The Evolution of Human to Whatever Interface

An IEEE Digital Reality White Paper

May 2020

Roberto Saracco, Co-chair, IEEE Digital Reality Initiative

DigitalReality.ieee.org

Contents

Introduction

A journalist working on a piece on the future of human-computer interfaces (HCI) recently posed the following four questions:

- What HCI technology is being developed and by what companies?
- What form will it take and will the future of the human-computer interface be intrusive or subtle?
- Can we trust companies like Facebook and Google to not use the output of such an interface to 'other' uses?
- How will the interface between human and computer be protected, and how will privacy be assured?

Figure 1 Human Machine interfaces have seen an expansion of technology support but have yet to move in the space of holistic interfacing. Image credit: Continental – Future of Motion, Holistic Human-Machine Interface

Human machine interfaces originated to direct a machine to perform a given task, like using a key to operate an engine or flip a switch to start the air conditioning. As machines became more complex they required more variety of commands but at the same time they could also accept more powerful commands, mostly in textual form. The point is that human machine interaction was designed, out of necessity, with the machine in mind. It did not mirror the way humans interact with one another. This situation has not really changed in the last fifty years, but it is starting to change now thanks to software and artificial intelligence.

Voice Interaction

Voice recognition has made impressive progress in the last few years so much, in fact, that some companies have turned to voice interaction. We can now talk to televisions and cars. The problem is that we are so used to typing (with a remote or a selection knob) that it feels awkward talking to a machine. Also, and this is probably the source of uneasiness, we need to learn how to talk to the television (and to the car). Basically these new voice interactions are a transposition of written command into voice, not necessarily in the same fashion as in written text. Besides, the interaction mimics those of robots talking in science fiction movies in the last century.

What is missing today, in spoken interaction, is the ability for the machine to understand a convoluted sentence and contextualize it to derive a meaning. This requires a higher level of intelligence to learn the possible meaning of the request. The required technology is already available, and there are several demonstrations of fluent human machine voice interaction, but this technology is not yet affordable for mass market use, e.g., in televisions or cars. It is expected, however, that in the coming years voice interaction will displace keyboards, relegating them to technology museums.

Notice how youngsters are starting to use voice messages instead of written messages, because they are more convenient. Voice messaging has little to do, basically, with human-computer interaction, but the spreading of voice messages is fostering the adoption of voice interaction with machines. Voice messages are different from a face to face voice interaction; the "interaction" part is lost, and it is a unidirectional stream of communication.

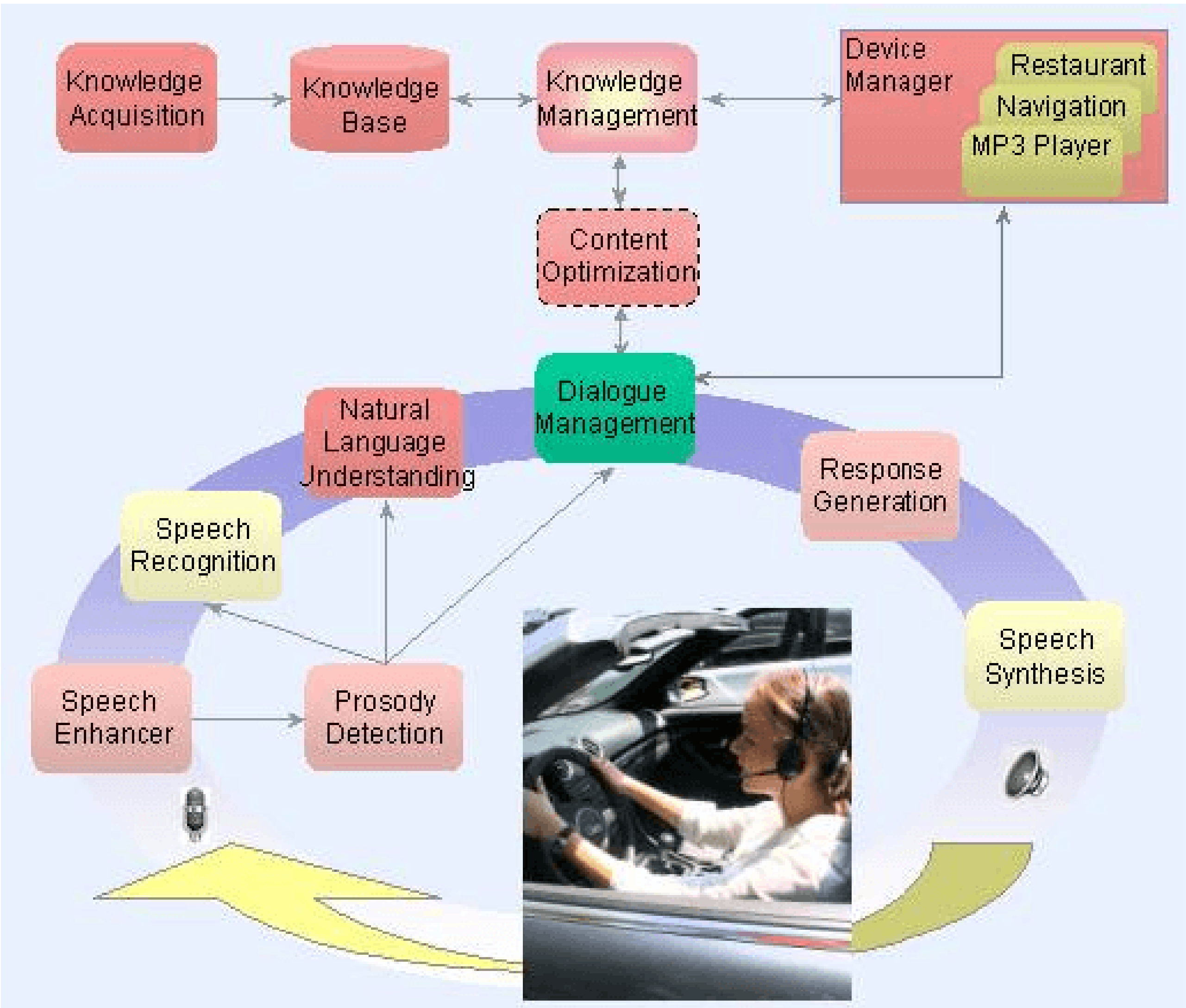

Figure 2 To have a meaningful interaction the machine has to have a context and a knowledge base. This has been recognised long time ago, this schematic is from 2006. It remains a challenge as of today but we are getting closer. Image credit: Heather Pon-Barry and Fuliang Wen

Voice interaction is considerably different from interactions via typing. One difference is the delocalization of the computer/machine. The person does not need to be physically connected (using fingers to type); the voice just needs to be heard. This has some interesting perceptual implications.

