10× — human-machine symbiosis

D Roy

10× is a new initiative at the MIT Media Laboratory with the goal of magnifying human abilities by an order of magnitude ('10×'), or more, along various cognitive and physical dimensions. Based on an understanding of human abilities and limitations, technologies can be shaped to extend human reach. By setting our sights on 10× gains, we deliberatively seek human abilities which can be dramatically improved with appropriate technology-based catalysts. We view the underlying challenges of developing assistive technologies for individuals with impairments and bionic technologies for unimpaired individuals as fundamentally similar in nature. Across this spectrum, our aim is to develop technologies that complement rather than replace human abilities.

1. Introduction

Consider our ability to visually search. There are many mysteries regarding the mechanisms underlying human vision and memory, but we can nonetheless make some simple observation about how we search. When searching for our keys, why do we often look repeatedly in the same places? We might be far more efficient in search if we could keep ourselves from looping to places we have already looked [1]. Imagine a lightweight device, perhaps woven into a pair of eyeglasses, that keeps track of where people have looked and steers them away from looking there again. This search aid would not need to know where the target object is — it would simply augment the person's visual search routine with a form of extended memory. With the appropriate interface for steering ongoing search, such a technology may lead to human-machine superperformance — more efficient search than either the human or the machine could have achieved on their own. This is just one of a wide range of technologies under development at the Media Lab that have the potential to '10×' human performance.

In this overview, I provide some historical context for the notion of human-machine symbiosis, identify a few basic research areas that are most relevant to the advancement of the 10× agenda, and finally sketch some of the active 10X research themes at the MIT Media Laboratory.

2. Licklider's vision — human-computer symbiosis

Human-machine systems have a long history at MIT. In 1960, J C R Licklider formulated a vision of human-computer symbiosis in which computers and humans would become fluidly interdependent, each providing complementary abilities towards some shared goal that neither could achieve alone. Licklider considered the mismatch between human thought and computer abilities along several dimensions — flexible versus logical analysis, speed, memory size and organisation. Based on the relative abilities of humans and computers, he suggested symbiotic roles that computers might play to enhance the abilities of humans [2]:

> *'Men will set the goals, formulate the hypotheses, determine the criteria, and perform the evaluations. Computing machines will do the routinizable work that must be done to prepare the way for insights and decisions in technical and scientific thinking. Preliminary analyses indicate that the symbiotic partnership will perform intellectual operations much more effectively than man alone can perform them.'*

Licklider predicted that computers would take over routinisable tasks, leaving the setting of goals and creative planning to humans. In significant ways, this vision has become a widespread reality. Computers have seeped into virtually every aspect of our everyday lives. In some forms the changes are clearly visible, such as Internet search engines that we turn to on an increasing basis to tap into a vast and growing collection of knowledge that respects no boundaries of space and time. In other forms, the changes are equally powerful yet invisible. Many are unaware, for example, of the complex network of computers that regulate the operations of our cars as we drive.

What is the appropriate metaphor for the role of technology in a symbiotic relationship? We might view technologies as extensions of humans (e.g. eye glasses), as tools (e.g. calculators), or as partners (e.g. autonomous robots). Each metaphor suggests a different level of autonomy on the part of the machine. We might want human extensions to have minimal autonomy, tools somewhere in between, and partners to be relatively self-motivated (but with shared goals). At the Media Lab there has been a tendency to focus on the extremes, by developing seamless technologies which augment humans and agents which autonomously act in collaboration with humans.

3. Beyond computers

In 1960, the year that Licklider formulated his conception of human-computer symbiosis, information technology was in its infancy. The RS-232 serial interface and the COBOL programming language were developed that year. The mouse would not be invented for another three years. The nascent field of artificial intelligence was firmly rooted in the symbol processing view of intelligence which was to guide the field for the next two decades. Within this context, Licklider highlighted [3] aspects of computing technology that needed advancement to support his vision of symbiosis:

> *'Prerequisites for the achievement of the effective, cooperative association include developments in computer time sharing, in memory components, in memory organization, in programming languages, and in input and output equipment.'*

It is interesting to ask which, if any, of these issues remain a bottle-neck to human-computer symbiosis in our current technological climate. The first two issues, speed of computers, and the speed and size of memory, no longer seem to be critical limiting factors. In comparison, memory organisation, languages for programming and communicating with computers, and human-computer interfaces all remain bottle-necks.

