX)	4	**Telia (Sweden)**	2

Table 1 Top 20 patent recipients in each of the six SL technical fields (covers patents published in the period 1976-2002). The companies listed in bold do not belong into the World's top 100 R&D spenders.

5 Matsushita's patents include those of Panasonic.
6 Verizon's patents include those of Bell Atlantic.
7 HP's patents include those of Compaq and DEC.
8 ScanSoft's patents include those of Dragon Systems, L&H, and Speechworks.

5. Regional patterns

Figure 2 depicts the nationality of the patent assignees in each of the six technical fields under examination. In all technical fields US holds the first place, followed by Japan. Patents issued to companies or individuals located in the US account for approximately 60-70% of the total patents issued in a field (with the exception of language generation). For this reason, the patents assigned to American companies are also broken down into states. Within the US, California is at the first place for the four out of six technical fields, after loosing speaker recognition and language understanding to New York. The states of Texas, New Jersey and Illinois share the following places. SL patent activity in Massachusetts is now lower than those of Texas and Illinois, whilst in older data (Koumpis and Pavitt, 1999) the reverse was true.

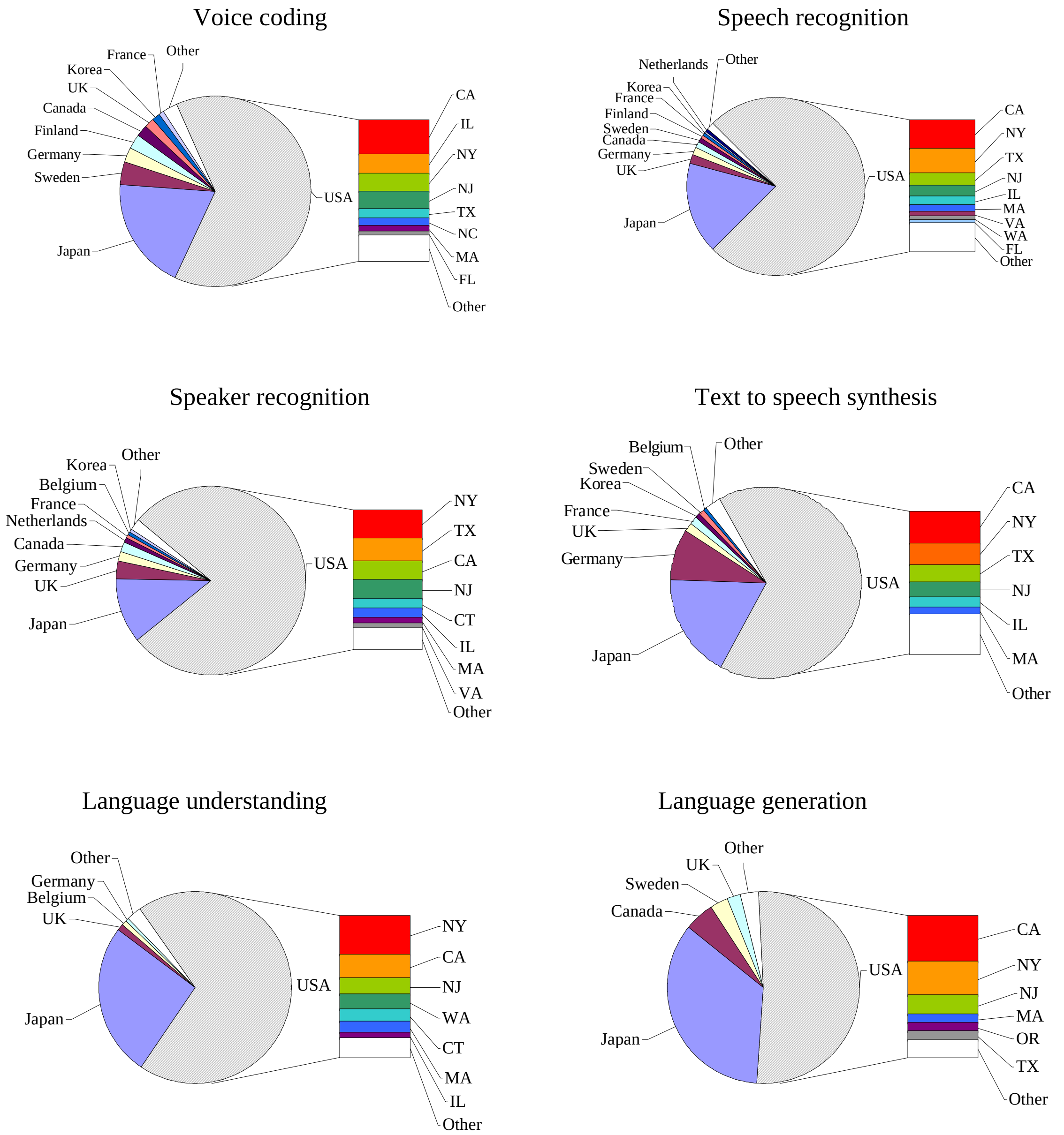

Figure 2 Regional distribution of patent assignees in each of the six SL technical fields.

Japan's average patent share is approx. 20%. The percentage is slightly higher in language understanding and generation but lower in speech recognition. The large patent portfolio of Ericsson in voice coding gives the third place to Sweden in this technical field. The same is true for Nokia and Finland in the fifth place. The UK has the third place in speech recognition, speaker recognition and language understanding fields. Germany has the third place in speech synthesis with a high percentage of the patents issued in this technical field. A similar case in the field of language generation is true for Canada. The analysis did not reveal a significant amount of patents assigned to companies or individuals in countries such as India or China which have recently become recipients of large amounts of IT outsourcing. This suggests that large multi-technology companies either do not outsource their SL R&D to those countries, or that they file patents using the addresses of their headquarters.

6. Technological trajectories