In spite of the image depicted in Figure 1, people do not perceive Alexa, Siri or whatever as embodied in the device being spoken to. These entities are ubiquitous in the ambient you are in; you are talking to a presence in the ambient, and actually the ambient has become responsive. A significant improvement in the interaction would be obtained if the computer becomes aware of the context; it doesn't just listen to what you say after the wake up word (Alexa, Hey Google…) but keeps listening to you and knows what you said (or others present in the ambient said) in the last 30 minutes or more. Companies supporting these voice interaction systems indicate that due to privacy concerns, the devices are only waiting for the wake up word and do not listen to anything you said before or after the command has been processed. However, would people be willing to trade part of their privacy for a much better interaction? Couldn't we trust the assurance from those companies that what the device (software) is learning about me and the ambient situation will only be used to ensure a better interaction and nothing else? Well if I have decided to trust them when they claim the software is only listening to what I am saying immediately after the wake up word, why shouldn't I trust them when they will be saying the continuous listening is used only to improve the interaction?

Figure 3 Alexa, Siri, Cortana, Google … they are all listening to what you say. Image credit: Shelly Palmer blog

Notice that today Alexa and its siblings interact with and control a limited number of appliances in the home but their number and diversity is going to increase in the future. Alexa will be used as the voice of the home, to make sure there is no leaking faucet, the cat returned home, that the maid did wash the curtains, as so on. Eventually, we will be talking directly to the ambient, forgetting we are doing so through an intermediary.

This will be a significant departure from the way we perceive the interaction with a computer today; interaction will be with the ambient and we are going to feel like our ambient is aware and responsive. Note the use of the general term ambient - it could refer to the home, the office, the car, a shopping mall or a hospital. The future of interface will be shaped around ourselves. The title of this white paper identifies the point; it is our human interface to whatever, not to a machine. And it will be tied to us, not to a specific machine, device, computer, exactly as today I am interfacing with other people using my interface.

Voice interaction technology, such as automatic speech recognition (ASR) and natural language understanding (NLU), will be instrumental in this transition affecting our perception. The more fluent the understanding is, and the more articulated the answers we receive are, the more convincing the interaction appears as a human like being. Technologies like sentiment analyses and affective computing will shift the perception towards a sentient being rather than a machine, getting really close to human to human interaction.

Touch Interaction

Clearly there is much more involved in human to computer interaction than voice based. Consider the sense of touch. Normally we don't use touch to communicate but there are several exceptions, like in dancing where the man steers the woman by using touch signals. This is particularly so with tango but it also goes with other dances. Sometimes we use touch to communicate emotions, like the various ways in which you can touch a hand or shoulder.

Additionally, touch is an important sense in providing our brain with a sense of reality; actually it is so important that when it is missing we immediately perceive that something is fake. This is evident when using virtual reality (VR) interfaces. These interfaces are not providing touch sensations, and this is one of the main reasons why our brain feel VR as … virtual and not real. Sony has just hinted that the new PlayStation 5, to hit the market at the end of 2020, will include an advanced haptic interface to make VR much more "real".

Technologies for haptic interfaces have progressed in the last ten years. We are seeing them in several gaming devices and in professional equipment, like in robotic surgery control. Haptic interfaces can be clustered into graspable, wearable and touchable. Graspable interfaces are based on bars or sticks that change their resistance through a motor counteracting your forces (that is also why some of these are advertised as "force feedback" like a joystick). The evolution has been in accuracy, sensitivity and in number of directions (also known as degrees of freedom). The very best graspable haptic can provide 7 degrees of freedom. To put this into perspective our human hand has 27 degrees of freedom; hence we can experience many more subtle sensations.

Figure 4 What does tango have to do with technology evolution? Not much really. It is just an example of the way we use touch interfaces to communicate, the (male) dancer by pressing his hands in different ways on the back of his partner communicates the intention for the next steps. Image credit: City Academy

Wearable haptics, like haptic gloves, are particularly useful in a VR context since they can provide a touch sensation in thin air. The problem with these technologies is their bulkiness that by itself sends a message of "fake" (or artificial) to our brain. It is an area where significant progress has been made but where the cost is still high (probably not affordable to the mass market).

Touchable haptics, like the Apple 3D touch interface, are based on a vibrating surface. Depending on the vibration frequency it can recreate specific touch sensation. Apple introduced it back in 2015 in their iPhones but it is not present in their most modern ones, like XR, iPhone 11 and iPhone Pro. (Apple claims to have replaced the 3D touch -that was a real haptic interface- with what they call a haptic interface -that is NOT a haptic interface since it just determines the length of time you keep your fingertips on the screen, it does not return any touch sensation.)

Augmented Reality (AR) and VR would greatly benefit from seamless haptics, and in the next decade further evolution in technology will likely bring this to our fingertips. By the way, this is an area where very low latency is needed, hence where 5G could make a difference.

Brain Computer Interfaces

The previous sections addressed seeing, hearing and touching applied to the human to whatever interface. There are three more, smell, taste and proprioceptors (these latter provide a sense of position and acceleration), but these will be skipped for now since they are not used, nor are likely to be used in isolation as interface, rather they may become part of a multisensory, multimodal interface, and therefore they will be considered in that context.