Licklider's vision of human-computer symbiosis was built upon certain assumptions about the nature of computers which are largely held to be as true today as they were in 1960:

- computers exist in a perceptual vacuum — they are deaf and blind,
- computers cannot act upon their physical environment — they are disembodied,
- computers 'think' by processing symbolic descriptions,
- computers know what to do based on what they are programmed to do.

Researchers at the Media Lab have challenged each of these assumptions, leading to a new view on what symbiotic technologies might be.

3.1 Perceptual computing — machines with a sense of the world

Computers today are largely unaware of their immediate physical environment. The only way for a person to directly influence the behaviour of a system is through keyboard and mouse. Most of the rich channels of interaction that we rely on to interact with others in social settings including voice and body language are lost upon computers. Perceptual computing is the development of signal processing and pattern analysis techniques for sensing and interpreting the environment. A particular emphasis has been placed at the Media Lab on interpreting the presence, identity, and activities of people using a wide variety of sensing modalities including touch, vision, audition, and biometrics [4—7].

3.2 Natural embodiment — human-friendly mechanical systems

Just as computers are unaware of their environment, they are also unable to act upon their immediate environment other than through conventional CRT/LCD monitors and audio output. Touch and physical action, traditionally the domain of robotics, may become an integral part of future human-machine systems.

intimate computer interfaces that can be worn all the time

Conventional robots are inflexible, heavy, and dangerous and thus inappropriate for human interaction. Researchers at the Media Lab are experimenting with compliant actuated systems, leading to new forms of touch-based interfaces [8], interactive robots [9, 10], and robotic exoskeletons [11]. Based on the metaphor of computers as clothing, the Media Lab has also developed a series of wearable computer prototypes (see, for example, Starner et al [12] or Picard and Healey [13]). These experiments explore the idea of intimate computer interfaces that can be worn all the time, that see what the user sees, knows where the user goes, and is always available no matter where the user is, serving as a persistent extension of a person's body.

3.3 Natural representation — giving symbols meaning

In artificial intelligence and cognitive science, intelligence is often equated with rational, logical processing of knowledge encoded in symbolic form. Recent psychological studies have shown, however, that affect and embodiment (the fact that our brains evolved to control physical bodies) shape virtually all aspects of cognition. The Media Lab is developing new frameworks for computational intelligence in which affect (emotional state) and grounding (connections to the physical) are taken seriously [14—16]. These efforts provide new perspectives on how machines may complement human performance, being sensitive to the role of emotion, goals, and physical situations.

3.4 Learning and expression — beyond programming

The most common way to instruct a computer today is to program it. With growing complexity of systems, providing explicit instructions is becoming increasingly unpractical. This situation increasingly motivates the need for machines that learn from humans through natural interactions such as by example or through guidance [17—19]. To complement learning systems, we must devise better ways to tell machines what to do, and for them to communicate back to us. The Media Lab pioneered a vision of multimodal human-machine interaction in which speech, gesture, and other modalities are seamlessly interwoven in the course of communication [20]. To advance fundamental understanding of communication along these lines, we must develop new ways of modelling semantics and social interaction. Researchers at the Media Lab are exploring new possibilities by cross-fertilising many of the themes mentioned earlier, including perceptual computing and natural representation.

4. 10×

The numerous challenges to the traditional view of what a computer is leads me to suggest that we are in a position to push to new heights of human-machine symbiosis. Moving beyond the conception of computers of the 1960s which motivated Licklider, we are now able to explore collaborative human-machine systems in which machines have become perceptually engaged, physically embodied, representation-grounded, learning systems. Within this expanded framework, we are designing symbiotic systems that magnify human performance along various cognitive and physical dimensions. Although it is beyond the scope of this introductory paper to survey all relevant strands of research related to 10× at the Media Lab, I have attempted to highlight some major directions of research and cite examples of work related to each.

we are designing symbiotic systems that magnify human performance

4.1 10× memory

A clear complementarity exists between human and digital memory. Computers can store practically unlimited quantities of information with highly accurate recall. In contrast, human memory is not as reliable, but has powerful modes of access (e.g. analogical recall) that are not well understood. Seamless interfaces that facilitate access to large stores of memory promise to expand cognitive abilities in fundamental ways (see, for example, Rhodes [21], DeVaul, Pentland, and Corey [22], and Vemuri and Bender [23]).

