Courtesy of Mandy Mobley, Lynn Qu, Eric Sit, Jessica Wong. Used with permission.

Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
6.933/STS.420J Structure, Practice and Innovation in EE/CS
Fall 1998
December 11, 1998

Dragon Systems

Mandy Mobley
Lynn Qu
Eric Sit
Jessica Wong

Abstract

This paper gives an analysis of our six-week study into Dragon Systems Inc, a technological leader in speech recognition systems. It begins with the history of speech recognition technology from introduction in the 1950s up to present day. Next we discuss the company's roots its subsequent transition into a rapidly growing, international company. Lastly, we put the company into the framework of its market, and consider the question of its success. We conclude that Dragon Systems has been able to achieve many of their goals and therefore have attained some measure of success, but face an uphill battle in the coming years.

Contents

1 Introduction

1.1 The Mission

Dragon Systems, the Natural Speech Company[tm], is a leading worldwide supplier of speech and language technology. It is currently ranked as the seventh largest publisher of business software in the United States (PC Data, August 1998).

Our group first became interested in this company after reading an article in the September issue of MIT's Technology Review. Written by Simson Garfinkel, it introduced Dragon Systems as "the startup that beat Big Blue to the market," referring to the fact that Dragon one-upped IBM by being the first to come out with a continuous speech recognition product for the personal computer (PC). Garfinkel reported that it was founded back in 1982 by the husband-and-wife team of Jim and Janet Baker, and that there was no venture capitalists involved. After a bit of preliminary research, we discovered that many start-up companies in the speech recognition industry fail in their infancy. We became intrigued by Dragon's success, especially since Garfinkel also mentioned that there was not even a business plan in the beginning. We wanted to find out how Dragon Systems got to where it is today, whether it really is as successful as it appears to be, and what it will take for it to remain a leader in this field.

To achieve this objective, we visited Dragon Systems and interviewed several of their employees, including a co-founder, the Vice President of Research, and the Chief Architect Engineer behind some of Dragon's most popular products. We also researched published press material and papers, including Jim Baker's Ph.D. dissertation and various articles about the company and the industry. In addition, we interviewed other speech recognition experts in academia as well as industry. Although not necessarily looking for a clearly defined answer, we were curious to see what the results of our detective work would tell us in regards to the notion of success for Dragon Systems. We also hope to apply the material learned in class to this project, so that we can have a more practical understanding of some of the rather abstract terms defined in our readings.

1.2 320 Nevada Street, Newton, Massachusetts

The company headquarters is located in Newton, Massachusetts, less than 10 miles away from MIT. On a sunny morning in early December, we pay Dragon Systems a visit. The building itself used to be an old rope mill. Constructed out of red brick and rather tall and majestic-looking, it stands out in a residential surrounding. Janet Baker's assistant, Carlin Folkedal, greets us and cheerfully gives a tour of the premises. She points out the numerous dragon memorabilia which are ubiquitous in the building. "Jim and Janet have been collecting dragon themes since the beginning of time, so now when people who visit often bring gifts of dragon (themed) items." [1]

Each conference room is named after a mythical dragon, with the fable pasted on the doors. Japanese bamboo screens cleverly conceal the gray divides of the cubicles, and colorful oriental fans hang from exposed brick walls. The tall windows flood the high-ceilinged rooms with light, making the place seem larger than it actually is. Dragon Systems has recently bought another smaller building next door, and has proceeded to move its research and engineering groups over there. The original building now houses the financial, marketing, user support, human resource, and quality assurance departments.

The age of the employees seem to vary quite a bit, we rub shoulders with people in their early 20's as well as those in their 50's. As expected, the male-to-female ratio in the research and engineering divisions is rather unbalanced, although in the other divisions it seems to be about even. As we

[1] Carlin Folkedal, 12/02/98.

walk by offices and cubicles, we notice most employees are typing away at their keyboard rather than dictating to the microphone headset. "Most of us here *do* tend to use the keyboard more," admits Folkedal. "People are just more used to it... the Quality Assurance folks are really the ones who use the microphone all the time."[2]

Luckily for us, the biggest software trade show of the year, Comdex '98, has just ended. Dragon newest products were a big hit. The engineers are happy – they had been working for months preparing for these demonstrations in Las Vegas and it had paid off. With their input, we hope to weave an accurate story of the development of Dragon Systems.

The body of this paper is divided into four main sections. The first section discusses the history of speech recognition; the next section is about the Bakers, specifically, the period of their lives from their first involvement in speech recognition up to the founding of their company; the third section is about Dragon Systems the company, focusing on two other key players and briefly covering the history of the business; the next section is about Dragon's most impressive product to date – NaturallySpeaking, and why it has made such a big splash; the last section is about the present and future challenges Dragon Systems must face. Finally, we conclude with what we have learned about Dragon Systems, the speech recognition industry, and what it means for the former to be successful (in the context of the latter).

[2]Ibid.

2 The Problem of Speech Recognition

2.1 Introduction

The problem of speech recognition has been studied actively since the 1950s, but no universal solution has yet been found. In 1971, the Advanced Research Projects Agency supplied the funding that greatly expanded the field.[3]

At a basic level, recognition requires translating the analog information of an acoustic signal into digital information which can be processed by a computer. The problem is similar to the general problem of pattern recognition and faces many of the same challenges.

Success at this task depends upon a wide range of factors, including the input signal, the quality of the signal, and properties of the language. Different languages also pose their own special problems; some languages are harder to recognize than others. Homophones (distinct words which sound alike), tonality, and phonemes (the smallest units of speech) that are hard to distinguish all make the recognition task more difficult. Joel Gould, Chief Architect at Dragon Systems, remarks that of the languages for which Dragon Systems currently writes software, French is the most difficult, and Italian is the easiest to recognize. English falls somewhere in the middle.[4]

2.2 Types of Problems

Discrete speech recognition is the problem where words are separated by pauses. This is the older speech recognition problem, whose use was developed for use with telephony in recognizing spoken digits. A user must pause between each digit in order for the recognizer to accurately parse the speech signal. Because the words are already divided up, the recognition problem is reduced to single-signal to single-word matching. Today, such technology helps phone directory users to specify names of cities or to choose menu options without the use of a touch-tone phone.

Continuous speech recognition is the problem where words run into each other without pause, much like "natural" conversational speech. A person who is just learning a foreign language can appreciate the increased difficulty of recognizing words in continuous speech. Such a person does not have the extensive understanding and training with the language to easily create the word divisions that come easily to the native speaker. Lacking the complicated contextual knowledge of the human brain, a speech recognition system faces a very similar task as the newcomer to a foreign language.

Continuous speech is by far the harder problem of the two, which many thought had no attainable solution until the next century.

2.3 Traditional approaches to SR

Earliest solutions to the recognition problem used techniques of Artificial Intelligence to extract information of spoken words from acoustic signals.

Template-based Data files contain "prototypical" voice patterns of individual words. The problem or recognition is then finding the best match for the signal among the possible "tem-

[3]D.R. Reddy. *Speech Recognition by Machine: A Review.*

[4]Joel Gould, 12/02/98.

plates." This is an effective method for very constrained recognition tasks with small vocabularies.

Knowledge-based This approach introduces some "intelligence" into the recognizer by utilizing knowledge of language rules. The idea is that this knowledge better models human speech processing, and therefore adds more information to the task than simply analzing the acoustic signal. From this we can argue that Knowledge-Based recognizers can handle more general recognition problems than Template-Based.

Stochastic Stochastic, or probabilistic approaches have gained wide popularity in the last decade. Of these, *Hidden Markov Models*, discussed in the next section, are certainly the most popular. These models are more general than Template-Based and easier to weigh different knowledge sources than Knowledge-Based approaches.

Connectionist The newest player to SR, Connectionist approaches such as *Neural Nets* are still controversial. They have a similar flavor to stochastic approaches, in that they require training to optimize their strategy. But they also consider the interactions between many computing units, hoping to reflect the computation done by the human nervous system. This is a possible future paradigm for SR.

All but stochastic are Artificial Intelligence (AI) approaches, while stochastic approaches are considered purely mathematical. In the 1970s, rule-based recognizers such as KEAL and HEARSAY performed reasonably well. But over the last two decades, speech recognition has shifted dramatically to HMM approaches.

2.4 Hidden Markov Models - Theory

The concept of the Markov Model was developed by Russian mathematician Andrei Markov in the late 19th and early 20th century as a technique for describing the probabilities of being in any of a number of discrete states. A Markov Model is a representation of a process which defines *states* and *probabilities* for transitions between these states. A further criterion for a representation to be Markovian is that the transitional probability from state A to state B depends only upon states A and B, and not the "history" or sequence of transitions that led up to being in that state. Figure 1 shows a simple version of such a model. An example of the Markovian property is that the transition probability p_{23} depends only upon being in state 2 and not anything that happens in the past or future.

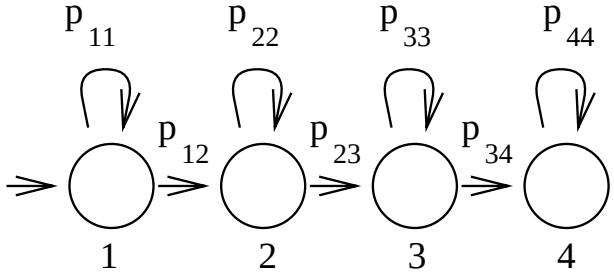

Figure 1: Simple Markov Model with four states. Each transition p_{ij} corresponds to the probability of moving from state i to state j in one step.