So now consider interfaces that do not engage our senses. There is a lot of work, and hope, to be able to interface our brain directly to the world, beginning with a connection to a computer. The IEEE Symbiotic Autonomous Systems Initiative, now merged inside the IEEE Digital Reality Initiative, addressed a 30-year horizon, and it is expected that ongoing research will deliver a brain computer interface in this timeframe. The technology is unlikely to fulfill the dream of a seamless Brain to Computer Interface (BCI), but non-seamless interfaces should be available in the coming decades. Note that BCI is available that allows a paraplegic person to communicate with an exoskeleton and walk again or allows a person to control a robotic arm with her thoughts to drink from a glass. These results are amazing but are not about an interface providing signals to a computer that can then decode them and "read the person's mind".

This existing interface needs to be completely understood in order to gauge the progress in BCI and the road ahead. In this case, electrical signals generated by the brain are captured by electrodes and are sent to a computer. The computer generates a visual rendering of these signals, and the person is trained to think in such a way that eventually the computer will understand his intention. Most of the learning is performed by the humans. The adoption of machine learning and better signal processing is now decreasing the time the human needs to learn how to interact with the computer. Also notice that there is a very strong tie between that person and the computer. The same computer that understood one person's intention will be at loss in trying to understand another person. The reason is clear. Most of the understanding is on the human side.

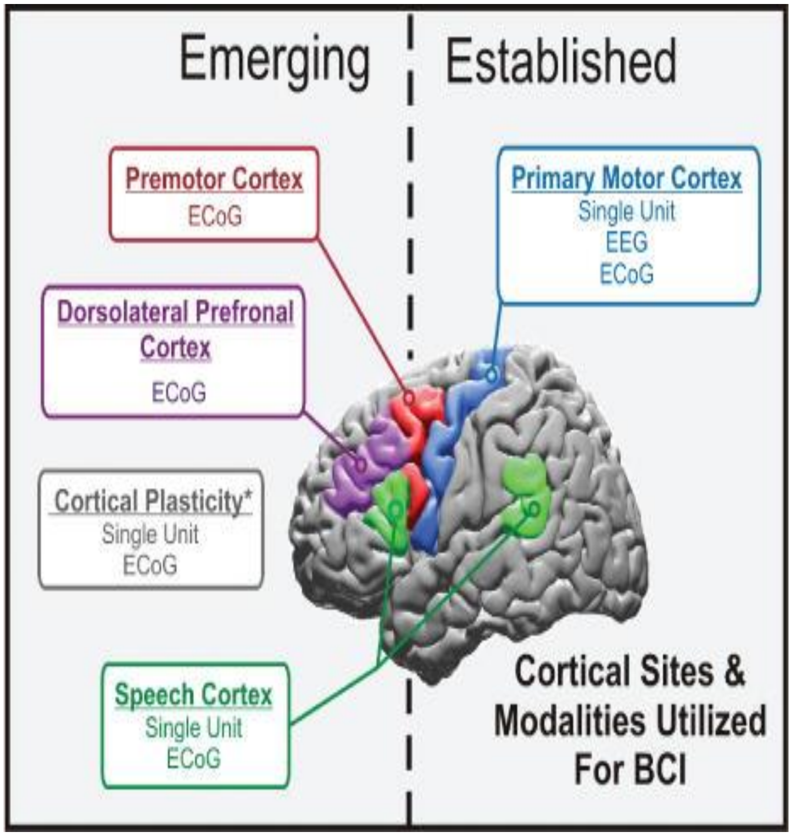

Figure 5 Interfacing the brain motor cortex with a computer is now an established procedure (although quite far from being common). New areas of the brain are being considered and studied as shown in the figure. Image credit: Eric Leuthardt et al, Journal of Neurosurgery

The evolution in BCI is occurring in several directions:

- less invasive physical interface and better electrodes to pick up the signals
- extending the brain area where signals are captured (see Figure 5)
- decreasing the training time for the person and improving computer sensitivity (intelligence)
- reverse communications, from the computer to the brain (currently still mostly science fiction).

The following addressed each of these evolutions.

Less Invasive Physical Interfaces

The challenge facing researchers and medical doctors today when looking for interfacing to the brain is the tradeoff between invasive procedures and implants versus lower resolution and increased noise. If an electrode is embedded in the brain, that electrode will be able to pick up electrical activity in the vicinity of its sensors since today's technology allows for having multiple sensors on a single electrode, hence providing high accuracy and very low noise. Current technology can be as selective as pinpointing electrical activity for as few as some hundred neurons. If one uses optogenetics the precision can be as good as a single neuron (however the problem becomes selecting the relevant neuron).

Technology evolution for implantable electrodes is moving towards providing smaller and smaller electrodes to minimize damaging the neurons. As shown in Figure 6, researchers are studying the use of nano-wires (nano-filaments of gold covered with zinc) as well as creating networks of these nano-wires to sense thousands of points in parallel. Recent studies (2019) have resulted in methods for producing these nano wires that can penetrate a cell, a neuron, without damaging it.

Implanting an electrode in the brain causes damage to the neurons; the brain is highly redundant so limited damage to neurons is tolerable but of course not ideal. To avoid any damage to neurons a brain interface option would be laying the electrodes on the cortical surface of the brain. This is considerably less precise since the electrode will capture signals generated by millions of neurons. However, coupling this electrical "mess" with artificial intelligence, as will be described later, may allow extraction of meaning and elimination of noise. Notice that choosing the electrical signal generated by a single neuron may not be significant at all, since the emerging functionality depends on the involvement of thousands, often millions of neurons. In a way it is like monitoring each single cell in the muscles of a hand to understand what the hand is doing. It is much easier, and more effective, to look at the hand as a whole.

