4

AUGMENTING OUR SENSES

Watson's performance on *Jeopardy!* was remarkable. The machine searched through millions of pages of information for possible answers and came up with solutions in about three seconds. But Watson was only searching one kind of data—textual information. All Watson could do is read and reason. But what if it possessed machine equivalents of all of the human senses. What if it could see, hear, touch, taste, and smell?

These new capabilities are crucial because of the explosion of new kinds of data. The rapid growth of low-cost sensors makes it possible to monitor everything from leaks in water systems, to the performance of electrical grids, to the location and movement of merchandise in stores and warehouses. The proportion of the world's data that comes from such sensors is expected to increase from 11 percent in 2005 to 42 percent in 2020. At the same time, the amount of video, still images, and audio content on the Web is exploding. It's estimated that seventy-two hours of video content are uploaded to YouTube every minute.[1]

Yet most images and audio on the Web are searchable only via metadata—words that are manually typed into forms by humans. So it's vital to develop systems that can recognize images and sounds more like humans do. As the world becomes more complex and as its complexity becomes increasingly decipherable via analytics, we will need machines that can comprehend what's going on in the physical world and provide an expansion of human senses.

SENSING THE FUTURE

It's also going to be mind-blowing. At IBM Research, just for fun, we recently conducted an exercise where we asked specialists in each of the machine-sensing fields to imagine advances that would likely come in the next five years. They envisioned a handful of scenarios. Some of them are practical; others, intriguing:

SIGHT

Computers will help us sort through the 500 billion photos that are taken each year, finding relevant images not only through metadata but also by understanding what's going on in the images. We'll train the computers to know what to look for, and they will learn more on their own as they interact with the images and information pertaining to them. Imagine a computer program that combs through all of the family photos you and your relatives have taken and assembles a family album, in chronological order, complete with descriptions of the who, what, and where of each photo.

HEARING

Computers will help us understand what we hear. We will record sounds that babies make when they're hungry, sleepy, in need a pat on the back, or in distress. Then the system will map the sounds to brain, heart, digestion, and lung activity. As a result, we'll be able to translate their cries and whimpers into messages adults can readily respond to, extracting language from prelinguistic sounds.

SMELL

Computers will be able to detect diseases via smell. Similar to how a breathalyzer can detect alcohol from a breath sample, sensors in hospital rooms or even smartphones will be able to identify molecules in the air and match them with known biomarkers associated with liver and kidney disease, diabetes, and tuberculosis. People won't have to wait for their annual physical to learn that something might be amiss. They'll get an early warning.

TOUCH

You will be able to reach out and touch through your smartphone. When you're shopping online, haptic technologies, which use electro-mechanical means to simulate the feel of things, will make it possible for you to distinguish among different materials, textures, and weaves. You'll be able to feel a sweater, jacket, or upholstery sample right through the screen.

TASTE

Using human beings as sensors, we're exploring the use of data analytics and learning technologies to create recipes that both taste good and are good for you. The technology analyzes food in terms of how chemical compounds interact with one another coupled with knowledge about which combinations produce pleasant tastes. Then it matches those factors with the health characteristics of the compounds. The results could not only delight people but help them lose weight and live longer and happier lives. The technology represents an advance in the field of computation creativity. It's the first time that chemistry has been combined with cultural artifacts (recipes) and psychological principles to create something that's brand new.

Today, robots with rudimentary sensing abilities generate a lot of excitement in the popular media. The Mars rover *Curiosity* operates on the surface of the Red Planet semi-autonomously and "sniffs" the air for methane gas searching for signs of life. Experimental self-driving vehicles are capable of using vision and other senses to navigate through cities without colliding with buses, running over pedestrians, or getting speeding tickets. Yet, as of now, robots remain firmly in the von Neumann computing paradigm. They must be programmed in advance by people to deal with nearly every situation they encounter.