2.5 Markov Models in Linguistics

As of 1950s and 1960s, Markov Models were already being debated as a viable method of parsing speech from phonemes. In this time, Institute Professor Noam Chomsky argued that the Markov Model could not be effective because there were flaws in the model as applied to recognizing grammatical English sentences.[5]

Chomsky discredits the Markov-based system as a model for speech with the following arguments:

1. The Markov model does not separate the clearly grammatical from the clearly ungrammatical.
2. Successive improvements in the Markov model will not change its status with respect to (1).
3. There are types of sentences in the language which the Markov model cannot generate.
4. It is impossible to collect the data necessary to build a Markov model.

Experiments conducted and compiled by Damerau and others did show that a recognizer of grammatical English sentences using a Markov Model was in fact feasible, though too slow to be practical.

2.6 HMM Ph.D. Dissertation at Carnegie Mellon

In 1975, Jim Baker completed his Ph.D. dissertation at CMU as a cumulation of his background in statistical mathematics and interest in SR. "The DRAGON System" used HMM as its theoretical basis.[6]Damerau does not cite Jim Baker's work in any of his findings. On the other hand, Jim Baker was the first to actually implement the theory and argue that it was practical. In this sense, he was a pioneer in the SR field.

2.7 Emergence of HMMs

HMMs make good models because they account for different types of knowledge of the speech problem in a smooth, integrated manner. Compared to AI approaches, their performance was shown to be superior.

2.7.1 Benchmarking

Vital to the emergence of HMMs in industry was the role of regular *benchmark* evaluations conducted for:

- National Institute of Standards and Technology (NIST)[7]
- Department of Defense Advanced Research Project Agency (DARPA)
- National Security Agency (NSA)

[5]F.J. Damerau, *Markov Models in Linguistic Theory*
[6]Baker, James K. *Stochastic Modeling as a Means of Automatic Speech Recognition.*
[7]http://www.nist.gov/speech

In 1984, Dr. Duane Adams, then of DARPA, approached Dr. David Pallet[8] of NIST about implementing benchmark tests for the benefit of the DARPA speech recognition research community.[9]

One important fact about these evaluations is that while competitors never shared code, they were required to describe how their systems worked. Over time, it became clear that HMMs were beating out other systems.[10] Soon, other recognition systems began adopting HMMs.

2.7.2 Paradigm Shift

As we remember from our readings of Thomas Kuhn, a *paradigm shift* in a scientific revolution is a change of an accepted model or pattern to another, usually prompted by a crisis.[11] The development of HMMs in SR technology can be phrased in terms of a paradigm shift in the accepted model of SR engines, brought about by the change in modeling power and clear demonstrations of superiority.

As a result of these performance evaluations, all of the competitive speech recognition products today use HMMs as their base recognition engine. The original notion that HMMs are not appropriate for the task of SR because they are based purely on mathematics is solidly refuted by the proliferation and success of HMMs today.

The shift from discrete to continuous speech is also entwined with the emergence of HMMs. Consider the difficult problem of continuous speech. The template-based approaches are unable to handle this problem, because they are limited to single-word pattern matching. In other words, they only have enough power to do the discrete problem. Knowledge-based systems are able to handle strings of words, but don't provide a mechanism for determining where words begin and end. HMMs, on the other hand, suffer neither of these limitations. Because they are so robust and adaptive, they are ideally suited to the continuous SR problem. The translation between forming word hypotheses between words is a smooth transition from the problem dealing with single words.

[8] http://www.nist.gov/speech/pallettd.htm
[9] Dr. David Pallet (E-mail), Group Manager of Spoken Natural Language Processing Group, NIST
[10] Paul Bamberg, 12/02/98
[11] Thomas Kuhn, *The Structure of Scientific Revolutions*

3 Prelude

3.1 The Bakers

Jim Baker and Janet Maciver met at Rockefeller University in New York as graduate students. Jim was a mathematician interested in probability and statistics. Janet was a biophysicist. They became interested in the problem of speech recognition in Fall 1970, married in 1971, and have been a husband-and-wife team working in SR ever since.

3.2 The DRAGON system

Jim Baker's Ph.D. dissertation presents the results of his HMM-based recognition system called Dragon, named after the same affinity for dragons which permeates the history of the Bakers.

DRAGON consists of five programs: MAKDIC, MAKGRM, MAKNET, GETPRB, and DRAGON. *MAKDIC* creates a dictionary of acoustic-phonetic knowledge. MAKGRM creates a finite-state grammar representation. *MAKNET* combines the results of MAKDIC and MAKGRM into one integrated network. *GETPRB* creates probability estimates for phonetic transitions. Finally, *DRAGON* computes conditional probabilities to find the best match for acoustic input, thus "recognizing" speech.

MAKDIC This is a phonetic dictionary, or the "vocabulary" of the model, broken down into phonetic components. Here we will skip over the details of the translation between the raw acoustic signal into these "phones." The dictionary contains every word that can be recognized by DRAGON, and thus may be tailored to specific tasks. A generalized speech recognition task must have a large dictionary.

MAKGRM This component adds grammatical rules about language. These rules are defined by a Context-Free Grammar (CFG). A CFG is defined by a set of variables, terminals, and rules.[12]Terminals are the letters or words in the language. Variables are used as place-holders in the construction of the language. Rules restrict the ways that variables can be combined to form words or sentences. Thus, a CFG representation is a compact way of recursively defining rules of grammar.

MAKNET combines the results of MAKDIC and MAKGRM into a single Markov Model network with associated probabilities.

GETPRB estimates the conditional probabilities of producing a set of acoustic values, given a phonetic value. This is important in classifying the acoustic signal as its phonetic equivalent.

DRAGON consolidates all the knowledge generated by the previous parts, and computes conditional probabilities to determine the best match of an acoustic signal to actual spoken words. Furthermore, it does so by searching all possible network paths and chooses the optimal one. This is different from most other models which use some variation of best-first search of a classification decision tree. It may seem like such a strategy must take much more time, but DRAGON compensates for this by keeping its search space small. This feature has important implications. Because DRAGON searches the entire space of classification possibilities, its search is inherently time-bounded. Other models cannot provide a linear time bound, and

[12]Michael Sipser. *Introduction to the Theory of Computation.*

thus cannot guarantee that their computation will finish before the time for the next computation to start. Such a consideration is crucial to continuous speech recognition, which tries to accomplish its task in real-time.

The application of Hidden Markov Models to the task of speech recognition was controversial as an innovative method which ran askew from the direction of most speech recognition systems at the time.

3.3 IBM

Jim and Janet left CMU to work for $5\frac{1}{2}$ years on HMM speech recognition at IBM, Yorktown Heights. Here, they made significant progress into continuous SR. But they were unhappy only doing development and not getting the product out to real users.

3.4 Verbex (Exxon)

In 1979, they looked elsewhere to see where they could do development work with a real product. Verbex, a research branch of Exxon, was doing work on discrete speech.[13]

[13] Janet Baker interviewed by S.L. Garfinkel

4 Birth of the Dragon

In 1982, after coming out with its first continuous speech recognition product (in collecting spoken data over the phone), Exxon decided to pull out its efforts in this research area. Once again, the Bakers were left jobless. They were not worried about being unemployed for any length of time, however. "This was 1982, and up through 1982 the computer industry had never had a recession. So we just sort of took it for granted that if we needed to find jobs, we could," says Jim Baker.[14] But at the same time, they were aware of the fact the funding of speech recognition projects comes and goes, so there was no guarantee that they would necessarily be able to continue on a project which interested them. The only insurance for longtime pursuit of their passion would be in starting their own company.

"The way we looked on it was, if (the company) failed, then we would have to go find work and do something else. If that is the worst that could happen to us...(then) it is what we would otherwise have done if we haven't tried," says Jim Baker.[15]

With that spirit, as well as two young children and a big mortgage, the Bakers started Dragon Systems. They did not have a business plan, nor did they have any investments from venture capitalists. They used the money saved from their salaries at IBM and Exxon, which they figured would last them 15-18 months. They ran the business out of the basement of their home in Newton, Massachusetts.

4.1 Paul Bamberg

" If your count summer school and extension school, I believe I've probably taught more courses at Harvard that anyone in history," Paul Bamberg tells us.[16] Stoutly built and loud-voiced, with a full graying beard and ruddy complexion, Bamberg may not look or sound like your typical tweed-and-bow-tie Ivy League professor. But he has been a member of the faculty at the Physics Department of Harvard University for over 28 years. He received his Bachelor's in Physics from Harvard, then went to study at Oxford as a Rhodes Scholar. In 1967 he returned to Harvard and joined the faculty, teaching a wide range of courses including Pre-Med Physics, Math for Physics, and Theory of Algorithm. In addition to being a full-time professor, he is a "Dragon Fellow" and one of the earliest employees of Dragon Systems.

Bamberg has always been keenly interested in the industry. "Most people work thirty to sixty-hours in a full-time job, I figured it's a better deal for me to work two thirty-hour jobs rather than one sixty-hour one, since even if you work sixty-hours you only get paid for forty," he claims, half-jokingly.[17] He first met the Bakers at Verbex, where he worked under Jim Baker in the research department. When husband-and-wife team left, Bamberg was asked to take over the research department at Verbex. Not one to follow the managerial track, he gladly left to help start Dragon when the Bakers approached him.