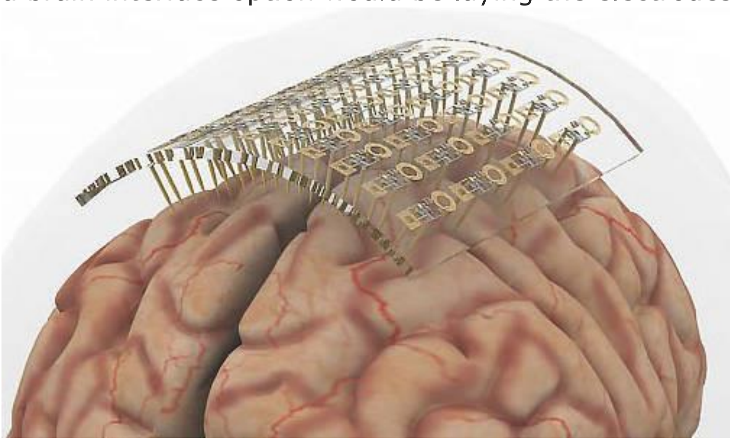

Figure 6 Graphical rendering of implantable brain electrodes based on nanowires. Image credit: Daegu Gyeongbuk Institute of Science & Technology

Also, it should be noted that even if we could manage to place an electrode by a neural circuit that corresponds to a specific activity of the brain, in a little while the brain will change and that location may no longer be the one involved with that specific activity.

So, using patterns captured at cortical level may not be bad after all. Still, capturing these patterns requires an invasive (surgical) procedure and should only be performed if it is absolutely necessary. Presumably, no one would choose to have brain surgery just to avoid typing on a keyboard (and that assumes, which is currently false, that by connecting your cortical electrical patterns to a computer the computer could understand what is going on).

A third level of interfacing is by placing electrodes on your skull. These electrodes can be embedded in a sort of cap and can communicate the detected electrical activity to a computer for signal processing and meaning extraction. There are a number of caps, like the one in Figure 7, that are embedding electrodes. The more electrodes they have and the more sensitive they are the more electrical patterns can be harvested.

These caps are still quite cumbersome to wear (but at least they do not require surgery to be used) and usually connect with the computer using a comet of wires. The evolution is moving towards making these caps easier to wear and wireless. Their price can vary from a few hundred dollars for caps used in video games, providing very low resolution and limited area coverage, to several thousand dollars for professional use. Notice that the more patterns the cap is able to detect; the more processing is required to make sense of it.

Figure 7 A skull cap embedding electrodes for EEG and brain monitoring. It is used in medical profession and research. A cap like this would cost over $10,000. Image credit: Shimadzu

This is a crucial aspect that permeates the whole brain-computer interfaces: making sense of the data being harvested. Notice that these sensors are picking up the electrical activity generated by the brain, which is the result of thousands of concurrent processes, and usually only one is interesting for the application of the interface. Sorting it out remains a major challenge. Besides, the brain activity is also modulated by chemical substances produced by the brain, and we currently do not have a way to detect these in real time, nor to gauge their effect. The electrical activity may be compared to Plato's cave, where we only see shadows of what is going on and by looking at those shadows we try to discover the reality.

As will be described later, the use of AI is now providing researchers with an important tool to create this mapping, shadow (electrical activity) to reality (what is the brain doing). So far, the best results have been obtained by monitoring the motor cortex because the electrical activity generated in that area has a closer correspondence to actions (signals that will reach the muscles causing their contraction or release). Current there is no way to place an electrode to detect when a person thinks she would like to eat a cookie.

Extending the Brain Area Where Signals are Captured

Although we like to identify specific areas of the brain and associate them to specific function, like speech, hearing, reasoning, emotion, the brain structure is massively diffused. There can be a prevalent area where a specific function is processed (emerge) like the Broca area in the frontal lobe of the dominant hemisphere where speech is formed, but that area is also contributing to other functionality and there are many other areas that are involved in speech. Hence, the possibility of capturing activities from a wider portion of the brain would be important. Two issues (among others):

- capturing electrical activity from a large area by inserting electrodes is not a viable option since many neurons will be damaged in the process, hence the need to use surface contact electrodes, like the ones shown Figure 8;
- the broader the area, the more confusing signals are obtained, hence the need for a much more sophisticated processing

The evolution is moving towards the use of a growing number of contact electrodes associated with machine learning and other AI techniques to make sense of the signals harvested identifying meaningful patterns.

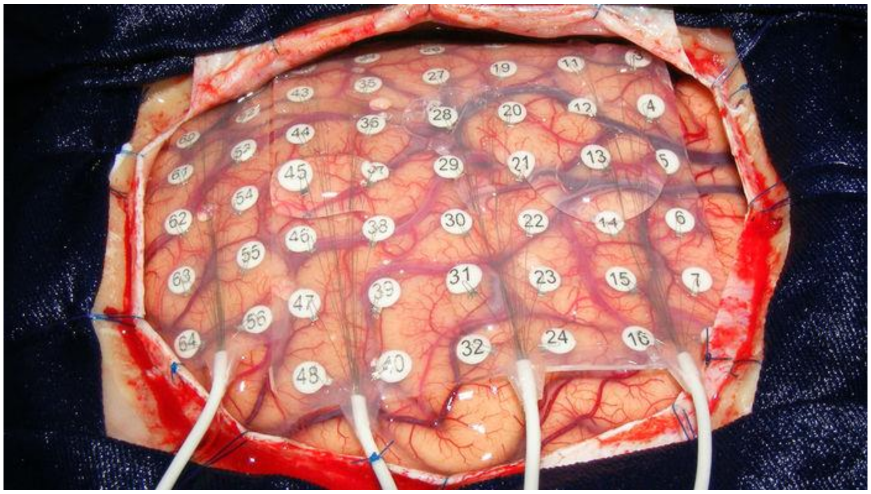

Figure 8 Extensive mesh of sensors to pick up electrical patterns in an epilepsy patient. Image credit: Wenht/Istock.com

One example is the one shown in Figure 8. Here doctors wanted to understand where the points generating an abnormal electrical activity were in an epilepsy patient. By pinpointing those seizure foci, it is possible to insert a few electrodes (normally there are very few areas that are generating the first wave of abnormal electrical activity) connected to an implanted chip to detect the insurgence of the anomaly and counteract it with an electrical spike to block the wave.

