Editors: Frank Steinicke
and Wolfgang Stuerzlinger

Future Reality: How Emerging Technologies Will Change Language Itself

Ken Perlin
New York University

Today, there is much excitement as different aspects of the mixed-reality continuum such as augmented reality (AR) and virtual reality (VR) start to reach the end of Bill Buxton's "long nose" of innovation.[1] The long nose theory observes that a particular innovation will often show up in consumer products perhaps 25 years or more after the initial innovation.

As an historical precedent, consider zoom- and swipe-based multitouch interfaces. The public generally associates this interface paradigm with the introduction of the iPhone by Apple in 2007. Yet, as Buxton has pointed out, it was actually first demonstrated by Myron Krueger 25 years ago in 1982.[2] In another example, in 1968 Ivan Sutherland gave the first demonstration of immersive AR with his groundbreaking "Sword of Damocles" prototype.[3] By today's standards, the computer technology available to him was incredibly slow.

Since then, Moore's law has vastly changed the available technology. Just as notebook computers once freed us to take our computers with us, smartphones freed us to walk around with computers in our pockets, and wearables will soon free us from needing to hold a screen at all. Today, as high-quality VR and AR begins to become available at consumer prices, the "screen" will soon be all around us.

But the largest long-term impact here may not merely be one of form factor, but rather one of language itself. Once wearables become small enough, cheap enough, and therefore ubiquitous enough to be accepted as part of our everyday reality, our use of language will evolve in important ways. By "language" I mean not merely verbal speech, but also gesture, as well as our use of the physical space around us and between us to communicate with each other.

To understand why, we first need to understand the influence of form factor upon widespread adaptation of new modes of communication. As mentioned earlier, in 1982 Myron Krueger demonstrated multitouch gestures such as pinch-to-zoom.[4] Yet it took a quarter-century more before advances in form factor allowed those gestures to become part of everyday language. Today, young children accept those gestures as part of everyday reality. Unlike their parents, they have never known any other reality. This is a phenomenon that can be followed back to every generation. A few representative examples are the use of Web browsers in the 1990s, TV in the 1950s, and radio in the 1920s.

There are already many examples in science fiction of the use of gesturing in the air to control virtual objects. But children who are born into a world where wearables are ubiquitous will begin to accept such gestures as part of reality itself. It is only then that such gestures can start to become integrated into natural language as part of everyday speech. Just as we now think of a younger generation as "digital natives" and their parents as "digital immigrants," we will begin to witness a new generation of "immersive natives," to whom a seamless merging of digital information with physical reality will seem natural and unremarkable.

Brief History of Advances in Human Communication

Our ability to transmit our presence has advanced over the last 5,000 years. Written language led to moveable type, and more recently the telephone, cinema/television, and now various forms of communication over the Internet. Each successive innovation brought a more powerful or immediate way for people to transmit or broadcast their thoughts and presence to other people.

 Published by the IEEE Computer Society 0272-1716/16/$33.00 © 2016 IEEE

Yet none of these advances would have made sense without our shared instinctual ability to learn natural language because each relies on our ability to communicate via natural language. In a sense, the nontechnology of language itself, part of our genetic heritage, is the original human power-up that enables and magnifies the power of all our successive technological advances in communication.

In Another 30 Years

In the 2006 novel *Rainbows End*, Vernor Vinge posits a future in which everyone is wearing cyber-contact lenses that let them see whatever they would like. He then spends much of the novel exploring the psychological, social, and cultural implications of such a technology.

Now, a decade later, such a future seems much more plausible. It is fairly clear, given recent technological advances supported by Microsoft, Sony, Facebook, Google, and others, that the futuristic "virtual reality" Oakley Romeo sunglasses worn by Tom Cruise at the start of *Mission Impossible II* are now less than a decade away, in a high-resolution wide-angle optical see-through incarnation. Within the next 10 years, many millions of people will be able to walk around wearing relatively unobtrusive AR devices that offer an immersive and high-resolution view of a visually augmented world.

This technology will take somewhat longer to get down to the form factor of a contact lens, although there has been some interesting work already. For example, Samsung recently announced an early prototype of a contact lens that reportedly contains both a tiny camera and a crude display (see Figure 1a).

The vision is that this research will lead to the eventual development of a kind of contact lens capable of creating an image upon the eye's retina. I described such a scenario in 2010 (see Figure 1b) and called it a *hololens*. (Coincidentally, Hololens is the same word now used to describe a forthcoming commercial product by Microsoft.)