Since then, Bamberg had worked on several of the main projects, most notably the original version of DragonDictate, the discrete speech-to-text software package, for which he wrote most of the code and used his own voice and vocabulary as training data. Until recently, he was the Vice President of Research at the company, although his lack of interest in management has not changed. Bamberg was the first to suggest a three-year rotation in positions, which Dragon Systems adopted.

[14] James Baker, 1997
[15] James Baker, 1997.
[16] Paul Bamberg, 12/02/98.
[17] Ibid.

Now Bamberg has relinquished his seat to Larry Gillick, and is happily heading the engineering team which produced the prototype for non-English recognizers.

Bamberg is also one of Dragon System's top recruiters. "When I hand back a hard test I'd given, I would say to the students, 'Now this test was difficult, if you got above eighty, that's pretty good; if you got above a ninety, that's really good; if you got above ninety-five, come and talk to me after class...'"[18] Additionally, he would also try to sell Dragon systems to those he works with. It is no wonder the company today employs a high number of Harvard graduates and former staff members.

Bamberg gives a rather interesting example of the practice of *heterogeneous engineering* in Dragon's early days, when resources were limited and the business itself was still located in the Bakers' home. The term is defined by Donald MacKenzie in *Inventing Accuracy* as "the engineering of the social as well as the physical world."[19] "The heads of these Fortune 500 companies would come to visit but we were still very small... once my wife dressed up as the 'company maid' so we could impress Alan Kay of Atari!" Bamberg recalls with great amusement. "We have more powerful partners now, but I'll never forget that."[20]

When asked about what changes he has seen in the company as Dragon grows large, Bamberg looks away, stares at the poster of an aerial shot of Sienna, Italy hung on his wall. He thinks for a moment before answering,"for a very long time, (Dragon was) a very small operation where everyone knew everyone else... we used to work holidays, but then took extended vacations. Now we can't do that anymore... have to act more like a real business."[21]

Although Bamberg has been in the industry for a long time, his background in the academia has led to a definition of success which is probably more modest than that of many other top engineers in a fast-growing company. "I really thought we had achieved success when we made big splashes across the world at the trade shows... I'm sure there are people who don't feel we're successful until we've driven IBM out of (the speech recognition) business. But I have much lower standards... In fact, Jim Baker and I decided early on that our goal was to set up an entity that will give a lifetime of interesting research, and we have already achieved that."[22]

4.2 Larry Gillick

Larry Gillick is the Vice President of Research at Dragon Systems. He oversees 50-60 research scientists, which includes both full-time and temporary staff. His background, like that of many at Dragon Systems, is in physics and mathematics.

Gillick got his Bachelor's degree in Physics from Swarthmore College. He had a hard time deciding between Molecular Biology or Statistics as something he would like to pursue afterwards. He proceeded to do research in protein synthesis at Columbia University for the next two years, but gradually his interest in this particular area waned. "I found it hard to keep the glass clean," Gillick jokes.[23] He came to MIT for a Ph.D in Mathematics, then taught at MIT as well as Northeastern. In addition to lecturing, he also analyzed clinical trials and consulted for the industry.

In late 1984, Gillick saw an advertisement placed by Dragon Systems in the MIT Gazette. The young company intrigued him and he called up Jim Baker and arranged for a site visit. Gillick was impressed with what he saw there, "I really thought what Jim Baker was working on seemed

[18]Ibid
[19]Donald MacKenzie.
[20]Ibid
[21]Ibid
[22]Ibid
[23]Larry Gillick, 12/02/98

like the perfect application for math," says Gillick.[24] He joined Dragon in January, 1985, and began working on isolated-word recognition. His main task was to figure out how to build an acoustic model for large-vocabulary speech recognition witout too much memory. In fact, in its early days, Dragon lacked the cash to purchase top-of-the line computation facilities and had to focus on making their underlying models extremely efficient. This strategy would prove itself to be a invaluable a decade later.

After fourteen years at Dragon, Gillick now heads the research department, where tasks are slightly different. "(There is) a strong project management element to my position," he says, "I've become used to doing stuff by remote control... I do miss writing code..." [25] He does still do some hands-on research – reading papers and analyzing algorithms for developing new recognizer models – as well as represent his department in the executive meetings.

Slenderly built, with thick glasses and dark, curly hair, Gillick is a self described "typical intellectual," In his spare time, he enjoys reading books on history and philosophy, plays the piano, and listens to jazz as well as musicals. "But I still like math," says Gillick with a smile.[26] Not surprisingly, the white-board in his office is filled with scribbled math formulas and graphs, stacks of paper lay around in loose piles on his desk and on the table in the reception area of his office (one of which is interestingly titled "Paranoia in Taiwan").

Gillick is quick to jump on the defensive side when asked whether he feels that the company has more bureaucracy now that it has gotten so much larger and have more divisions. "We *are* more departmentalized now, but Dragon as an organization has very permeable departments... I would say we have a much richer intellectual environment now... when Dragon was small we couldn't afford to have a director of marketing, for example..."[27]

Gillick thinks Dragon Systems is successful, but not quite in the same sense Bamberg does. "We have been pioneers, we did (isolated-word and continuous speech recognition) long before anybody else did... that's been an important part of our culture." However, this is a culture that requires a lot of work to maintain. "We need to continue to come out with imaginative products and get there before others do... and continue to be a major player in this market, which is growing rapidly."[28]

His group is going to help to contribute to such a definition of success. Gillick says the atmosphere in the research department is very much like that of a graduate school. "We try to place high value on innovation, and not let bureaucracy get in the way." The key difference, he points out, lies in the ultimate goal. "Although we are... academic in culture, (we are) also very much a business, Our research is not so much for publishing interesting papers, but more (for) producing a product."[29]

The fact that their research is product-oriented is a key element to their current success. There have been companies who did not survive because they did research for research's sake. Gillick agrees, but he has other ideas as well when it comes to what makes a business successful. "I also think it is important that the employees are happy... We enjoy our work a lot. I think it is a lot of fun to develop the technology... although it is research people don't feel like that they are slaving away. (We) are having a good time."[30]

[24] Ibid.
[25] Ibid.
[26] Ibid.
[27] Ibid.
[28] Ibid.
[29] Ibid.
[30] Ibid.

4.3 Early projects and technologies

Dragon Systems stayed very small for the first 8 years. They did not pay any salaries in the beginning, and even turned away potential employees for fear of endangering these people's employment. Gradually, as the revenues started to come in, they started receiving salaries of $100 each.

The focus in the early days was on conducting research and licensing their technology to customers. Dragon Systems landed their first major customer in 1984. Apricot Computers, Ltd. of the UK, boasted a wireless PC integrated with a microphone which, with Dragon's recognizer, offered limited speech recognition for command and control. Unfortunately, Apricot went out of business soon after the product release. However, Dragon continued to form more alliances with companies such as IBM, releasing IBM VoiceType in 1991; as well as with Microsoft, who licensed Dragon's technology in Windows Sound System, which was shipped in 1992.

In 1986, Dragon was awarded the first of a series of ARPA contracts and began working on large vocabulary, speaker-independent continuous speech recognition. Government funding constituted a large percentage of Dragon's research budget in the first decade or so, until, according to Bamberg, the focus was shifted to broadcast communications.

Also in 1986, Dragon released its own product – VoiceScribe 1000, which was a 1000-word speech recognition system for IBM PC-compatible computers. The system was hardware-dependent, the package included an IBM PC/XT or PC/AT compatible peripheral board. It was well-received by critics. PCWeekly said it was "a complete and versatile isolated-word voice-recognition system" and "bridged the wide gap between usefulness and exploratory potentials." The magazine also lauded its documentation as "one of the best introduction to voice-recognition technology available." A year later, DragonWriter 1000 was released. It had a more advanced package which claimed to reduce speech recognition errors by one-half. Both of products would later be replaced by more sophisticated systems running on a PC.

In 1990, Dragon Systems came out with DragonDictate-30K, a 30,000-word speech recognition system. It was the first comerically available larte vocabulary sppech-to-text system for general purpose dication running on a PC. It was the company's most high-profile release to date, and gained instant recognition in the industry. [31]

[31] see Appendix

5 Bending the Trajectory: CSR

Although the Dragon Dictate software was an amazing technical feat, it was not able to penetrate the general market. One of the main reasons was that although many people could enter text with Dragon Dictate faster than they could type, nobody enjoyed being forced to pause between each spoken word.

What users really wanted was continuous speech recognition (CSR); that is, users wanted software that could recognize natural speech. However, experts in speech recognition thought that continuous speech recognition would be reality for a long time. These experts had defined a *natural trajectory*, or "a direction of technical development that is simply natral... corresponding to the inherent possibility of the technology."[32]

Dragon Systems did not believe in the natural trajectory of speech recognition that these experts had declared. They set out to prove these experts wrong.

5.1 Joel Gould

With a bushy mustache and dark, dusty hair, Joel Gould enters the room. He walks with intrepid determinism and a unalterable, strong ego. When he speaks, his voice is loud and clear with authority as if there could be no dispute on what he has to say.

He graduated from MIT with a Bachelors in Electrical Engineering in '83 and a Masters of Engineering in software in '84. He worked in Cambridge Scientific Center until they closed. In 1992, a friend introduced him to Jim Baker.