Already, learning systems are playing important roles in advances in speech recognition and image recognition. Some voice-recognition systems, for instance, get better at understanding an individual's manner of speech the more they interact. But the scientific community is just at the beginning of making machines that learn like humans do, so we're just scratching the surface of the potential of computer sensing. "The excitement with machine learning is figuring out how we build a computer to recognize the things in the world and name them," says John Smith, a senior manager in intelligent information management at IBM Research. "A machine needs to understand something about the world to understand pictures and videos, and one way that it learns about the world is by examining pictures and videos."[2]

All sensing and learning technologies have basic elements in common, starting with the sensor, a device or capability that can detect something about the physical world. Then there's the ability to extract features from the sensory input, such as colors or edges in visual data or phonemes in voice recognition. The next step is modeling. The computing systems use a variety of analytical tools to group together features that constitute an entity, such as a person, a building, a song, or an odor. Then they apply semantics to place the entities in categories, such as locations, events, and behaviors.

Machine learning is employed at each step. The system is instructed by the software to recognize features, entities, and semantic categories. Ultimately, because we can't teach machines to identify everything and every relationship among things, they have to be capable of inductive reasoning, in which they derive propositions from specific examples. That's the point at which the cognitive systems begin to advance beyond the confines of von Neumann computing. They don't just do what they are told; they figure out what to do on their own.

In order to make this great leap and become true thinking machines, the cognitive systems of the future will integrate information from multiple sensing technologies. Today, as scientists labor to create machine technologies to augment our senses, there's a strong tendency to view each sensory field in isolation as specialists focus only on a single sensory capability. Experts in each sense don't read journals devoted to the others senses, and they don't attend one another's conferences. Even within IBM, our specialists in different sensing technologies don't interact much. Yet if machines are to help humans understand the world, they have to make sense of it and communicate about it in a way that's familiar and comprehensible to humans. This integration of data from various sensing technologies is beginning to happen in multimedia and visual analytics, where vision and sound are correlated. But that's just the start of what will be required in the next era of computing.

Imagine a security system in a major port city a dozen years from now. On the docks, video cameras record the movements of people and equipment and ships. Microphones record sounds. Sensing devices sniff for smoke or noxious gases. It would be prohibitively expensive to have people constantly monitoring all of these streams of information, but computers can integrate the information in real time and detect patterns and correlations that indicate, say, that an arsonist has boarded a ship late at night, broken a window, and set a fire. The authorities might be alerted so quickly that they arrive on the scene not only before the fire has spread but also before the arsonist has fled the scene.

All of these new tasks will require a tremendous amount of data processing. Millions of bits of digital information are contained even in a single frame. A computer has to extract various features, classify them, and build a semantic understanding to make sense of that single image—and a full video contains thousands of these. That's a lot of data to process.

Yet we constantly size up the world around us as a matter of routine. The human brain is a marvel. A mere 20 watts of energy are required to power the 22 billion neurons in a brain that's roughly the size of a grapefruit. To field a conventional computer with comparable cognitive capacity would require gigawatts of electricity and a machine the size of a football field. So, clearly, something has to change fundamentally in computing for sensing machines to help us make use of the millions of hours of video, billions of photographs, and countless sensory signals that surround us.

In order to move forward, we need a new kind of chip, now on the drawing boards, that embodies the principles of cognitive computing in the realm of the senses. The goal is not to reproduce a human brain in electronic circuitry; rather, it is to take inspiration from the way the brain works and use it as a model for designing chips and systems. By combining the principles of neuroscience with techniques from nanotechnology, scientists can design chips and systems that not only analyze massive amounts of complex information from multiple sensory modalities at once, like brains do, but also can dynamically reprogram themselves as they interact with their environments.

So how would a brain-based technology like this work? In the simplest terms, the brain operates like this: Specialized cells, the neurons, receive, integrate and send messages, which are electrochemical pulses of varying amplitudes. The messages are passed along axons, the paths between the neurons' cell bodies, to the synapses, the junctions between neurons. Through this messaging, neurons collaborate to form networks that are capable of sophisticated information processing.[3] This network of neurons is the organic circuitry of the brain, and the neurosynaptic chips now under development mimic that design.