Unfortunately, the optics for this is more difficult than it might seem because the hololens is situated at an awkward place within the optical system. An image that originates within a contact lens—pressed against the eye's cornea—is too near to be imaged by the eye's own lens via conventional optics. Either a collimated light would need to originate inside the hololens (most likely in the form of a tiny laser embedded inside the hololens) or the lens would need to incorporate a fine array of micro-scale LEDs, each with its own tiny collimating lens, and each on the order of few microns in width. This is certainly possible, but the engineering required places this perhaps some years away.

(a) (b)

Figure 1. Unobtrusive augmented reality (AR) contact lens display technology: (a) The contact lens (b) may be able to create an image upon the eye's retina.

Future Reality

Why do I call this "future reality" rather than simply virtual, augmented, or mixed reality? Consider shoes.

In a hypothetical thought experiment, imagine that you could travel back in time to 1863 to discuss some finer points of the Emancipation Proclamation with Abraham Lincoln. To your surprise, as soon as he meets you, the 16th president of the United States looks down at your feet and says, "Where did you get those shoes?"

At this point you realize that the shoes on your feet are, in 1863 terms, impossible objects. They rely on materials, manufacturing and assembly methods, and global shipping practices that have yet to exist. So to Lincoln, they're going to look like future shoes—because they are.

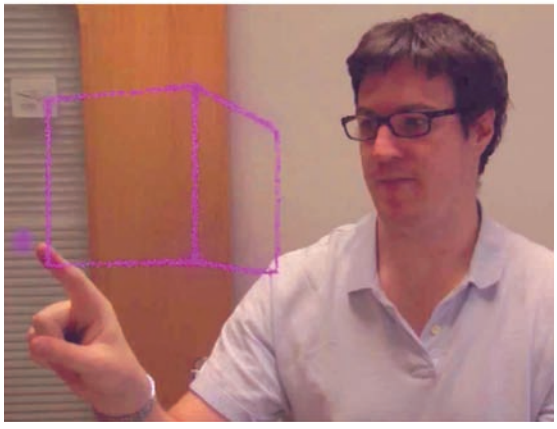
But to you, they're just shoes. And that's the point. In the future, the available technology will be so ordinary that nobody will not even think about it. It will be as ordinary as the shoes on your feet. What Abraham Lincoln might once have called future shoes, we now just call shoes.

Likewise, future reality, and everything it will make possible, may seem exotic now. But one day, we will just call it reality, whatever form that takes. Fundamental to that future reality will be the ability to communicate visually in a way that will come to seem natural and intuitive. We will take for granted that we can augment our face-to-face communication by gesturally manipulating visual shapes that appear to float in the air between us.

Prototyping the Future

To better understand what this kind of future might feel like, for much of the last decade, New York University's Media Research Lab has been

(a)



(b)

Figure 2. Arcade project aerial display mockup: (a) Free-hand drawing in the Arcade system and (b) "drawing" a predefined 3D shape.

creating a succession of working prototypes. Here are some of the projects we've been working on.

Arcade

In 2010–2011, our lab used the then-new Microsoft Kinect together with real-time video overlays in the Arcade project to create a mockup of what a "gesturing in the air" interface might be like (see Figure 2). The focus was on understanding how an audience would respond to a real-time presentation that made use of a hypothetical aerial display. We imposed the constraint that all performances needed to be given in real time, so the performers always saw the results of their gestures. We used our own gesture-recognition system to analyze the Kinect's depth image to map hand shape to "pen down" and "pen up." Some gestures were simply drawn free-hand (Figure 2a), while others were morphed in real time to predefined 3D shapes and behaviors (Figure 2b).

The speakers could only see the results of their performances on a video monitor, but the audience saw the superimposed graphics on a large projection screen behind the performer, as though those graphics were floating in the air in front of the performer. Performers mimed the process of looking at what they were drawing.

We used real-time graphics processing to convey an impression of lines of energy floating in the air, inspired by the holographic displays in the Star Wars films. The author used this method to give the keynote address at the 2011 SIGGRAPH Asia Conference. The talk was extremely well received, with attendees reporting that they felt as though they had experienced a glimpse into the future.

The positive response we received to this project encouraged us to begin work on the two projects that now constitute our major research focus in future reality: Chalktalk and Holojam.

Chalktalk

Chalktalk is a rich shared visual language we have been developing since February 2014, whereby people draw freehand, and the computer interprets those drawings to create virtual characters, procedural descriptions and processes, 3D objects, music, or anything else in its large shared vocabulary.[5] Our long-term goal is to integrate Chalktalk into a shared future reality experience so that people who are conversing with each other face to face will be able to draw in the air to support their communication, using a shared language that allows the drawings they create to "come to life" in ways that support their discussion.