His motivation for joining Dragon Systems was to "develop products, especially highly visible products which made a difference." "My goal two years ago was to develop a product that I would find on the shelf of Egghead," Gould said.[33]

Joel Gould was the man who was faced with the supposedly impossible task of continuous speech recognition. Given the challenges described earlier, Gould needed to keep close tabs on both the technical developments and the marketing developments in speech recognition.

However, Dragon's quest for continuous speech recognition was not without problems as will be described shortly.

5.2 Technical Difficulties

Continuous speech recognition had long been a holy grail in the field. It was a challenging problem that modern AI techniques seemed unable to solve.

5.2.1 Processor Power and Memory

Conventional wisdom at the time thought that computer speed and memory would retrict the development of continuous speech recognition for at least five or ten years. However, the Bakers calculated their own rate at which computer speed and memory were improving and concluded that top-of-the-line desktop machines should be able to do continuous recognition within a few years.

Also, Dragon Systems did not have the same computational resources as its much larger competitors. Gould says, "Dragon Systems is poor so we have to make our algorithms more efficient to accomplish [our goals] in a limited amount of time"[34]

[32] MacKenzie, Donald, *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance*
[33] Joel Gould, 1997.
[34] Joel Gould, 12/02/98.

5.2.2 Insufficient Speech Recognition Models

Using classical pattern matching techniques to solve the continuous speech problem proved to be nearly impossible. One of the main difficulties was that these techiques had difficulty in distinguishing words. Also, searching for the correct match in speech patterns in a database took an exponential amount of computation time.

However, the usage of Hidden Markov Models became an increasingly popular methodology in discrete speech recognition. The advantage of the Hidden Markov Model is that it can be expanded to do continuous speech recognition while pattern matching cannot. Hidden Markov Models turned the speech recognition problem into one that could be solved in polynomial time, not exponential. Thus, the Hidden Markov Model was the technology that enabled continuous speech recognition.

Luckily, Dragon Systems already had a head start in this field with their Dragon Dictate product which used Hidden Markov Models. Also, Dragon had been working on developing continuous speech recognition for the government for a while using DARPA funding.

5.3 Social Difficulties

Although there were many technical difficulties in developing Dragon NaturallySpeaking, there were perhaps even more social difficulties involved. Heterogeneous engineering of these social problems would play a large role in the success of Dragon NaturallySpeaking.

5.3.1 Baby Dragon

Although successful with their Dragon Dictate product, the Dragon Systems company was still a small company. In order to realistically ship Dragon NaturallySpeaking on time, they needed to beef up their research department. However, Dragon had very little capital in order to make such an investment.

The Bakers realized this was critical problem. Therefore, while Jim Baker was forming his new development team to build Dragon's first continuous speech recognizer, Janet Baker brokered a deal with California-based hard disk manufacturer Seagate Technologies to buy 25 percent of Dragon's stock. The company used the cash to expand its engineering, marketing, and sales forces. Within a year, Dragon had the largest speech research team in the world with more than fifty scientists and engineers. This research team would prove critical in developing Dragon NaturallySpeaking.

Today, Dragon Systems is still a very small company and could easily be eaten alive by its much larger competition. According to Peter Ffoulkes of the market research firm Dataquest, "There is only so much coverage they can do with the people they have got and the funds available to them. And therefore they have got to partner. It is sort of crazy advice: just be bigger. Which they are trying to do as much as they can."[35]

5.3.2 Military Spelling versus Natural Spelling

Often times in dictation, it is necessary to spell out words. However, speech recognition systems have a difficult time distinguishing between letters like *v*, *b*, *e*, and *d*. It is the same reason why people have difficulty understanding names over the telephone.

Dragon Dictate solved this problem by using the military alphabet. This seemed to solve the problem having individual letters output. For example, normally, if you dictate the sound 'be' out

[35]Peter Ffoulkes, 1997

of context, "B.", "b", "be", or "bee" can show up. To force the program to output "b" or "B", saying 'Bravo' instead works. See Figure 2 below.

A=Alpha B=Bravo C=Charlie D=Delta E=Echo F=Foxtrot
G=Golf H=Hotel I=India J=Juliet K=Kilo L=Lima
M=Mike N=November O=Oscar P=Papa Q=Quebec R=Romeo
S=Sierra T=Tango U=Uniform V=Victor W=Whiskey X=X-ray
Y=Yankee Z=Zulu

Figure 2: Military Alphabet

The only problem with using this system is that customers do not like it. They feel uncomfortable in spelling their words using the military alphabet. What they would like is to use natural spelling to accomplish this.

Dragon System's implementation of natural spelling required a complete rethinking and redesign of the algorithm and the interface as a whole, Gould said. It required a mental shift to say that the goal is not to solve the accuracy of spelling, but to solve the customer's program. If the software gets a word wrong like a customer's last name, you want to be able to spell it naturally.

Natural spelling is inherently a difficult problem. High accuracy in natural spelling was achieved by combining techniques of the overall recognition of letters, considering only the sequences which match words in the dictionary.

The recognition and solution to this problem required a shift in design philosophy. Dragon worked to alter their product so that it solved what the problems of the customer, not just the problems of the engineer.

5.3.3 Re-engineering the engineers

According to Gould, if you had gone to any of the major speech research houses two years ago and asked them if something like Dragon Naturally Speaking was possible, they would say no. In fact, Gould says, "The hardest thing I did working on Naturally Speaking was convincing the research department that it was possible."[36]

Research is paid to be conservative, Gould says. Dragon's research department worked on large vocabulary general-purpose continuous speech recognition systems for decades, and each year they made noticeable improvements towards the ultimate goal of being able to recognize anybody's voice with complete accuracy in real-time.'

However, when Dragon calculated the time when Dragon NaturallySpeaking would ship, the recognition performance would not be adequate given the projection of expected improvements based on past history.

"It was my job to remain optimistic," Gould said. "I firmly believed that if research was focused on the problem of improving recognition performance for a commercial product, and not on all the problems of meeting the needs of the government, then te recognition performance would improv significantly more than would be predicted by historical evidence," Gould said.[37]

Gould said to the engineers, "We will do this... this is possible." "We cannot sell a product with less than X percent accuracy so we *will* get the accuracy better than X percent."[38] He worked with

[36] Joel Gould, 1997.
[37] Joel Gould, 12/07/98.
[38] Joel Gould, 12/07/98.

his team to design additional algorithms to work around the known limitations in the continuous speech recognition.

5.3.4 Keeping It Away from the Competitors

Because Dragon was a small company determined to be an innovators in speech recognition technology, being first in the continuous speech recognition market was extremely important. They would have difficulty in competing with a large competitor such as IBM.

"Dragon is an innovator, but small," states Gould. "Should we pre-announce our [future] plans, our bigger competition could [get us]." Some of Dragon's bigger competitors have the manpower to clone new products once they're released. Gould explains, "Take NaturallyOrganized, for example. Every time we release a new product we cut our own throats."[39]

While engineers at Dragon Systems were developing a continuous speech recognition product, they needed to work closely with Microsoft's code in order to insure that their program would run on Windows. However, when Dragon Systems asked Microsoft to fix their code, Microsoft engineers wondered why Dragon Systems needed these "tags" to be available so soon, since a continuous speech recognizer was years or decades away.

Dragon knew that they had to be careful about how they handled this relationship with Microsoft. So they actively hid the fact that they were very close to having a continuous recognizer. They created an e-mail address for a fictional research department to create the illusion that their queries did not revolve around an actual product.[40] In reality, the people on the e-mail list were the exact people who were developing NaturallySpeaking.

5.3.5 Making It Natural

Since the beginning of the modern computer revolution, users have interacted with their computers through the keyboard. The mouse was invented in 1968 by Douglass Engelbart, but it was not until the advent of graphical user interfaces many years later that the mouse became truly popular.

Similarly, in the 1980s, nobody really *talked* to their computers. Speech recognition was a novelty and nothing more for the majority of computer users. Thus, Dragon was faced with the task of making the idea of talking to your computer seem natural.

This might suggest why all the Dragon products after Dragon Dictate begin with the word *Naturally.* This word implies Dragon's intention of making their speech recognition product natural and ubiquitous.

5.3.6 Change in Market

With the move to developing a continuous speech recognition product came a new market. The market for continuous speech recognition products would be very different from that of the discrete recognition product. Thus, Dragon had to tailor its products for the new market.

Although Dragon Systems did not have an official company focus for DragonDictate, the development team designed Dragon Dictate primarily for the accessibility marketplace, Gould says. There was an overriding emphasis on making sure that Dragon Dictate for Windows would allow someone to use their computer 100 percent hands free. This overriding emphasis impacted significant aspects of the development of the products and shaped its characteristics.

[39] Joel Gould, 12/02/98.
[40] Mike Elkins, 11/20/98

Later, in Dragon Dictate for Windows, the product was designed to be flexible enough to be used in a variety of environments. This was because Dragon did not know exactly what markets (beyond the accessibility marketplace) in which Dragon Dictate would sell. Therefore, Dragon designed the program to be configurable by end-users so that it would be able to be adapted to markets which Dragon did not consider during development.

Finally, in Dragon NaturallySpeaking, the focus even more. Gould says, "I like to say that with Dragon NaturallySpeaking, the goal was to design a program that I would use." To that end, Dragon deferred issues of accessibility to later versions and instead concentrated on creating a product which would appeal to a broader audience for general text creation.