Companies follow technological trajectories, which are set by their choice of innovation strategies given their accumulated skills, and by the opportunities that they are capable of exploiting. The proliferation in patent activity (Figure 1) contradicts the fact that the fundamentals of SL science have not greatly changed in the past decade. At least it is not clear that any major technological 'breakthrough' took place during this period[9]. Some of the major SL fields have matured to the point that the efforts of many engineers are more oriented toward improving the effectiveness of existing methods rather than investigating radical new ones. This suggests that the increasing number of patents reflects the numerous follow-through inventions of the rapid diffusion phase of innovation rather than the crucial original inventions and innovations (Mowery and Rosenberg, 1979). However, companies continue to put a lot of effort in patents because in a new market such as SL engineering these can discourage newcomers and may be cross-licensed with their competitors.

It is widely accepted that a great amount of the technological effectiveness in the recent years is due to data intensive methods that have become available and affordable as a result of advances in computing[10] and large scale data collection efforts[11]. This trend started in the speech research community and was soon adopted by the natural language research community. The performance of systems employing statistical machine learning techniques to automatically acquire acoustic and linguistic characteristics from data continues to improve with training data size. However, there are concerns whether such orientation is sufficient to offer major progress in the future (cf. Bourlard et al, 1996, Greenberg, 2001), especially when scientists are expected to demonstrate consistent progress to justify their funding[12].

Technology s-curves drive the researchers to revisit the same themes decade after decade (cf. Foster, 1986). With each return technology is more sophisticated. Church (2003), suggested that SL science follows such cycles of consecutive empirical and rationalist approaches with a period of 20 years. Given that it took about 20 years to move from the empiricism of the 1950s to the rationalism of the 1970s and another 20 years to move back the other way to the empiricism of the 1990s, rationalism should be back in the 2010s. Although the practice has shown that

---

[9] Of the significance of dynamic programming, linear predictive coding (LPC), hidden Markov modelling (HMM) and stochastic language modelling.