To detect these foci, the surgeon layered a mesh of sensors over the cortex. Researchers took advantage of this, with the cooperation of the patient who is fully awake during the procedure, by asking the patient to read several sentences. As the patient reads the sentences the electrical activity detected by the mesh sensors was processed by AI software that used machine learning technologies to try to identify patterns that would correspond to the words being read (the software at this stage was aware of the sentences being read and the time at which each single word was read). After several training sessions the patient was asked to read other sentences, and at this point the AI software was asked to guess the words being read. The software was developed by researchers at the University of California in San Francisco.

Similar experiments were performed in other research centers sometimes with patients reading the sentences silently (hence not affecting the motor area of the brain). The results so far have been disappointing in the sense of detection of the sentences read. However, some progress has been made. It should be noted that the software training has been necessarily limited since the patient has an opened skull. Usually, the training of AI software requires a much longer period of time.

If it were possible to use electrodes on the skull, rather than on the cortex, it would clearly be possible to have much more extended, and repeated sessions through which AI software could be trained. However, as noted previously, placing electrodes on the skull, with the present technologies, creates much more confusing signals than those that can be harvested by placing the electrodes directly on the cortex. For an easier kind of detection, like the intention to move a pointer on a screen, contact electrodes on the skull are already working (here again extensive training is required, mostly of the person).

Decreasing the Training Time and Improving Computer Sensitivity

As noted previously, current technology is not sufficient to translate the brain electrical activity, captured by sensors, into meaning. Significant progress has been made, thanks to better signal processing capability and more recently by applying AI but a "plug-and-play" interface is still in the future. The quality of the signals is also critical as well as the purpose of detection. As an example it is relatively easy to indicate the direction (up/down, left/right) you want to move a cursor on a screen through an interface between your brain and a computer but it is impossible to communicate the idea that you like the Autumn foliage to the computer.

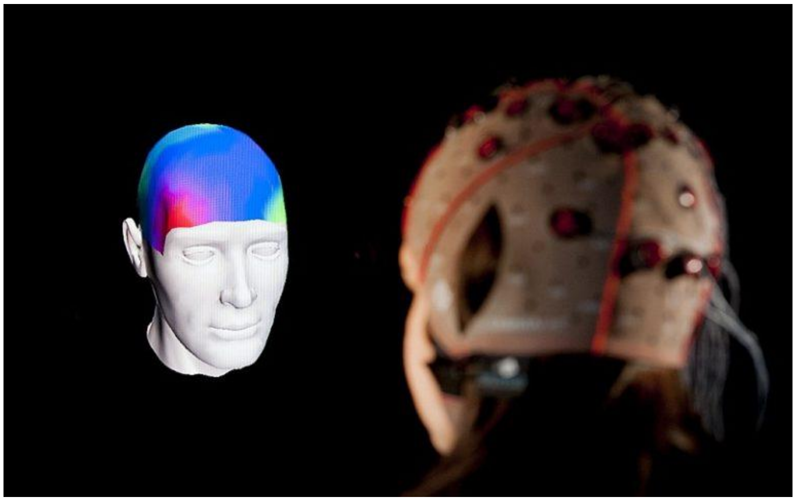

Figure 9 Using visualization techniques to increase the ability of a person in generating specific EEG patterns that can be interpreted by a computer. Image credit: Potioc

In the case of moving a cursor, the sensors embedded in a cap on your head (even the standard ones used in an EEG) can capture your brain electrical activity, particularly derived from your motor cortex. It only takes a short time (in the order of a few hours or less, it depends on the person) to learn how to generate specific patterns that the computer can interpret correctly. Usually you are told to think about moving your right hand, then your left hand and the computer is instructed to associated the electrical patterns generated to a right and left movement. When you think about moving your hand, even though you are not actually moving the hand, the motor cortex starts to prepare for the movement creating specific electrical patterns. Notice, however, that these patterns are specific to the way you think about the movement; a different person will generate different patterns, so a computer that can understand you will not be able to understand me. A re-training will be required.

Researchers are working to simplify the training, both for the person and the computer. For the person, researchers are studying visualization techniques that can provide effective feedback to the person about what the computer understands from the detected electrical activity. Figure 9 provides an example of rendering where the person is seeing the intensity and localization of the detected brain electrical activity. By thinking different thoughts, the person can see how the detection changes and can select a specific pattern to indicate the intention to perform a specific action.

The training of the computer is becoming an important area of research and results are very encouraging thanks to the adoption of artificial intelligence. Through artificial intelligence the computer can learn specific patterns and adapt to a specific person and to changes occurring in the way that person's brain creates electrical patterns as time goes by. This is crucial when implanted electrodes are used since the software needs to cope with the changing electrical patterns of the brain to avoid the need to re-implant the electrodes (and related invasive procedures with limited but unavoidable damage to neurons).

So far, particularly in case of invasive procedures (implant in the brain or on the cortex), the studies are focusing on interfaces to control prosthetics (for disabled people that cannot use alternative interfacing approaches, like voice). In prosthetics, a brain to computer interface is also pursued to increase the effectiveness. As an example, an interface based on eye movements to identify letters on a screen and then form a word works but the speed is very slow. It may take up to ten seconds to identify a single letter on a screen using eye gaze/tracking. The hope is to be able through BCI to achieve a communication speed that can compare with our normal one. This technology is still quite far away.

Several researchers working in the BCI field feel that the adoption of ever more sophisticated AI can be a game changer in this area, moving the training from the person to the computer and, even more important, allowing the computer to keep learning and adapting to changes in the brain's electrical activity patterns.