Conventional chips are limited by the fact that a single transistor can communicate with a maximum of four other transistors at the same time. In neurosynaptic computing, each tiny electronic element speaks to thousands of others. The chip circuitry is networked like the brain. Further, logic and memory are intermixed within a single chip, breaking the von Neumann bottleneck that we talked about in chapter 1. That allows computing systems based on the cognitive chips to store enormous amounts of information and act on it efficiently.

There's a debate in the scientific community, including within IBM, about whether it will be better to use a digital or analog approach to designing cognitive chips. Digital chips process information in ones and zeros. They compute using mathematics. But some researchers, including IBM's Mark Ritter, argue that the best approach for designing neurosynaptic chips will be to make them analog, meaning they will transmit information via electric pulses of varying amplitude, similar to the messages in the brain. In that scenario, the elements representing neurons will "remember" what happened to them by storing information about the strength and routing of the pulses they receive and transmit. At the elemental level, these mechanisms will make it possible for the machine to think. "This is how we'll make machines that truly learn—that understand and digest data like the brain does," Mark says.[4]

JOURNEY OF DISCOVERY: A CHIP MODELED ON THE BRAIN

On January 13, 2011, just days before the first broadcast of Watson on *Jeopardy!*, a box arrived at IBM Research–Almaden in San Jose, Calif. It had been mailed from the IBM chip fabrication plant in East Fishkill, N.Y. Inside the box, cushioned by Styrofoam squiggles, were 200 experimental neurosynaptic computing chips modeled on the human brain.

Dharmendra Modha, the manager of cognitive computing at the Almaden lab, and John Arthur, one of the research scientists on his team, were thrilled as they cut open the box and lifted out its precious cargo. To be sure, the chips weren't very smart compared to a human brain. Each had approximately the intelligence of a worm, in that it had a number of artificial neurons comparable with a worm's actual neurons. But they demonstrated that chips could be made that to an extent re-create the interaction of neurons and synapses in the brain through the combination of advanced mathematical algorithms and silicon circuitry.

It was a significant step toward their ultimate goal: to create a cognitive machine that can sense and draw conclusions from the things that happen in the world—not through the abstraction of numbers but more directly, as humans do via our senses. "This is not a science project. It's a quest," says Dharmendra. "It involves many people, many years, and lots of unknowns. But if we can engineer a computer to exhibit the versatility and robustness of the mind, it would have huge ramifications for society."[5]

Their cognitive chips were the first harvest of the SyNAPSE (Systems of Neuromorphic Adaptive Plastic Scalable Electronics) project, a collaboration of the Defense Advanced Research Projects Agency (DARPA), IBM, and faculty members from several universities, among them Columbia, Cornell, Stanford, and the University of

Wisconsin. It illustrates yet another key element of successful innovation in the cognitive era. Researchers have to keep an open mind. They can't stubbornly pursue a path because they're invested in it personally. They have to be open to making drastic changes.

Dharmendra grew up in Mumbai, India, and his early education came in a vocational school. But after a tutor took interest in him and gave him his first lessons in trigonometry, he came to believe that he could go far in life if he put his mind to it. He eventually got an undergraduate degree from the prestigious IIT Bombay and a Ph.D. in electrical engineering from UC San Diego, where, among other things, he studied the mathematics of neural networks, one of the foundations of cognitive science. Dharmendra landed a position as a researcher at the Almaden lab, where he specialized in writing algorithms for storage devices. When another company tried to recruit him, Dharmendra instead went to lab director Mark Dean with a radical proposition: he wanted to build a brainlike computer.[6] Dharmendra convinced Mark to permit him to organize a conference on cognitive computing in 2006 where leaders in the field—including Nobel laureate Gerald Edelman—presented their latest insights. Dharmendra came out of that event convinced that it was possible to, in a sense, reverse-engineer the brain.