[10] For instance, some of today's mobile phones combine more computing power than laptops a decade ago with the power of integrated high-speed wireless communication.

[11] For the organisations that could not afford large in-house data collections the contributions of the Linguistic Data Consortium (LDC) in US and the European Language Resources Association (ELRA) in Europe have been invaluable.

[12] As an element of their participation in annual evaluations of speech recognition, speaker recognition and spoken document retrieval tasks, funded by the US Department of Defense.

empirical and cognitive models are to some extent complementary rather than in conflict, whatever their origin (experience or reason) technological advances are the key to future success.

7. Market opportunities and challenges

Predicting the SL market size is not easy. Predictions from industry analysts have often proven too optimistic. But then again, the revenue growth of mobile telephony, for example, illustrates that technology markets can grow exponentially. Only the future will tell if SL remains a small market with niche applications, or if it will grow in a massive scale.

Vertical markets such as customer service, travel, insurance, medical, legal and financial have used speech recognition and speech synthesis profitably for a few years now and are enjoying a good return on their investment. Advances in voice coding are also related to the massive success of mobile telephony and to some extent of the Internet telephony. Standards such as VoiceXML, defining how to make Internet content and information accessible via voice and telephone, allow reduced time to market for technology adopters. But SL technology has the potential of applications well beyond the above.

Hands-free technology can enable the automation of many enterprise processes, such as manufacturing and inspection. SL technology can also have a positive impact in telematics. It not only offers convenience but also enhances safety by enabling hands-free applications for communication, control or navigation systems. Opportunities exist for both client/server and embedded solutions, based on whether access to remote services and information is required. The latter, known as voice portals, have not been able to build a viable business model that will take a considerable share of the dominant graphical web paradigm. The area with probably the greatest potential is indexing and retrieval of audio and video archives that would enable rich media to be searched using text keywords and phrases.

Though SL technology can reduce costs, improve customer relationship management, and provide fast and extensive access to data and information, it can also, if companies implement it for the wrong purposes or in the wrong way, pose challenges that extend far beyond technological issues. Many users may have difficulty adjusting to particular SL applications and may decide they prefer human operators or graphical user interfaces. Finding workable revenue models and demonstrating return on investment are major challenges for players in the SL industry. This is particularly true in the new economic reality, characterised by intense competition, informed and demanding customers, and relentless pressure to cut costs. A good way to avoid the traps of investing development efforts too heavily in technology is to start with the people who will use the SL product: What types of users are to be served? What problems do they encounter which the SL technologies will overcome? What are their goals and what tools and information they need to accomplish those goals? Answers to these questions can be used to develop personas, or fictional characters, who represent the needs of the various members of a target market. These personas will help companies make intelligent, well-reasoned decisions when it comes to defining and developing new products and services.

Effects of the recent economic downturn

Most SL developers and suppliers got caught up in the Internet/Telecom boom, and their business plans for new technologies focused less on actual any business and more on gaining investors. But, in markets regularly led by investors looking for overnight gains, rocky days were inevitable while the technology had to be worked out. Many companies found themselves under enormous pressure to generate income by exploiting internally generated technology or uniting externally generated technology. The pace of progress that markets subsequently demanded or could absorb was different from the progress offered by the technology. The SL engineering industry learned the hard way that it is important not to grow too fast. Gearing up to manage fast growth proved to be very painful, and companies later realised that they should have taken a more measured approach. Further, the SL engineering industry still tries to rebuild the public confidence from the Enron-like fraud by L&H, until then an industry leader and holder of some of the best technologies around, that drove the company into bankruptcy (cf. Baer, 2003).