Notice that in order to work, AI software needs training; that is, it has to be told what certain patterns are related to. Hence the need for lengthy and repeated sessions with the person where the software is notified via textual communication what the person is thinking and what the person's intention is. It requires full cooperation (focus) from the person, particularly in the first phases where the AI software has to work out significant patterns from the mess of detected electrical activity. In a second stage, the software will look to isolate those patterns in electrical activity that is no longer focused.

Development of BCI is available now using OpenBCI and acquiring, at an affordable price, the basic hardware pieces needed to start.

Reverse Communications, From the Computer to the Brain

If it is difficult to communicate from the brain to a computer, the reverse is close to impossible. There are many hurdles to overcome to make this technology feasible. The brain is a massively distributed system that receives input from several channels (the 6 senses: sight, sound, smell, taste, touch and proprioceptors). The signals arriving from our senses are distributed to many areas in the brain, and it is the parallel "processing" in each of these areas that modifies the status of the whole brain, creates awareness, and prompts the emergence of intelligence (and consequent decision taking whenever needed).

Consider, as an example, the pathways of sight. The data coming from the retina(s) are sent to several cortical areas as well as to both hemispheres through the optic chiasma and to the amygdala and lower parts of the brain. If one of these pathways is broken the person will lose part of the sight *experience*. As an example, the interruption of the pathways to the amygdala hampers the perception of danger. You can see a truck approaching but you no longer perceive it as a danger, hence you don't take action to move out of the way. Interrupt the pathways to the occipital cortex area, and you no longer see the truck but if the pathway to the amygdala is functioning, you will move away from the truck path (this is called blindsight, you are technically blind but still perceive an approaching danger).

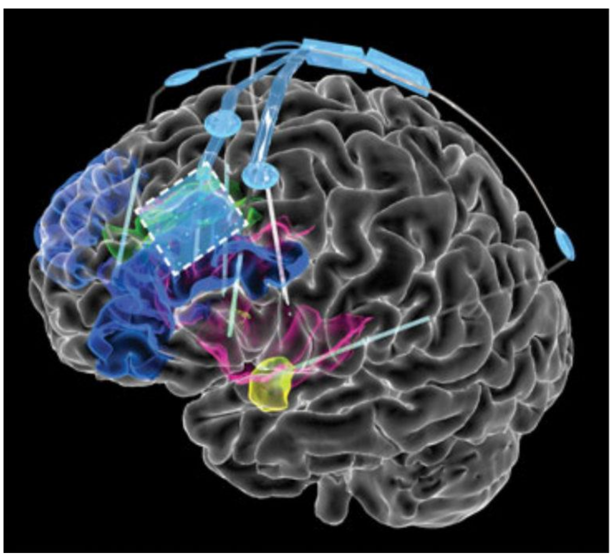

Figure 10 Rendering of brain implants for treating depression. The chips intercepts mood swings and detect the insurgence of depression. By releasing electrical spikes researchers are working to stop depression or at least decrease it. Image credit: UCLA

This preamble demonstrates the importance of activating several (many) neuronal circuits in different parts of the brain to establish communication with the brain. This is clearly impossible with the present level of technology and seems to be an elusive goal for many decades to come.

Notice that we have technologies to stimulate the brain at the level of a single neuron as well as a whole. Optogenetics has been used in this decade to stimulate specific neurons activity. However, activating a single neuron does not result in the activation of functionality in the brain, nor in the transfer of information to the brain. Similarly, the global activity of the brain can be conditioned by sending electrical spikes to the cortex or in specific areas inside the brain, such as stopping an epileptic seizure. Or, chemicals can be used to alter the functionality of the brain. Chip implants are being experimented to relieve depression and Post Traumatic Stress Disorder (PTSD). For this latter DARPA has funded a five year research project, SUBNETS, looking to implant chips that could help relieve PTSD in soldiers.

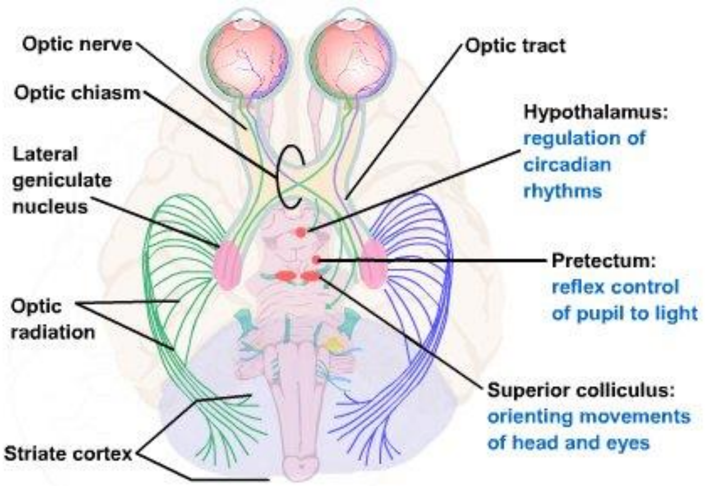

Figure 11 The inferior surface of the brain illustrating the visual pathway. The termination sites of the retinal ganglion cell axons in three nuclei that are not considered a part of the visual pathway are also illustrated. They include the hypothalamus, pretectum and the superior colliculus. Image credit and caption source: Neuroscience online

In all these cases technology is used to affect the brain but it is not an interface since information is not being transmitted to the brain.

The dream of having a chip on the brain that could be uploaded with any type of knowledge is still a dream today, and it will remain so for quite some time. What is reasonable to expect in the coming decade, as more and more knowledge on our brain is harvested and better technology becomes available, is the possibility to influence the brain to counteract some disabilities. We might even hope to stimulate the brain in ways that can improve its performance, like making it better at remembering things.

The transfer of information from a computer to a brain will rely on stimulating the senses, like bringing images to the retina or sounds to the aural nerves, and then letting the senses communicate with the brain.