Initially, the goal for Dragon NaturallySpeaking was to make it faster for people to create documents and e-mail messages when they did not type fast, Gould says. In particular, Dragon targeted professionals such as lawyers, doctors, etc. Because of that, the design of the product was different. Dragon also tended to be less worried about being able to do absolutely everything by voice since Dragon discovered that even in the accessibility market, people are comfortable with using the mouse to manipulate their computers.

Instead, Dragon concentrated on making the text creation process as easy and simple as possible, Gould says. In fact, the first version of Dragon NaturallySpeaking did not even have user configurable commands.[41]

It is interesting to note Dragon System's attitude towards its Dragon Dictate product. For example, Bamberg has said that there will always be a market for its discrete speech recognition product. One such market would be the assistive needs market who considers Dragon Dictate to be a savior in their day-to-day computing needs.

On the other hand, Gould strongly asserts that Dragon Dictate is "dead." He says that the continuous speech technology of NaturallySpeaking is far superior and obliviates the need for discrete. Perhaps one way to explain this discrepancy is to notice Gould's background as lead architect of NaturallySpeaking.

5.3.7 Change in Marketing Channels

The change in customer type created yet another problem. The traditional marketing techniques for Dragon Dictate would not apply to Dragon NaturallySpeaking because they would reach the wrong type of customers. Thus, the change in customer type necessitated a complete change in marketing technique.

DragonDictate for Windows was primarily sold through a value added reseller network (VAR). Each VAR specialized in a small number of high maintenance customers, such as the disabled. The VAR would customize the product for its customer. The price of Dragon Dictate for Windows was set to allow the VAR sufficient profit to justify small volume sales.

However, with Dragon NaturallySpeaking, the primary marketing channel was retail. With retail products, the price has to be lower (although most of the downward price pressure was a result of competition rather than marketing decisions). But Dragon NaturallySpeaking does also have a VAR product and a VAR channel. This is the Professional Edition which has many of the same characteristics as DragonDictate for Windows such as configurability. Thus, resellers can customize the software for its specific customer need.

[41] Dragon had always intended to add this. User configurable commands later appeared in Dragon NaturallySpeaking Professional Edition version 2.0 in the fall of 1997.

5.3.8 Change in Corporate Attitude

Thus, as described earlier, in order to develop Naturally Speaking, Dragon Systems needed to make the transition from a company that is research based to a company that is market driven, Gould said.

Up to this point, from a technology standpoint, the research department, research climate, and the corporate culture were geared toward solving certain classes of problems which were characteristic of the government projects that Dragon worked on. In fact, some of design decisions in Dragon Dictate are suggestive of the DARPA funding such as the usage of the military alphabet for spelling.

This argument can also be applied to Dragon Dictate in the assistive needs market. People in this market used Dragon Dictate out of necessity. Customers came and willingly put in the time to learn all the quirks of Dragon Dictate, such as its military alphabet. Whatever design decisions Dragon's engineers made, the customers had to deal with them.

However, the market for Dragon NaturallySpeaking, or continuous speech recognition, is much different. Although the government was interested in this technology, it was not nearly as important to them as the discrete technology. Also, the assistive market did not really have as much need for continuous speech recognition as that of discrete speech recognition. The customers now were mainstream professionals who had been using the keyboard to interact with their computer all their life. These professionals had no immediate need for speech recognition technology.

All of a sudden, Dragon Systems actually had to find its customers. Moreover, they had to *convince* them to use Dragon NaturallySpeaking. This necessitated a change in corporate attitude and direction. The marketing department became even more important. As a result, Dragon began to shift from a primarily research-centric organization to a customer-centric organization.

5.4 Success

Dragon NaturallySpeaking made its first major debut at the computer industry trade show COMDEX 1997. To the surprise of many in the computer industry, Dragon Systems stole the show. Its NaturallySpeaking turned out to be an incredible success. In fact, it won *PC Week's* Best of 1997 COMDEX award.

Since then, Dragon Naturally Speaking has won more than 44 awards worldwide, including *PC World's* World Class Awards Most Promising Software Newcomer (May 1998), *PC Week's* Best of COMDEX, (Nov. 1997), 1997 *Popular Science's* Best of What's New Awards, Grand Winner: Computers and Software (Dec. 1997) and *PC Magazine's* Technical Excellence Award (Dec. 1997).

In the calendar year 1998, Dragon Naturally Speaking sales exceeded competitor's offerings in both revenues and units according to PC Data (April 1998).

Thus, the computer industry was clearly shocked at how this small start-up company could perform such an incredible technological feat and sweep one the industry's most prestigious trade shows. Dragon Systems had defied the speech recognition experts and bent the natural trajectory of continuous speech recognition to their own favor.

However, Dragon System's stardom and early market advantage in the speech recognition field would soon be compromised by the emergence of large, aggressive corporate competitors.

6 Competition Emerges

6.1 IBM

IBM, commonly referred to as "Big Blue," is a Goliath in the computing industry. With more than 25 years of research and close to 100 patents, IBM has built a team of speech recognition specialists worldwide, dedicated to the development, support, marketing, and sale of IBM speech solutions in ten languages.[42]

IBM is currently a close competitor of Dragon Systems; however, in the early days, IBM was a customer of Dragon's. In 1991 IBM announced IBM VoiceType Version 1.0, a 7,000 word speech-to-text system, under license from Dragon Systems. [43] Regarding the licensing deal with IBM, Paul Bamberg exclaims, "Jim [Baker] made very good deals for us."[44]

IBM also receives support from government agencies like DARPA, similar to the funding they give to Dragon Systems. "IBM is richer and doesn't have to sell products," explains Joel Gould, so they put a lot of money and effort into research. "The core technology [behind speech recognition] comes out of IBM's research division. People live and breathe phonemes and HMMs. However, their products reflect that they focus on recognition and not usability. They don't stretch beyond speech recognition."[45]

Dragon takes a different approach. "We develop and engineer [products] very closely," states Gould. He adds, "Dragon's products are good because our developers talk to the customers." [46] Gould, although he likes to take the credit for most of this 'public relations work,' was not the driving force behind listening to Dragon's customers.

Nancy Gardner, a former employee of Dragon Systems, was Dragon's Usability Engineer and worked in Dragon Systems' Human Factors division. She spent over 2 years with Dragon Systems before leaving to go to Harlequin, a software company in Cambridge doing criminal investigation projects. Recently, she was laid off from Harlequin, but is planning to start up her own human factors consulting business. A Master's degree graduate from MIT, her husband, Richard Shyduroff, is one of the co-founders and co-directors of the MIT Entrepreneur's Club.

Dragon System's Human Factors division is responsible for thinking about the "out of the box experience," which is the first 5 minutes for the user. "Janet Baker's underlying interest is in the user's experience," Gardner reveals, "but the company's position is going after the mass market."[47] This shift in priority focus may explain why, after Gardner left, Dragon didn't hire another Usability Engineer. Gould decided when she left that he would do it all himself. Whether he puts as much time as Gardner did into getting customer feedback is left to be determined. However, he most certainly claims all the credit for understanding the customers and satisfying their demands.

6.1.1 ViaVoice

According to Patri Pugliese, Research Operations Manager and Government Contracts Manager at Dragon, "When Dragon announced [NaturallySpeaking], IBM announced [ViaVoice] before Dragon Systems went to sell."[48] Dragon Systems' announced Dragon NaturallySpeaking, the world's first general purpose large vocabulary continuous speech recognition product, at a press conference held

[42] IBM Software - Speech Recognition, http://www.software.ibm.com/speech/
[43] see Appendix.
[44] Paul Bamberg, 12/02/98.
[45] Joel Gould, 12/02/98.
[46] Ibid.
[47] Nancy Gardner, 12/06/98.
[48] Patri Pugliese, 12/02/98.

simultaneously in New York City and San Jose, CA on April 2, 1997.[49] When it released in mid-June, IBM was pressured to get their own competing product out on the market. Six weeks later, IBM released version 1.0 of ViaVoice.

ViaVoice, at that time, had many advantages over NaturallySpeaking. ViaVoice released at$99, a much lower price than NaturallySpeaking's $695. Instantly, a 'price war' was waged between the two competitors. After all the price-slashing, Dragon NaturallySpeaking Standard 3.0 now sells for $99.95 and IBM ViaVoice currently sells for $49.95.[50] ViaVoice also offered the attractive ability to dictate into several word processors and spreadsheets. However, Michael Caton stated in his January 21, 1998 PC Week article, "The ability to dictate directly into mainstream business applications, including Lotus Development Corp.'s WordPro 97 and Lotus 1-2-3 97 and Microsoft Corp.'s Excel 7 and Word 97, provides a small but significant productivity boost and more convenience for those with a general interest in using dictation."[51] Dragon, however, responded to IBM's challenge and in early 1998 released Dragon Point & Speak, a sub-$59 speech recognition dictation solution that enables users to easily create text by speaking into their favorite Windows application, including word processing applications, email and online chat rooms.[52]

Compared with NaturallySpeaking, ViaVoice placed a close second. "Dragon Systems has a much more integrated product [than ViaVoice]," claims Gardner, who conducted her own comparative analysis between Dragon NaturallySpeaking and IBM's ViaVoice after leaving Dragon Systems. "Joel Gould is a very forceful personality in the development team. He makes sure that Dragon's products are more integrated. IBM, because it is a large organization and due to their corporate culture, they came up with a less integrated, more awkward product. The initial experience [for users] is less positive."[53] Despite voluntarily leaving Dragon Systems due to one-too-many conflicts with Gould, Gardner states, " I would go [with] Dragon Systems because they are a smaller company [than IBM]. That's where the success is."[54]

6.2 Lernout and Hauspie

Lernout and Hauspie, although a fairly new competitor in the continuous speech industry, has a history working with and competing against Dragon Systems. In the past Dragon Systems did business with L&H. Dragon struck up a deal for L&H to develop a continuous speech recognition product written for European-speaking users. However, at the conclusion of their work, L&H decided to put out a product of their own, despite the contract with Dragon Systems. According to Paul Bamberg, "Janet was quick to straighten them out."[55]

Kurzweil Educational Systems, a leading developer and provider of educational software and systems for reading assistance, competed in the discrete speech market against Dragon through most of the 1980s. L&H recently acquired Kurzweil on September 1, 1998, and have continued to market Kurzweil's existing products targeting vertical markets while using Kurzweil's technology to come out with new products of their own.