Consider the traditional blackboard (or its close variant, the whiteboard). Sketching on a blackboard while explaining a mathematical concept lets a teacher tell a story in a compelling way. As each part of the concept is introduced, the emerging sketch can be timed to complement and emphasize parts of the emerging narrative.

Unfortunately, once a shape is drawn on a traditional blackboard or whiteboard, it cannot come to life to illustrate the principles in action. Chalktalk supports the ability to create freehand sketches and then link these sketches together quickly and naturally to create a working example of the concept being explained.

For example, in Figure 3 we see example screenshots from a Chalktalk lesson that is used to explain the mathematics underlying a matrix transformation algorithm for performing spatial transformations such as translations, scaling, and rotations. The full lesson includes the following procedural steps:

1. Sketch a roughly triangular shape.
2. Click on the shape. The system recognizes it as a triangle.
3. Sketch a shape: horizontal and vertical lines.
4. Click on the shape: The system recognizes it as coordinate axes.
5. Click then drag on the axes to rotate them.
6. Click then drag from triangle to axes to create a data link.

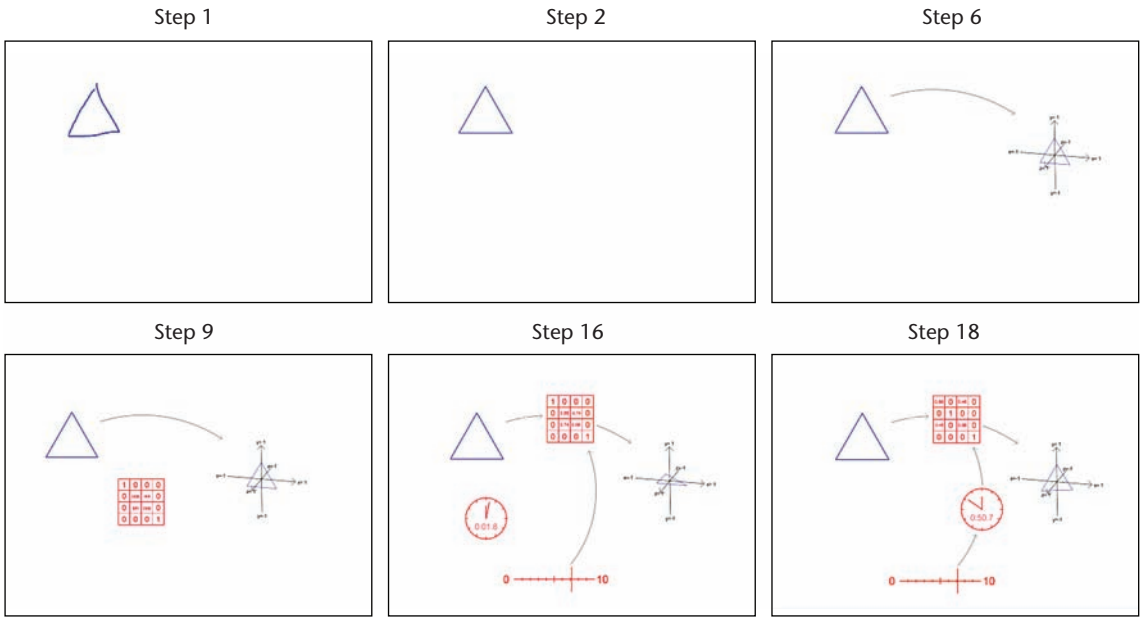


Figure 3. Excerpts from Chalktalk lesson explaining the mathematics underlying a matrix transformation algorithm. The user begins by sketching simple shapes, such as the triangle in step 1, which the system recognizes in step 2. The user can then click on the newly recognized shapes and drag them (see step 6) to form more complex, combined shapes (such as the rotation matrix in step 9 and the timer in step 16). In the final step, the user changes the axis of what has become an animated scene.

7. Sketch the shorthand shape for a matrix.
8. Click on the shape. The system recognizes it as a matrix.
9. Click on the matrix to change its state to a rotation matrix.
10. Sketch a slider-like shape.
11. Click on the shape. The system recognizes it as a slider.
12. Click then drag from slider to matrix to create a data link.
13. Drag on right-most value to change range.
14. Drag the matrix onto the first data link to create transformation.
15. Sketch a clock-like shape.
16. Click on the shape. The system recognizes it as a timer.
17. Drag the timer onto the second data link to create a rate control.
18. Change the axis of what is now an animated scene.