4.2 10× expression

Human expression is achieved through the translation of intentions into physical action. The expressive process may be augmented by designing technologies that infer and translate intentions naturally, and extend the range of physical actions that a person can use as forms of expression — for example in the domain of music, see Machover [24, 25], Paradiso et al [26], and Farbood, Pasztor, and Jennings [27], and Maeda [28] describes a computational language for design, while Vercoe's Csound is a language for expressing music [29].

4.3 10× listening

We have access to increasing amounts of information as audio, either streaming from the Internet or recorded digitally. Computers can create synthetic listening experiences which allow for more effective listening. For example, we can leverage the 'cocktail party effect', our ability to selectively attend to a single sound source, to create audio browsing environments in which sound sources are selectively amplified on the basis of the listener's head position (see, for example, Schmandt and Mullins [30]). Automatic acoustic analysis of speech combined with search and browse interfaces can enable nonlinear access to speech recordings, similar to the way we can skim and search visually (see, for example, Arons [31]).

4.4 10× learning and understanding

Computers can bring models to life, enabling people to see and understand situations in new ways. To learn in this way, people must convey their mental models to computers and be able to explore the implications of those models when put into action, leading perhaps to revisions of their mental models. Efforts at the Media Lab take this approach through the development of technologies that support constructionist styles of learning [32, 33].

4.5 10× physical skills

Mechanical systems with finely tuned control promise to extend our physical abilities. Bionic technologies may range the full spectrum of symbiotic relationships, from extensions of the body to autonomous partners. Robots may be reconceived as 'skins' that supercharge bodily movements (see, for example, Herr and Langman [11]), or as partners that complement an individual's reach through social and conversational interaction (see, for example, Breazeal et al [34] or Roy et al [10]).

4.6 10× awareness

Awareness of one's immediate environment, that is, a sense of the 'here-and-now' is essential to everyday life. With a combination of sensor networks and wearable computers, it is possible to extend human awareness to include events that are not in a person's immediate physical environment, and to include events that are not detectable by the unaided human senses [12, 35, 36].

5. Conclusions

The development of human-machine symbiosis must be grounded in an understanding of both human and machine elements. In many ways the situation is akin to architectural design in which the architect must be attuned to both human nature and the technology of building.

It is thus not entirely surprising that the Media Lab, situated within a school of architecture, provides a test bed for mixing various aspects of art, science, design, and technology to push the boundaries of human-machine symbiosis.

Acknowledgements

This paper is a synthesis of ideas that have been developed by many of my colleagues at the Media Laboratory. Pattie Maes, Rosalind Picard, and Chris Schmandt made specific suggestions which helped shape this overview.

References

1 Roy D, Ghitza Y, Bartelma J and Kehoe C: 'Visual memory augmentation: using eye gaze as an attention filter', To appear in the Eighth IEEE International Symposium on Wearable Computers (2004).

2 Licklider J C R: 'Man-computer symbiosis', IRE Transactions on Human Factors in Electronics, HFE-1, pp 4—11 (1960).

3 Licklider J C R: 'The computer as a communication device', Science and Technology (1968).

4 Pentland A: 'Smart rooms', Scientific American, 274, No 4, pp 68—76 (April 1996).

5 Picard R: 'Affective computing', The MIT Press (1997).

6 Vercoe B L: 'Computational auditory pathways to music understanding', in Deliège I and Sloboda J (Eds): 'Perception and cognition of music', London, Psychology Press, pp 307—326 (1997).

7 Paradiso J A: 'Tracking contact and free gesture across large interactive surfaces', Communications of the ACM, 46, No 7, pp 62—68 (2003).

8 Ullmer B and Ishii H: 'Emerging frameworks for tangible user interfaces', IBM Systems Journal, 393, No 3, pp 915—931 (2000).

9 Breazeal C: 'Designing sociable robots', The MIT Press (2002).

10 Roy D, Hsiao K and Mavridis N: 'Mental imagery for a conversational robot', IEEE Transactions on Systems, Man, and Cybernetics, Part B, 34, Issue 3, pp 1374—1383 (2004).

11 Herr H and Langman N: 'Optimization of human-powered elastic mechanisms for endurance amplification', Journal of the International Society of Structural and Multidisciplinary Optimization (ISSMO), 13, pp 65—67 (1997).