[49] see Appendix.
[50] CompUSA: The Online Superstore, http://www.compusa.com
[51] Michael Caton, IBM Takes Dictation Further, PC Week, January 21, 1998.
[52] See Appendix.
[53] Nancy Gardner, 12/06/98.
[54] Ibid.
[55] Paul Bamberg, 12/02/98.

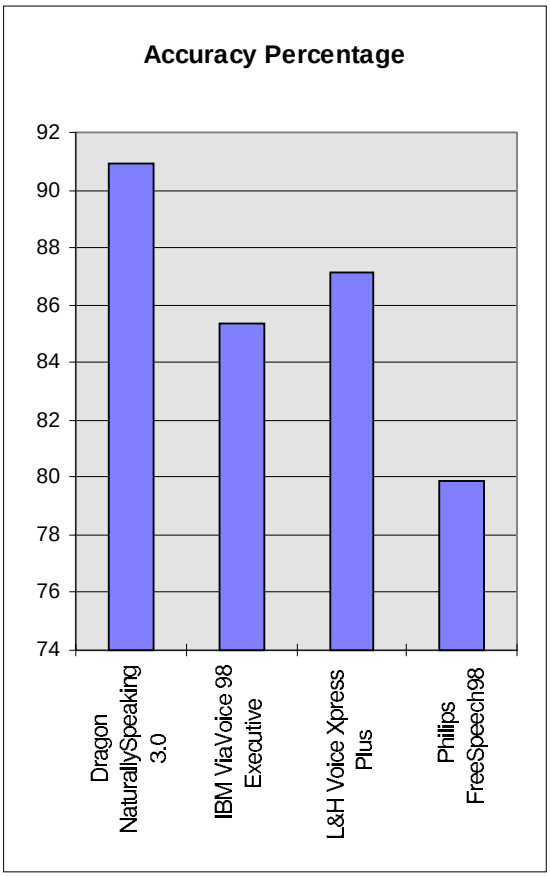

Figure 3: Results from the accuracy tests performed by PC Magazine.

The highlight of Voice Xpress is its natural-language command capabilities. Though NaturallySpeaking and ViaVoice also offer this feature, Voice Xpress provides more flexibility.[58] Instead of using the typical 'command and control' discrete speech commands, such as "File, Print" to print a document, VoiceXpress allows the user to give the computer commands in a more natural way, such as "Print pages 3-5 of this file in landscape layout," thus making it appear that the computer

[56]L&H Press Release(19980428) Voice Xpress(TM), http://www.lhs.com/news/releases/19980428_VoiceXpress.asp
[57]Paul Bamberg, 12/02/98.
[58]PC Magazine: Speech Recognition, http://www.zdnet.com/pcmag/features/speech98/rev3.html

actually 'understands' what the user is saying. However, there are some drawbacks to this feature. "L&H's stuff is very impressive, but there's a cost," explains Pugliese. "The more free-flowing the commands, the harder to tell if it is a command."[59]

Although L&H's technology is gaining respect, their marketing department needs some work. "[L&H is] slow on uptake in terms of marketing and getting a continuous speech product out," Gardner explained. "[However,] I know they're a serious contender."[60] Once they did boost up their marketing strategy, they made a few mistakes which came back to haunt them. "L&H set expectations too high by advertising 'no training required,'" states Gould. "[As a result,] about 60% of their customers got frustrated and sent [the product] back."[61] Gould also mentioned that L&H's product claimed speech understanding, a feat which no one has yet been able to master.

6.2.2 Alliance With Microsoft

On September 11, 1997, Microsoft Corporation and Lernout & Hauspie Speech Products announced a broad strategic alliance designed to accelerate development of the next generation of voice-enabled computing on the Microsoft Windows platform.[62] Specifically, L&H will develop applications for currently available and future versions of Microsoft's speech application programming interface (SAPI). L&H will also continue to pursue its goal of providing the broadest range of speech technologies in multiple languages for horizontal, and, in particular, vertical markets such as the medical and legal sectors, and will embed speech applications into special purpose hardware devices.[63] "A horizontal market is the broadest kind of market – for example, Windows is a horizontal market – anyone who buys a PC can buy the product and use it. A vertical market is one that is designed for a particular domain – for example, a medical dictation product," explains Kevin Schofield, a Senior Program Manager in Microsoft Research.[64] Kevin states regarding the L&H alliance, "Microsoft, as you are aware, has a deep expertise in building global platforms. L&H has a great depth of expertise in languages, and in particular speech technologies for many languages. There was a natural fit between our companies because we both wanted the same thing: a worldwide speech platform."[65]

However, the motives behind the alliance are questioned by many. "Microsoft will come out with their own technology, not L&H's," predicts Gould. Bamberg adds, "Microsoft's own recognizer is better than L&H's." Paul speculates that Microsoft, since they are now L&H shareholders, probably aligned with L&H to make a tidy profit off the stock which rose as a result of the alliance.

6.3 Microsoft

"I don't worry about anyone except Microsoft," Paul Bamberg bluntly states. Bamberg fears the day when "every Windows user is forced to pay an annual license fee, and for only an additional fifteen dollars, the user can have a complete dictation system integrated into their operating system. We can't compete," says Bamberg of this sticky, but realistic, situation.[66]

[59] Patri Pugliese, 12/02/98.
[60] Nancy Gardner, 12/06/98.
[61] Joel Gould, 12/02/98.
[62] Microsoft and Lernout & Hauspie Announce Strategic Alliance In Support of Voice-Enabled Computing, http:/www.microsoft.com/presspass/press/1997/sept97/msl&hpr.htm
[63] Ibid.
[64] Kevin Schofield, 11/24/98.
[65] Ibid.
[66] Paul Bamberg, 12/02/98.

Microsoft, the "sleeping giant" as Jim Glass of MIT's EECS department so referred to them, has been slow to challenge speech recognition product manufacturers with their own competing product. However, they do hold the reigns on the competition by forcing them to comply with Microsoft's Speech Application Programming Interface (SAPI). "Microsoft wins in the middle-ware," explains Gould, "because they require everyone to comply with SAPI in order to produce a product that will operate in the Windows environment."[67]

In addition to SAPI, Microsoft has its own plans to introduce a speech recognition product on the market. "Windows 2000 plans to include our [Text-To-Speech] engine," states Schofield.[68] And Microsoft Research is busy at work on 'Whisper', their proprietary speech recognition product. Microsoft is currently researching advanced acoustic and language modeling algorithms to improve accuracy, usability, and efficiency for spoken language understanding and conversational UI, developing conversational systems for telephone and home platforms, and investigating advanced dictation systems for multiple languages.[69]

"Our business plans do not assume we will replace Microsoft," Gould exclaims realistically. "However, Microsoft does not address vertical markets, such as the law and medical industries. Maybe [in the future] we will make products compatible with Microsoft's [software] that will address vertical markets. Dragon System's software will probably be on most operating systems in the short-term. Microsoft will most likely replace us in the long-term."[70] Even though Dragon might not be directly competing with Microsoft in the vertical markets, they will be competing against L&H, a 'Microsoft-preferred' company.

Despite the threat that Microsoft poses, Larry Gillick hopes "to be a major player in the developing speech recognition market. But we can't aspire to have it all. That's not healthy for the industry." Microsoft, as we may verify by the anti-trust law suits, takes a different approach. Microsoft may force Dragon out of the PC market and into other markets where speech recognition has an application. Some of these applications will be discussed in the next section.