Because all gestures are simple and intuitive, the teacher can concentrate on the topic itself, rather than needing to focus on the interface itself. Also note the complete absence of menus in Figure 3. All visuals are directly related to the topic itself, so neither the teacher nor students are distracted by extraneous visual elements.

Figure 4 shows two other example Chalktalk lessons. In Figure 4a, several freehand sketches were linked together to create a working on-screen musical instrument. This example was used in a classroom as part of a lecture on the use of equal pitch ratios in the musical scale. We have also used Chalktalk to teach NAND gate logic; the physics of pendulums; GPU accelerated procedural texture; surfaces of revolution; gravity, velocity, and acceleration; procedural animation techniques; and modeling 3D shapes with polygon meshes.

One question that might reasonably be asked is, how general purpose is such a system? When addressing ever-more topics, would the vocabulary that Chalktalk needs grow unreasonably large? Our goal (and a major focus of our research in this area) is to understand how such a system might evolve so that it has a finite, yet general-purpose vocabulary. This system would need a grammar that can be generative, analogous to how natural languages (such as English) can generate a vast variety of meaningful sentences using only a finite vocabulary of words.

To achieve this goal, we suspect that the vocabulary needs to evolve and stabilize through use by a community. Toward this end, we have begun extending Chalktalk so that users can expand its vocabulary through purely gestural means, without needing to write any code. In this way, a user community could introduce and cooperatively choose to adopt (or not) new candidate words and grammatical constructions. This research focus builds on prior work that studies how communities of children spontaneously evolve natural languages through shared use and adaptation, such as the work on the evolution of Nicaraguan sign language (NSL).[6]

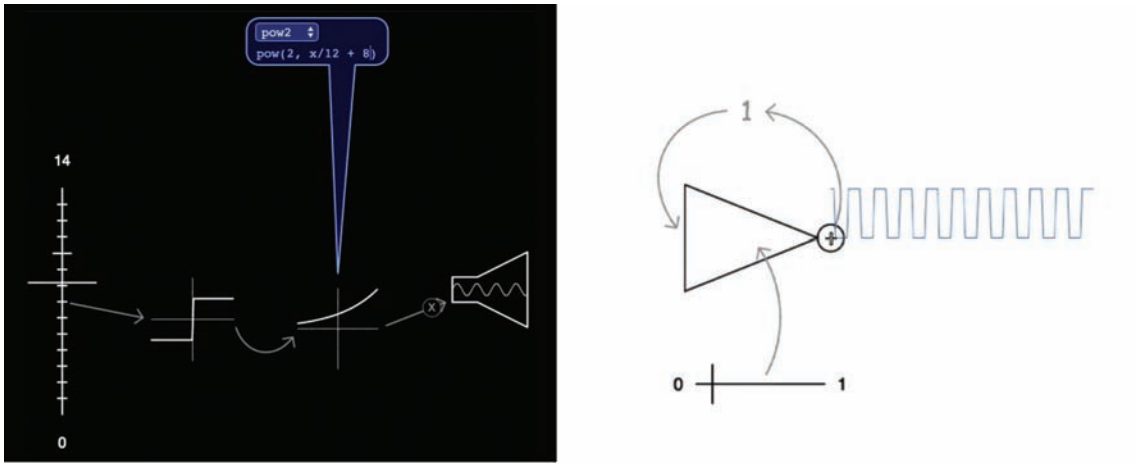


Figure 4. Example Chalktalk lessons: (a) Creating a musical instrument from sketched components and (b) using a NAND gate to create an oscillator.

Figure 5. Holojam shared immersive experience. Holojam (a) as seen from the outside world and (b) as seen by its participants in the shared virtual world.

We are currently integrating Chalktalk with another of our future reality projects, Holojam, so that people who share knowledge of a common visual vocabulary and grammar can communicate with each other face to face in powerful new ways.

Holojam

Holojam is a shared immersive experience that we first presented at the ACM SIGGRAPH 2015 conference.[7] Each participant puts on a lightweight wireless motion-tracked GearVR headset as well as strap-on wrist and ankle markers. These devices allow them to see everyone else as an avatar, walk around the physical world, and interact with real physical objects. Participants see the physical world around them, but it is visually transformed. People can draw in the air and collaborate by freely mixing physical and virtual objects.

The Holojam architecture also allows people to be in different, remote locations, yet still feel as if they are in the same room, particularly as we add mobile robotic proxies to move objects around to simulate a shared changeable physical environment.