Principal players have cut their R&D expenditure and now make modest investments so that they can be effective in niche markets. This trend has not yet been reflected as a slow down in patent activity as it takes on average 2-3 years from the time a patent application is filed till the time a patent is issued. A limited number of companies in the present tough economic environment can afford the resources required to attack more than one of the target markets which are currently segmented into three areas, i.e. server, desktop and embedded. It has also been confirmed that it is not necessarily a great advantage to be a technological leader in early stages of development of radically new products (Tidd et al, 1997), especially when the features valued by users are not clear either to the developers or to the users themselves. Companies without a previous involvement in SL technologies take advantage of the rock bottom prices and acquire other companies with long tradition in SL industry (the most notable case is the acquisition of SpeechWorks and the SL technologies division of L&H by ScanSoft). At the same time, the strategy that gains momentum among the leading specialised companies is to license their technologies to channel suppliers and to form alliances and partnerships. Small companies that have not managed to build capabilities and exploit them will most likely have to limit themselves to system customisation, e.g. to regional languages. In addition, few companies and university departments keep specialising on making fundamental advances in the core SL technology, and focus on challenging research tasks. However, given the significant amount of job losses across the SL engineering industry, many engineers and scientists returned to universities and this may act as a catalyst for the revival of rationalism in SL science.

8. Concluding remarks

Spoken language technologies are characterised by incremental innovation and have made steps toward widespread commercial acceptance. After few years of market uncertainty, crashing stock values and a significant amount of job losses, it is back to basics for the companies still involved in these technologies. Engineers continue to strive to increase accuracy and robustness, and minimise the memory footprints of their systems. As some fields are maturing (e.g. voice coding), major advances are also inevitable in the fields of natural language understanding and generation as a result of resource reallocation. More years of R&D are required before highly accurate interactive systems are available to drive applications for mass-market consumer products. As the range of technologies necessary for innovation has expanded and technologies have become more complex, companies can no longer cover all relevant disciplines. Companies will also need to determine when it is wise not to listen to investors and customers, when to invest in developing approaches that might initially harm performance, and when to pursue niche markets at the expense of seemingly larger and more lucrative ones. Finally, the engineers and those who fund them will need the patience to consider radical deviations from the scientific and technical mainstream in order to produce superior and commercially successful solutions, i.e. true innovation.

9. References

M. Baer, *"Immortal Code"*, Wired Magazine, Issue 11, Feb. 2003.

H. Bourlard, H. Hermansky, and N. Morgan, *"Towards Increasing Speech Recognition Error Rates"*, Speech Communication, pp. 205-231, May 1996.

R. N. Foster, *"Innovation: The Attacker's Advantage"*, Simon & Schuster, London, 1986.

S. Greenberg, *"Whither Speech Technology? - A Twenty-First Century Perspective"*, Proc. Eurospeech, Vol. 1, pp. 3-6, Aalborg, Denmark, Sept. 2001.

K. Koumpis and K. Pavitt, *"Corporate Activities in Speech Recognition and Natural Language: Another 'New Science' - based Technology"*, International Journal of Innovation Management, Vol.3, No.3, pp. 335-366, Sept. 1999.

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

K. Pavitt, *"The changing usefulness of university research: Dangers as well as opportunities"*, Nobel Symposium 123, Science and Industry in the 20th Century, Stockholm, Nov. 2002.

K. W. Church, *"Speech and Language Processing: Where we have been and where are we going?"*, Proc. Eurospeech, Vol. 1, pp. 1-4, Geneva, Switzerland, Sept. 2003.

D. Mowery and N. Rosenberg *"The Influence of Market Demand Upon Innovation: A Critical Review of Some Recent Empirical Studies"*, Research Policy, 1979 reprinted in N. Rosenberg, Inside the Black Box: Technology and Economics (Cambridge University Press, 1982).

IEEE Spectrum, *"Top 100 R&D Spenders"*, Vol. 40, No. 10, pp. 33-36, Oct. 2003.

J. Tidd, J. Bessant and K. Pavitt, *"Managing Innovation"*, Willey, 1997.