6.4 Others

6.4.1 Philips

"Philips FreeSpeech98 is a limited speech recognition solution. Its accuracy lagged far behind [NaturallySpeaking, Voice Xpress, and ViaVoice]. And FreeSpeech98 lacks many of the features found in other programs, such as support for multiple users, true modeless operation, and voice macros," states PC Magazine in a review on Philips FreeSpeech98.[71]

FreeSpeech98 aims for the 'at home' users, brandishing a limited 30,000-word working dictionary. Upon announcing the release of FreeSpeech98 in October of 1998, Philips also announced an agreement with UbiQ, a leading sales and marketing organization that helps PC OEMs incorporate new technologies into their products, allowing UbiQ extensive rights to market FreeSpeech 98 to the world-wide OEM community. UbiQ has predicted that one million FreeSpeech 98 licenses will be sold to PC vendors for inclusion in their desktop PCs and laptops by the end of 1999.[72]

Despite Philip's plans to take a larger share of the speech recognition market, PC Magazine's Craig Stinson doubts its anticipated customer acceptance. "Although [Philips FreeSpeech98] is

[67] Joel Gould, 1997.
[68] Kevin Schofield, 11/24/98.
[69] MSR Research Areas: Speech Technology, http://research.microsoft.com/srg/srproject.htm
[70] Joel Gould, 12/02/98.
[71] Craig Stinson, Philips FreeSpeech98, PC Maganize, October 20, 1998.
[72] Philips Speech Processing - Press Office - Press Release, http://www.speech.be.philips.com:100/bin/owa/psp_s_press?xid=773

the lowest-priced product in [the speech recognition] category, FreeSpeech98 does not include a microphone, is the least accurate of the programs we tested, and lacks many features. You'll be better off spending the extra money to get a more powerful solution."[73]

6.4.2 Future Competitors

The speech recognition field is still in its early stages and the market is just now continuing to grow. "After 3 or 4 years, there will be a lot more speech recognition players," predicts Gould. However, Dragon is prepared to take on the competition. Janet Baker, the President of Dragon Systems, states, "Multiple competitors are very important in improving and accelerating the rate of technology advancement. Building technology is therefore building the market. There's plenty of room for more players. I welcome the competition."[74] Pugliese adds, "The main thing we have going for us is higher accuracy with the standard user. Overall, people want accuracy. We don't spend time with features that people won't use."

[73] Craig Stinson, Philips FreeSpeech98, PC Maganize, October 20, 1998.
[74] Janet Baker, 12/06/98.

7 Speech Recognition Applications of the Future

Research in speech recognition is still in its early stages. According to Larry Gillick, researchers will continue to build and improve recognizers which are capable of recognizing different kinds of speech, from "careful" to "natural" to "broadcast." "Continuous speech recognition is a technology that more people find useful. However, there continues to be a role for [discrete speech recognition] technology in some consumer devices," notes Gillick.[75]

One of Janet Baker's long-term goals is "to make speech recognition truly ubiquitous."[76] Gillick adds, "[Janet is] unusual because she understands the technology and sees the possibilities for its application."[77] Janet's clear vision makes her somewhat ahead of her time since she is able to predict the value of speech recognition development. Maybe IBM was right when they referred to the Bakers as 'premature.'[78]

Unfortunately, speech recognition systems are still hampered by the rate of growth of the hardware. Although hardware is advancing in astronomical proportions, it still does not allow 'resource-hogging' speech recognition algorithms to run efficiently. "Speech recognition isn't exciting if it can't be made 'real time,"' according to Gillick. "Speech recognition took off only when continuous speech recognition in real time was possible."[79] Thus, Dragon sometimes has to shelf new improvements in the recognizer until the hardware needed to run the algorithm is readily available in the common household PC. "Often we have to scale things down because of hardware limits," explains Gillick. "However, we usually have more stuff 'waiting in the wings."[80]

7.1 PC Desktop Integration

Both Joel Gould and Janet Baker compare the adoption of the mouse to the adoption of speech. "When I was in college [roughly 15 years ago], we were considered 'keyboard-centric' computer users," explains Gould. "Today most college students are 'mouse-centric.' I predict that ten years from now, the [college student population] will be largely 'speech-centric' users."[81] Janet adds, "A [Boston] Globe article today [December 6, 1998] talks about the adoption of the mouse taking thirty years. Speech recognition is way past the halfway point."[82]

7.2 Handheld Devices

Olympus came out with the D1000 Digital Voice Recorder, which can be purchased with IBM's ViaVoice speech recognition software included in the package. "Using this package ($300 street), you can plug the recorder into your computer's audio card line-in adapter and have your computer type what you recorded," states Matthew Gravern in a PC Magazine article.[83] Dragon has taken the mobile market one step further with Dragon NaturallyOrganized.

"We have started to move off into the mobile space," says Joel Gould. "Janet often says, 'Keyboards are getting smaller, but fingers aren't.'"[84] Dragon NaturallyOrganized, the world's first

[75] Larry Gillick, 12/02/98.
[76] Janet Baker, 12/06/98
[77] Larry Gillick, 12/02/98
[78] Simpson L. Garfinkel, Enter the Dragon, Technology Review, September/October 1998.
[79] Larry Gillick, 12/02/98.
[80] Ibid.
[81] Joel Gould, 12/02/98.
[82] Janet Baker, 12/06/98.
[83] Matthew Gravern, New Speech Technology, PC Magazine, October 20, 1998.
[84] Joel Gould, 12/02/98.

Natural Speech Productivity Assistant, was announced November 16, 1998 by Dragon Systems at COMDEX 98. NaturallyOrganized runs on Dragon NaturallyMobile digital recorder, the world's first digital recorder specifically designed for speech recognition.[85] This device functions much like a PalmPilot, except the user interacts with it via voice. The user is able to record notes to himself, such as "Remember to buy milk" or "Send an email to Professor Mindell", and then later attach the recorder to his PC to process his tasks. NaturallyOrganized is able to distinguish between tasks, emails, and appointments, and then executes the user's commands after his final approval. The first version of Dragon NaturallyOrganized supports the ACT! contact management system, and future releases are planned to support a additional applications, including Microsoft Outlook, Goldmine, and Timeslips.[86]

7.3 Language Translation

Many companies are interested in the language translation area of the speech recognition market. Lernout & Hauspie have made a significant impact with the L&H iTranslator, which offers a complete range of translation on the internet to and from English, German, Spanish, French, Arabic, Japanese and Korean via a web server and client/server architecture.[87]

Dragon has also been engaged in a long-term project with DARPA concerning language translation products. Paul Bamberg recalls, "I got an email from a long-time supporter at DARPA saying 'We have a phrase transaction system that we are interested in using in Bosnia. It could use a speech interface. We would like you to demonstrate the interface next week."[88] This translation device, referred to as Babelfish, allows U.S. soldiers to speak selected war-time phrases, such as "Do you know where any dangerous materials of any kind are stored?", into the device. The device recognizes the phrase via Dragon Dictate, enters the keystrokes which corresponds to the correct phrase, and plays the phrase back to the user for verification before vocalizing that same phrase in the Bosnian's native language. "We actually had a contract in place about two weeks after the initial demo," exclaims Paul. "This system was quite well known around DARPA. [It] was on someone's list at DARPA of the major achievements of the previous year."[89]

Paul Bamberg is optimistic about this area of research. He predicts, "Ten years from now [we will have] one spoken language in, another spoken language out."

7.4 Asian Languages

Victor Zue, head of the Spoken Language Systems Group at MIT's Laboratory for Computer Science sees the Asian market as a major future playing field for speech recognition. "If you have a language in which keyboard entry is difficult, then it's not just a matter of RSI. This is why Intel and Microsoft are setting up research centers in China," Zue explains.[90] Keeping up with their larger competitors, Dragon is quick to respond to this possible market my conducting their own research into a Mandarin recognizer.

[85] Dragon Systems Current News, http://www.dragonsys.com/frameset/currentnews2.html
[86] Ibid.
[87] L&H iTranslation, http://www.itranslator.com/
[88] Paul Bamberg, 1997.
[89] Ibid.
[90] Victor Zue, 11/25/98.

7.5 Speech Understanding

Speech understanding, a feat that L&H falsely claims with their natural speech command structure, is a goal for which every speech recognition organization is working. Right now, most of the products on the market are simply 'dumb' dictation products, products which can recognize what a user says and transcribe it, but has know idea what the user 'means.' Many companies, as well as government and educational institutions, have research departments doing various research in the area of speech understanding.

Victor Zue is working on JUPITER, a conversational system that provides up-to-date weather information over the phone. A caller may ask questions like "What's the weather like in Beijing?" or "Is is raining in Los Angeles?", and JUPITER will interpret and respond to the questions with 80% accuracy for first-time users (over 95% for experienced users).[91]

"Speech understanding is going to happen," assures Joel Gould[92]; however, Larry Gillick seems a bit more pessimistic. "Technology as good as a person may never arrive. It could take hundreds of years."[93] Whenever this next step in the speech recognition industry arrives, it will most certainly create another paradigm shift which will reinvent the way in which society interacts with its computers.

[91] JUPITER, (617) 258-0300, http://www.sls.lcs.mit.edu/sls/whatwedo/applications/jupiter.html
[92] Joel Gould, 12/02/98.
[93] Larry Gillick, 12/02/98.

8 Conclusion: Can Dragon Claim Success?

Dragon Systems today is considered a leader in the speech recognition industry. Jim and Janet Baker have pioneered the dominant HMM technology and are well-respected in the SR field. Dragon's products continue to win numerous awards and rave reviews at trade shows across the globe. Do all of these facts determine that Dragon Systems can rightly claim success?

Success is a very subjective concept whose definition varies from person to person. "Success, for a company, is measured sustainable growth and profitability," states Joel Gould. "For me, I'm successful if 5 years from now, the way [people] interact with their computer via speech is the way I invented it."[94] Janet Baker laughed when she learned of Joel's personal definition of success, she adds her own definition: "Success is reaching your goals. Jim and I have been successful in creating a large-scale vocabulary continuous speech recognizer. But that's not to say that we've solved all the problems. We haven't worked ourselves out of a job."[95]

A company's success can often be measured by their mission statement's goals. Janet Baker states, "[Dragon Systems'] mission is to create leading speech technology which will become ubiquitous."[96] However, measuring a company's success by what is listed on their mission statemet is sometimes an unrealistic measurement. Many "lofty goals," such as the one aforementioned, are listed, intended to serve as a long-term goal toward which the employees will work. Often these long-term goals are what unite employees in a corporate environment.