Interfacing with Our Senses

As mentioned above there is a low expectation on the feasibility of establishing a direct connection from a computer to the brain, at least in the foreseeable future. A much more promising approach, already followed, is to take over the nervous pathways connecting our senses to the brain. Two aspects of computer to brain interaction mediated by senses will be presented:

- sense augmentation
- nervous pathway hijack

Before diving into these areas, it is important to understand that from the point of view of the brain the nervous pathways bringing data to the brain are basically equivalent. This may be surprising both at an intuitive level and at a structural level. At an intuitive level we know very well that hearing is completely different from seeing, taste and smell are two different things, touch is very different from the other senses and so on. At the same time anatomists indicate that the sensorial pathways end up in different places of the brain and activate different neural structures. Actually, this is a macro view of the pathway terminations. Researchers that have taken a finer view discovered that those terminations actually go, almost, everywhere, as shown in projects like the human connectome. This video shows the myriad connections inside the brain white matter.

Figure 12 Neil Harbisson is seeing colours through his aural nerve. The camera over his head picks up the colours and a chip transform them into electrical stimuli to his aural nerve. Image credit: Moogfest photographer Carlos Gonzalez

For the brain, a neuron spike is like any other neuron spike, and chemical and electrical signals flowing on the sensorial pathways produce spikes in millions of neurons. It is the whole activity, parallel and sequential, of activation and repression of neurons that generates our perception of the world, what we call seeing, hearing, smelling.

There is proof of these basic equivalence of sensorial pathways by looking at what happens in cases where there is something unusual, like a malfunction of a pathway. Over time the other pathways (connected to other senses) are used by the brain to make up for the malfunctioning ones. This is the phenomena of synesthesia. A person can start seeing colors by hearing sounds or tasting things by touch. A well-known example is the one of Neil Harbisson. He was born color-blind and received an implant where colors picked up by a digital camera were translated by a computer into stimulation of his aural nerve. After a while Neil started to "see" (or better described as "perceive") colors with his ears. This means that technology could, at least in principle, use existing sensorial pathways to stimulate the brain with data that are not harvested by our senses.

Sense augmentation

Our senses have evolved to capture certain parts of our ambient, and evolution has been biased by the ambient we have been living in. As an example, our retina is more sensitive to green (plenty of green around in places where the human journey began). Our sight can convert a tiny range of the electromagnetic spectrum (the one between 380 and 740 nm). We do not perceive infrared (like snakes) nor ultraviolet (like bees). However, technology exists that can sense a very broad range of the electromagnetic spectrum (broader than any living thing), and that technology can be used to extend our senses, as with infrared goggles (detecting longer wavelength) or night goggle (increasing the sensitivity). Rather than using goggles, seamless contact lenses can capture the desired range of wavelengths and convert them into visible ones. Research has developed electronic contact lenses that can serve as seamless interfaces; the last one announced is from Google.

Likewise, a computer could hijack one or more of our senses to communicate with the brain. Of course, this is exactly what computers do by creating images, sounds, haptic forces that are picked up by our senses. However, this may be done in a seamless way with more advanced technology, like using electronic contact lenses, ear implants, or fingertips implants. This communication, however, is limited by the capability of our senses to process the signal.

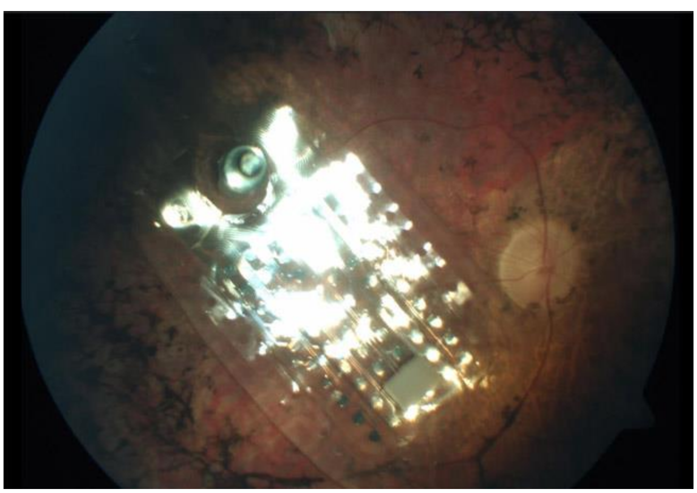

Figure 13 An Argus II implant on the retina. The implant detects the light rays entering the eye and converts them into electrical stimulation of the optical nerve, skipping the retinal cells no longer responsive to light. This establishes a direct connection to the brain. Image credit: Cleveland Clinic

Neural pathways hijack

A work around to this limitation would be to use implants that connect directly with the nervous pathways, something that is done today to overcome a sensorial disability. As an example, retinal implants are restoring a minimal level of sight to hundreds of people with retinitis pigmentosa today. The implant skips the retina and directly connects to the optic nerve. Similarly, there are many cochlear implants restoring hearing to people with a broken tympanum.

The implant procedures are still cumbersome, and the technology is far from perfect. In the future, technology will get better to the point that it can be used not just to restore a lost function but to complement existing functions, augmenting them. One potential augmentation will provide a seamless interface to a computer and to the ambient.

It will take at least two decades of technology evolution to reach that point, but the constant drive to create technology to overcome disabilities and using technology in niches (including the big one of military applications) to provide a strong competitive advantage, will make this happen.

For sure, it looks much more feasible to hijack a neural pathway then to have implants in the brain. These latter will continue to evolve, they will not likely be used as "interfaces", rather as way of affecting the whole (or a whole area of the) brain to condition macro phenomena, like the insurgence of an epileptic attack, increasing memory retention, counteracting Parkinson tremors and so on.

An intermediate step is already occurring, through the use of augmented (and virtual/mixed) reality technology, a topic being addressed in the IEEE Digital Reality Initiative. Notice that today AR/VR/XR are basically forms of linearly-advanced interfaces, but once the technology supports a truly seamless interaction (through some form of non-intrusive wearable or implants) then our communications with the ambient will change radically.