Unlike much shared VR, this is a highly physical experience. Participants see each other in their true physical locations, re-created as avatars in an alternate "magical" world. They are free to talk to each other and physically interact with each other while they walk or run around in the shared space. Using a magic wand, they can draw sculptural shapes in the air.

In this way, each participant can contribute to an ongoing 3D sculptural art work, which is collectively created by all participants over the course of the day. At any given moment, participants see a time-slice of the emerging space-time sculpture. After a few minutes, earlier portions of the sculpture fade away, as those portions recede into the past.

An outside observer looking at the participants will see people running around, talking, laughing, looking at each other, or drawing in the air. However, if that observer looks at a computer graphic view of the VR world, they will see participants as avatars, drawing in the air, looking at each other, and having conversations about the art they are creating together. Figure 5 shows pictures of each view.

Holojam was designed to be a highly social experience for participants. Participants can talk to, observe, and physically interact with one another in the space. More participants provide more activity to observe.

The Long Term

Young children who grow up in a world where Vernor Vinge's contact lenses are ubiquitous will build on these capabilities to create their own shared gestures. But will these be extensions of natural language, much as communities of deaf children spontaneously generate rich natural language such as NSL[6]?

When we talk to each other verbally, we are using natural language. One of the defining qualities of natural language is that it is "naturally learnable," meaning that it doesn't need to be explicitly taught. Nearly all children born into any society will, in the first seven or so years of life, master much of the vocabulary and grammar of that society's verbal speech. This is in contrast to formally defined languages, such as mathematics and computer programming, that need to be explicitly taught.

Our lab's research has advanced to the point where we can put people into a shared simulation of social future reality. We can now begin to simulate those futuristic contact lenses that will allow us to graphically augment gesture. Other people are able to see the shapes that you make in the air with your hands, and you can see theirs.

It may at last be possible to find empirical answers to some important questions. For example, will the shapes that future children make in the air end up being "naturally learnable" like natural languages or must they be explicitly taught, as is the case for formal languages? This question is important because it will tell us whether communities of children will be able to spontaneously evolve and adapt future descendents of Holojam Chalktalk.

In other words, will the resulting visually enhanced communication evolve toward natural language that children will learn (and contribute to) in the normal course of their normal growth and development? Although it takes approximately seven years for the average child to master a natural language such as English, that process of mastery is indeed reliable and repeatable by individuals across entire populations.

Similarly, in a future world where children can visibly draw in the air between them as a speech act, we should not expect that some future Holojam Chalktalk would be naturally learnable by individuals in a week or a month. But we can ask whether it can be learned over the course of the first several years of a child's life and whether that shared visual language can be passed on and spontaneously evolved through shared use by communities of children.

If the capability to share visual communication under gestural control leads to a change in natural language, then as these future children grow up, they will live in a richer world, with powers of natural-language-based communication that we can only begin to imagine. The resulting communicative power-up for coming generations may be as fundamental and paradigm changing as was the development of written language itself 5,000 years ago.

References

1. B. Buxton, "Why eBay is a Better Prototyping Tool Than a 3D Printer, The Long Nose, and Other Tales of History," video, Interaction South America, Recife, 2013; https://www.youtube.com/watch?v=H5Zm7E3atW8.
2. R.S. Kalawsky, *The Science of Virtual Reality and Virtual Environments: A Technical, Scientific and Engineering Reference on Virtual Environments*, Addison-Wesley, 1993.
3. I.E. Sutherland, "A Head-Mounted Three Dimensional Display," *Proc. AFIPS Fall Joint Computer Conf., Part I*, 1968, pp. 757–764.
4. M.K. Krueger, *Artificial Reality II*, 2nd ed., Addison-Wesley Professional, 1991.
5. K. Perlin, "The Coming Age of Computer Graphics and the Evolution of Language," *Proc. 2nd ACM Symp. Spatial User Interaction* (SUI), 2014, pp. 1–1.
6. A. Senghas, "Nicaragua's Lessons for Language Acquisition," *Signpost: J. Int'l Sign Linguistics Assoc.*, vol. 7, no. 1, Spring 1994, pp. 32–39.
7. C. DeFanti et al., "Holojam," SIGGRAPH 2015 VR Village, 10 Aug. 2015; http://s2015.siggraph.org/attendees/vr-village.

Ken Perlin *is a professor of computer science in the Media Research Lab and the director of the Games for Learning Institute at New York University. Contact him at perlin@mrl.nyu.edu.*

Contact department editors Frank Steinicke at frank.steinicke@uni-hamburg.de and Wolfgang Stuerzlinger at w.s@sfu.ca.

Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.