Jim Baker and Paul Bamberg agree that the goal of creating Dragon Systems was to "set up an entity [containing] a lifetime of interesting research problems."[97] The environment so created is "sort of like an academic environment," according to Larry Gillick. "We talk about ideas and have lots of books on the shelves."[98]

The goal of creating the entity of which Bamberg spoke has been achieved, but the ubiquity of speech recognition has yet to arrive. Simply judging the success of Dragon Systems by looking only at the completion of its goals, we cannot come to a definitive conclusion. However, success is not some kind of threshold which needs to be reached. It also needs to be maintained. Dragon Systems has reached a point where they can be viewed as a successful company, but to stay that way is no small feat.

For one thing, the alliance formed between Microsoft and Lernout & Hauspie means there is a more priviledged outside party when it comes to developing speech recognition products to run on the Windows platform. In fact, Microsoft itself poses a serious threat. The software giant has an excellent research staff in speech recognition and has claimed that they will soon been shipping speech technology applications (most likely their own) with Windows 2000. Once customers can get Microsoft's application for free with Windows, will they be willing to pay for Dragon System's packages, even if it's a superior software? Probably not. In response to this inevitable threat, Dragon has already begun to shift its business model to target new markets where speech recognition has a place, such as mobile, hand-held and language translation devices.

From our overall analysis of Dragon Systems and the speech recognition industry, we feel that Dragon has what it takes to "fight the smart fight." Dragon Systems continues to be a pioneer in the speech recognition industry, laying claim to some notable 'firsts,' starting with Co-Founder Jim Baker. Jim was the first to introduce Hidden Markov Models as a useful algorithm in speech recognition technology. In 1997, Dragon Systems had the first Hidden Markov Model-based product

[94] Joel Gould, 12/02/98.
[95] Janet Baker, 12/06/98.
[96] Ibid.
[97] Paul Bamberg, 12/02/98.
[98] Larry Gillick, 12/02/98.

on the market. 1998 brought with it the release of Dragon NaturallyOrganized, the world's first productivity assistant.[99] Gillick describes Dragon Systems as "pioneers" because they "tend to do things first. This is an important part of our culture. We win over our competitors because we're clever, think carefully, and get there first. We will continue to come out with very imaginative products and get them out first."[100] According to Bamberg, "Dragon has achieved success in big splashes all over the world with their products at various trade shows. People line up to look at our booth."[101] Joel Gould adds, "Dragon is very aggressive in what we try to do in speech."

Another key reason for Dragon's current success deals with their relations with their employees. Dragon keeps their employees happy so that they enjoy going to work every day and don't mind staying that one extra hour. Cornelia Sittel, a Quality Assurance representative who regularly brings her dog Heidi to the office, states of Dragon's environment, "It's nice to get a lot of work done and not be stressed out."[102]

In addition to allowing pets of all kinds to accompany their owners at work, Dragon imposes no dress code and allows most employees to exercise flex hours. The current company culture is strongly related to the company culture that existed ten years ago. "Everyone is driven by the desire to develop technology that **sells**," explains Gillick. "[The research department] won't pursue an idea only if it's interesting. It must also be valuable in the marketing sense." A high priority is put on ideas and innovation, and bureaucracy is not allowed to get in the way of their progress. "I

[99] See Appendix.
[100] Larry Gillick, 12/02/98.
[101] Paul Bamberg, 12/02/98.
[102] Cornelia Sittel, 12/02/98.

have a good time building products," continues Gillick. "We enjoy our work, and that's one reason why we've done so well. I like to create an atmosphere of research where they're having a good time."[103]

"Dragon's growth has paralleled the growth of the PC," claims Gillick, regarding the recent recruiting efforts of Dragon Systems. "We are tracking it as it gets more powerful."[104] Despite Dragon's rapid growth in the past few years, they still claim to hire the best people. "We are picky about who we hire," Gould exclaims. "We go for the cream of the crop. We hire intelligence, not skills."[105] And within that 'cream of the crop,' there are many employees whose backgrounds represent many different fields of expertise. Having an eclectic group is a luxury limited mostly to companies of Dragon's current size and larger. "A small company can't afford to have a division of marketing," explains Gillick.[106]

Although growth created departments within Dragon Systems, thus creating a stronger division between research and applications, Gillick and others claim the departments communicate well with each other. "They're very permeable," claims Gillick.[107] "The research and engineering departments all talk to each other rather easily," says Patri Pugliese. "There is an openness to innovation."[108]

Dragon Systems is effectively meeting their short-term goals, as well as making the necessary moves to reach their long-term goals in the future. These moves include expanding across various markets, pushing themselves to be the first to introduce new products, keeping their employees happy, creating an environment conducive to working effectively, hiring bright new talent, and communicating freely between departments. However, there is no guarantee of perennial success, and indeed, there is not even a clear definition of success in this context. So we are almost left with a 'cliff-hanger.' Joel Gould sums up the 'Dragon' outlook on the future with this last quote: "Dragon is not a place you go to get rich quick. For people who want to work with hot technology and bright people, and to see products on shelves, Dragon Systems is 'it.'"[109]

[103]Larry Gillick, 12/02/98.
[104]Larry Gillick, 12/02/98.
[105]Joel Gould, 12/02/98.
[106]Joel Gould, 12/02/98.
[107]Ibid.
[108]Patri Pugliese, 12/02/98.
[109]Joel Gould, 12/02/98.

A Appendix

References

[1] Baker, James K. "Stochastic Modeling as a Means of Automatic Speech Recognition." Carnegie-Mellon University, 1975.

[2] Caton, Michael. "IBM Takes Dictation Further", PC Week, January 21, 1998.

[3] Damerau, Frederick J. "Markov Models and Linguistic Theory." Mouton & Co., The Netherlands, 1971.

[4] Garfinkel, Simpson L. "Enter the Dragon", Technology Review, September/October 1998.

[5] Gravern, Matthew. "New Speech Technology", PC Magazine, October 20, 1998.

[6] MacKenzie, Donald. "Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance"

[7] Pallet, D. Group Manager of Spoken Natural Language Processing Group, NIS. (Email)

[8] Reddy, D.R. "Speech Recognition by Machine: A Review." IEEE Proceedings, 64(4):502-531, April 1976.

[9] Sipser, Michael. "Introduction to the Theory of Computation." PWS Publishing Company, Boston, 1997.

[10] Stinson, Craig. Philips FreeSpeech98, PC Maganize, October 20, 1998.

[11] Baker, James. Interviewed by S.L. Garfinkel, 1997.

[12] Baker, Janet. Interviewed by S.L. Garfinkel, 1997.

[13] Baker, Janet. Co-Founder and President, Dragon Systems. 12/06/98.

[14] Bamberg, Paul. Interviewed by S.L. Garfinkel, 1997.

[15] Bamberg, Paul. Vice President and Dragon Fellow, Dragon Systems. 12/02/98.

[16] Elkins, Mike. Senior Engineer, Dragon Systems. 11/20/98.

[17] Ffoulkes, Peter. Interviewed by S.L. Garfinkel, 1997.

[18] Folkedale, Carlin. Assistant to the President, Dragon Systems. 12/02/98.

[19] Gardner, Nancy. (former) Usability Engineer, Dragon Systems. 12/06/98.

[20] Gillick, Larry. Vice President of Research, Dragon Systems. 12/02/98.

[21] Gould, Joel. Interviewed by S.L. Garfinkel, 1997.

[22] Gould, Joel. Director of Emergining Technologies, Dragon Syste ms. 12/02/98.

[23] Gould, Joel. Email, 12/07/98.

[24] Lettvin, Jerome. Professor Emeritus, Massachusetts Institute of Technology. 12/06/98.

[25] Pugliese, Patri. Research Operations Manager and Government Contracts Manager, Dragon Systems. 12/02/98.

[26] Schofield, Kevin. Senior Program Manager, Microsoft Corporation. 11/24/98.

[27] Sittel, Cornelia. International Quality Assurance, Dragon Systems. 12/02/98.

[28] Zue, Victor. Head of the Spoken Language Systems Group at MIT. 11/25/98.

[29] CompUSA: The Online Superstore, http://www.compusa.com

[30] Dragon Systems Current News, http://www.dragonsys.com/frameset/currentnews2.html

[31] IBM Software - Speech Recognition, http://www.software.ibm.com/speech/

[32] JUPITER, (617) 258-0300, http://www.sls.lcs.mit.edu/sls/whatwedo/applications/jupiter.html

[33] L&H iTranslation, http://www.itranslator.com/

[34] L&H Press Release(19980428) Voice Xpress(TM), http://www.lhs.com/news/releases/19980428_VoiceXpress.a

[35] Microsoft and Lernout & Hauspie Announce Strategic Alliance In Support of Voice-Enabled Computing, http:/www.microsoft.com/presspass/press/1997/sept97/msl&hpr.htm

[36] MSR Research Areas: Speech Technology, http://research.microsoft.com/srg/srproject.htm

[37] PC Magazine: Speech Recognition, http://www.zdnet.com/pcmag/features/speech98/rev3.html

[38] Philips Speech Processing - Press Office - Press Release, http://www.speech.be.philips.com:100/bin/owa/psp_s_press?xid=773