At about the same time, Henry Markram and his team at the Swiss Federal Institute of Technology conducted an experiment where they simulated a slice of the brain on a supercomputer—which led a few years later to the European Union's Human Brain Project. The race was on to create a virtual brain within a computer. Dharmendra's big break came in 2008, when DARPA launched SyNAPSE. That program would provide funding for Dharmendra's project beyond what IBM would invest.

The goal was to design electronic circuitry that would be based on the workings of the brain. It was a difficult journey. Realizing that the brain couldn't be easily mapped to today's chips, which are designed and manufactured using an approach called complementary metal-oxide-semiconductor (CMOS), the team planned on inventing a new technology that would mimic the functions of neurons and synapses. They intended to use the analog approach, rather than digital, because biology is analog. But, after a time, it became clear to Dharmendra that to develop a new technology rivaling CMOS was an effort that would take too long and cost too much to meet the deadlines and budgets of the DARPA project. On December 16, 2009, he arrived home from work feeling like a defeated man. Tears welled up in his eyes as he told his wife, "SyNAPSE is toast. It's over."

Fortunately, that was not the case. A few weeks later, Dharmendra was struck with a serious case of the flu that left him feverish and in bed. He couldn't even muster the energy to answer e-mail. But being stuck at home gave him a lot of time to think over the SyNAPSE dilemma. He concluded that while it was futile in the short term to try to invent a new technology for a cognitive machine, that didn't mean the project should be abandoned. Instead, the team needed to refocus on CMOS chip technology and on digital circuitry rather than analog circuitry. They would create an entirely new non–von Neumann architecture in both silicon and software that would simulate the functions of neurons and synapses. Using that architecture, they would produce chips for sense-making tasks that would be vastly more efficient than today's standard digital processors.[7]

Dharmendra called a meeting of all of the participants in the project. On January 21, 2010, about twenty-five people crowded into the library at the Almaden lab. Others were on the phone from IBM labs in New York and from the universities. He knew that it would be difficult to convince some team members that his new approach was the way to go. To do so, he would have to get people to abandon emotional attachments to fiercely held beliefs, primarily the thought that the project could only succeed through a new technology and a new kind of analog chip. So he decided to use a brainstorming exercise called "six thinking hats." The technique requires participants in a contentious discussion to identify, using colors, the essential nature of the points they make in discussions, with six different colors standing for facts, emotions, and so on.[8]

After Dharmendra laid out his proposal an intense discussion ensued. Some of the scientists argued that they should continue with the long-term exploration even though it would not be possible to meet DARPA's deadlines. But Dharmendra made a convincing argument that they needed to adopt a strategy that could succeed within the DARPA timeframe. He argued that, in the future, armed with more knowledge, they'd be able to continue the quest for the novel technology. Also, he pointed out that by using

CMOS and digital circuitry, they would be able to test their designs thoroughly by simulating them in software using supercomputers before going to all the work of designing and manufacturing a physical chip. "Learn through simulation. Burn on the chip," he told them. People gradually came around this way of thinking. Using the "six thinking hats" technique made it easier to separate gut feelings from rational arguments. The exercise "gave us structure and clarity, and made it easier to solve things," says Bill Risk, an IBM Research physicist who is one of the project managers for SyNAPSE.[9]

Out of that meeting came agreement on the way forward. Soon afterward, on February 17, 2010, DARPA enthusiastically endorsed the new direction. Within a year, the team had designed the chips that arrived in the box from East Fishkill. Rather than processing data sequentially, as in classic von Neumann computing, the chips mimic the brain's event-driven, distributed, and parallel processing. Using those chips, team members were able to write applications, including teaching the system to play Pong, one of the earliest computer games, which demonstrated capabilities such as navigation, machine vision, and pattern recognition.[10]

In mid-2013, IBM produced a second generation of chips, with the design name TrueNorth. They possess approximately the same number of neurons as a bumblebee. If you put one hundred of the chips together in a computing system, you're got something that occupies less than two liters of volume and consumes less energy than a lightbulb. Dharmendra calls it a "brain box." The coming generations of brain boxes will be able to combine information from different sensors and draw conclusions from it, much as humans combine what they see and hear to understand what's going on around them.[11]