12 Starner T, Mann S, Rhodes B, Levine J, Healey J, Kirsch D, Picard R and Pentland A: 'Augmented reality through wearable computing', Presence, Special Issue on Augmented Reality, 6, No 4 (1997).

13 Picard R and Healey J: 'Affective wearables', Personal Technologies, 1, No 4, pp 231—240 (1996).

14 Minsky M: 'The Emotion Machine', (forthcoming) — http://web.media.mit.edu/~minsky/

15 Picard R: 'Affective Computing', The MIT Press (1997).

16 Roy D: 'Grounding language in the world: signs, schemas, and meaning', (forthcoming) — http://web.media.mit.edu/~dkroy/

17 Maes P: 'Intelligent software', Scientific American, 273, No 3, pp 84—86 (1995).

18 Lieberman H (Ed): 'Your Wish is My Command: Programming by Example', Morgan Kauffman, San Francisco (2001).

19 Ivanov Y and Blumberg B: 'Developmental learning of memory-based perceptual models', Second International Conference on Learning and Development (2002).

20 Bolt R: 'Put-that-there: voice and gesture at the graphics interface', International Conference on Computer Graphics and Interactive Techniques (1980).

21 Rhodes B: 'Using physical context for just-in-time information retrieval', IEEE Transactions on Computers, 52, No 8, pp 1011—1014 (August 2003).

22 DeVaul R, Pentland A and Corey V: 'The memory glasses: subliminal versus overt memory support with imperfect information', 7th IEEE International Symposium on Wearable Computers (2003).

23 Vemuri S and Bender W: 'Next-generation personal memory aids', BT Technol J, 22, No 4, pp ...... (October 2004).

24 Machover T: 'Hyperinstruments: a composer's approach to the evolution of intelligent musical instruments', CyberArts, pp 67—76, Miller Freeman (1992).

25 Machover T: 'Shaping minds musically', BT Technol J, 22, No 4, pp ...... (October 2004).

26 Paradiso J A, Pardue L S, Hsiao K and Benbasat A Y: 'Electromagnetic tagging for electronic music interfaces', Journal of New Music Research, 32, No 4 (2003).

27 Farbood M, Pasztor E and Jennings K: 'Hyperscore: a graphical sketchpad for novice composers', IEEE Computer Graphics and Applications (January—March 2004).

28 Maeda J: 'Design by Numbers', The MIT Press (1999).

29 Boulanger R (Ed): 'The Csound book: perspectives in software synthesis, sound design, signal processing, and programming', The MIT Press (1999).

30 Schmandt C and Mullins A: 'Audio streamer: exploiting simultaneity for listening', CHI 1995, pp 218—219 (1995).

31 Arons B: 'SpeechSkimmer: a system for interactively skimming recorded speech', ACM Transactions on Computer Human Interaction, pp 3—38 (1997).

32 Papert S: 'Mindstorms: Children, Computers, and Powerful Ideas', New York, Basic Books (1980).

33 Resnick M: 'Rethinking learning in the digital age', in Kirkman G (Ed): 'The Global Information Technology Report: Readiness for the Networked World', Oxford University Press (2002).

34 Breazeal C, Brooks A, Gray J, Hoffman G, Kidd C, Lee H, Lieberman J, Lockerd A and Mulanda D: 'Humanoid robots as cooperative partners for people', (forthcoming).

35 Marmasse N, Schmandt C and Spectre D: 'WatchMe: Communication and awareness between members of a closely-knit group', to appear in UbiComp 2004, The Sixth International Conference on Ubiquitous Computing, Nottingham, England (2004).

36 Barton J, Kelaney K, O'Mathuna C and Paradiso J A: 'Miniature modular wireless sensor networks', in Holmquist L E and Ljungstrand P (Eds): 'UbiComp 2002 Adjunct Proceedings', Viktoria Institute Press (Gothenburg Sweden), pp 25—26 (2002).

Deb Roy heads the Media Lab's Cognitive Machines group, which develops machines that learn to communicate in human-like ways. This work combines elements of spoken language understanding and generation, machine learning, computational linguistics, computer vision, robotics, and developmental psychology. Other active research interests include multilingual technologies, multimedia search and retrieval algorithms, and interactive robotics. He holds a Bachelor of Applied Science in computer engineering from the University of Waterloo, Canada, and an MS and PhD from MIT's Program in Media Arts and Sciences.