"Thoughts" to Machine Interface

The current status and limitations of BCI were previously described, showing that at the moment there is a tradeoff to be taken between higher precision and sensitivity provided by implanted electrodes/chip and noninvasive interfaces. Also, at present, interfacing with our sensorial pathways seems to be the most effective way of communicating to the brain.
What is not possible today, and even what might look impossible may become reality in a future. This is the approach taken by DARPA that in the past has funded several research projects on BCI (focusing mostly on implants) and that now has launched a research initiative for a non-invasive BCI that would allow a soldier to control a military drone (or a swarm of drones).

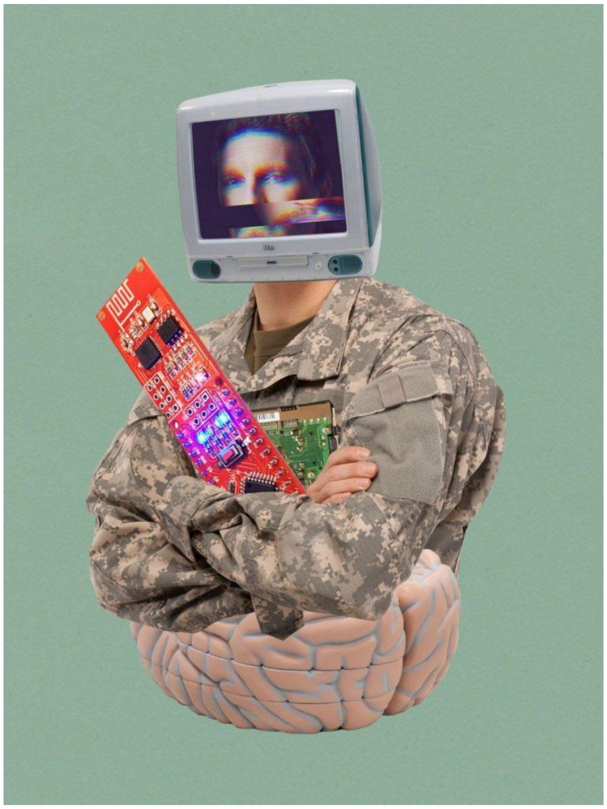

Figure 14 An intriguing image on MIT Technology Review to introduce the article on the DARPA initiative for a non-invasive BCI. Image credit: Enrico Nagel

Today there are six university research teams involved in the N3 DARPA initiative (Next generation Nonsurgical Neurotechnology program), with a $104 million funding, including Massachusetts Institute of Technology, Carnegie Mellon University, and John Hopkins University. The team principal investigator, Pulkit Grover, points out that "nothing like this is possible today and it is really hard to do", but this is what research is for.

This will likely not be developed in the next ten years, however working on this problem is likely to result in an increased understanding of the issues and will provide some practical applications in the medical domain. It is not just about the feasibility per sé; it is also about what would happen if this becomes reality.

Imagine controlling the stove from the couch with thoughts and getting distracted by what is on the television. Would dinner be ruined because the thoughts are getting misinterpreted? Clearly, this situation is a joke, but what about a soldier controlling a deadly drone and getting distracted in his mind? The implications here are huge. It is true that a soldier may get distracted as he is controlling a drone today, but the level of distractions possible inside a brain, where many thoughts are running and overlapping is much greater and more difficult to control.

Multimodal Interfaces

To conclude this white paper on human to whatever interfaces, multimodal interfaces will be described for three reasons:

- Humans have been using multimodal interfaces throughout our evolution history;
- Multimodal interfaces will be the future of interfaces;
- Multimodal interfaces are already being used in a number of situations.

Indeed, when we interact with one another, as well as when we interact with the ambient, we use multimodal interaction. We look (see), we talk (even when interacting with a dog), we touch, and sometimes we smell. When interacting with a machine, technology has forced us to use one single channel, mostly our hands (with the addition of sight used to monitor what is going on, like seeing

Figure 15 Operator at an assembly line using a, (limited) multimodal interface supported by AR glasses. Image credit: Umeå Institute of Design, Sweden

the letters we are typing on the keyboard showing up on the screen), but more recently our voice. We have adapted so well to the way we currently interact with a machine that when an alternative comes up (like speaking to the car navigator rather than entering letters) we may feel uneasy. In some cases, we have even been told that a single interaction channel is more efficient, promoting focus. In the past, pilots used to fly planes with the seat of their pants, feeling the acceleration and vibration on their body which gave them important information on what was going on and how the plane responded to their commands. Today this is no longer the case. Pilots now get information through the glass cockpit; they have lost the direct connection with the plane.

Multimodal interfaces, both as "input and output" will eventually become the norm. As devices get smarter and equipped with a variety of sensors it will become natural to have a more articulated interaction space: voice, gesture, touch, sight. Interacting with a robot will be (perhaps alarmingly) indistinguishable from interactions with another human. We will look at "its" expressions as we talk; we might take its arm to show how to move it. Sawyer from Rethink Robotics is near this type of interaction. Multimodal interactions will become a "must" for the interaction with "whatever", i.e., with a smart environment. Our environment, be it the home, the office, a department store or a hospital room will consist of several smart objects that will need to coordinate the way they interact with us while also being flexible in the way we interact with them. Digital twins, flanking each object, and possibly you and me, will play a significant role in the management of the interaction. While today our interaction with a machine is basically happening at a syntactic level (while the one with a person we are with, face to face, happens mostly at a semantic level), in the future the interactions will happen mostly at the semantic level. These kinds of interactions exploit multimodal, multichannel interfaces.

Now for the last point: multimodal interaction is already being used today when AR is adopted as an interaction interface. AR can be a powerful means for interacting with the real world assisted by cyberspace, as in the example of an operator in an assembly line. AR is providing an additional channel, a virtual one so to speak, supplementing the physical channels of our senses.

Acknowledgement
The editing and polishing of the text of this paper by Theresa Cavrak is gratefully acknowledged.