Dharmendra's team has developed a number of scenarios to help people understand how their technologies might be employed, using anywhere from a single chip to dozens of them. They worked with industrial designers to create physical models of potential devices:

- Assistive Vision, eyeglasses designed to help people with impaired vision. The glasses would be outfitted with video cameras that can identify obstacles, dangers, and objects the wearers might be looking for. The glasses would plot a pathway and tell the wearer how to navigate via embedded speakers or ear buds.
- Tumbleweed, a robotic ball capable of rolling through a disaster scene to find injured people. The ball, perhaps the size of a volleyball, would have numerous video cameras embedded in it, and would concentrate on the images that are most useful. The cognitive system would evaluate the video streams and other sensor outputs to find injured individuals and send messages to the authorities reporting their location via a radio transmitter.
- Conversation Flower, a device that captures both images and sounds during business meetings. The device, with petal-like elements extending from a central core, would sit in the middle of a conference room table and monitor the proceedings. The flower might be designed to give visual cues to the participants in meetings, opening if

the conversation is dynamic and useful and closing if it is boring and repetitive.[12]

The cognitive chips and Watson are complementary technologies. For the sake of simplicity, you can think of them as the right brain and left brain of the era of cognitive systems. Watson, the left brain, focuses on language and analytical thinking. The cognitive chips address senses and pattern recognition. Over the coming years, IBM scientists hope to meld the Watson and TrueNorth capabilities together to create a holistic computing intelligence.

SCENARIO: MEDICAL IMAGING ASSISTANT

Over the past century, tremendous progress has been made in technologies designed to help spot evidence of disease inside the human body. X rays, ultrasound, computed tomography, nuclear medicine, and magnetic-resonance imaging make it easier to identify and treat everything from heart disease and breast cancer to complex gastrointestinal maladies. Yet innovation has lagged when it comes to helping physicians interpret the information captured in such images. They must routinely look through hundreds or even thousands of images to find the ones that are relevant and telling. And, in the process, they may miss signs of diseases they're not explicitly looking for.

Imagine a new cognitive technology, a medical-imaging assistant, that can automate the process of examining medical images and also spot evidence of disease that physicians might miss. The technology doesn't replace radiologists; it relieves stress and eyestrain and helps make them better at what they do. Today, a radiologist typically reports for work at a hospital or clinic and plows through two dozen cases during a shift. Often, all she knows about each patient is a tentative diagnosis and instructions on what to look for in the images. Radiologists are highly skilled people who are required to do too much manual inspecting.

The medical-imaging assistant will help transform the way radiologists do their jobs. When a specialist begins considering a cardiology case, for instance, she will be presented with a digital-image storyboard created by the computer that shows the most significant twenty-five images in a series. Write-ups created by the computer will explain why each frame has been selected for close examination. These will alert the radiologist when evidence of disease is found, whether or not it is the particular disease being sought. The medical imaging assistant will have been trained by examining millions of images along with the diagnoses, treatments, and outcomes associated with each of them. It will have a huge memory packed with encyclopedic medical information, medical guidelines, and rules established by the hospital or medical group.

Even more importantly, the system will make all of the health-care information about the particular patient available to the radiologist, including his family history and entire personal medical record. The assistant will offer up annotated summaries of the results of other tests. The radiologist will be able to carry on a spoken dialogue with the system to delve deeper into the particulars of each case. In this way, the radiologist gets a full picture of the patient and his problems.

The automation inherent in this system could make health care much more efficient. A coronary angiogram is typically made up of 25 short video sequences, contains about 5,000 images, and takes a radiologist about 30 minutes to examine. Using the medical-imaging assistant, the time spent looking at images could be reduced to perhaps five minutes. The impact on patient outcomes could be even more significant. A radiologist supplied with all of the contextual information and given the ability to explore the case through a conversation with the cognitive system is likely to be more effective at providing sage advice to treating physicians.