SIP (2017), vol. 6, e4, page 1 of 8 © The Authors, 2017.

doi:10.1017/ATSIP.2017.4

INDUSTRIAL TECHNOLOGY ADVANCES

Responsive media: media experiences in the age of thinking machines

BO BEGOLE

This discussion of responsive media *provides a perspective on the future of media experiences that are increasingly responsive to users' preference, alertness and their physical, digital, and social environment. By examining a range of future scenarios combining virtual-, remote-, and augmented-reality, autonomous vehicles, digital assistants and robots, we see that the responsiveness of media is what provides the key value. To reach the ultimate goal of* augmented innovation *in which thinking machines supplement humans, there are a number of technological and user-experience challenges that the research community needs to resolve. These challenges fall into a few key categories: throughput, latency, perception, intelligence, and interaction. While some challenges may be tackled purely technologically, others require insights from sociology and psychology to break new ground. The paper concludes that intelligent, responsive media will not fully supplant human intelligence, but will increasingly serve as augmentation to human creativity.*

Keywords: Responsive media, Interactive media, Remote reality, Virtual reality, Augmented reality

Received 9 May 2016; Revised 29 March 2017

I. INTRODUCTION

The future of digital media is more exciting than ever. While pundits search for the next big thing among a dizzying array of shiny ideas (drones, virtual reality, digital assistants, autonomous vehicles, and more), we should also notice that many technologies have achieved critical mass to enable a new world of audio and video experiences – media that responds dynamically to your attention, preferences, and context. This new era of *responsive media* differs from interactive media of the past by not asking the audience to deliberately control the media, but by creating intelligent media experiences that sense the user's environment and situation and indirectly infers the optimal branch of content to present to the user in real time.

As with many computing paradigms of the past, the notion of responsive media originates at Xerox PARC [1, 2] and incorporates a number of emerging technology concepts all of which have media responsiveness as the central element.

Virtual reality (VR) – immersion within a synthetic or remote reality (RR) (i.e. other than the user's current, physical reality) – responds to user's head and body position.

Augmented reality (AR) – overlaying supplemental digital information over the user's perceived reality (i.e. can be a media stream or a virtual or remote reality) – responds to user's gaze, needs and situation.

Remote reality – capturing all perceivable sensations (light-field, sound-field, other data) and transmitting it in real-time to remote humans who respond across geographies. (Remote Realities may also be recorded, as in 360° video capture, but the responsiveness is then triggered by the viewer's head and body motion, similar to that of virtual reality.)

Semi-autonomous vehicles – partially automated vehicle control – responds to driver's alertness and traffic context.

Digital assistants – assists people in digital tasks – responds to user's physical, social, and digital context.

Social robots – assists people in physical tasks – responds to humans within a physical environment.

The term responsive media subsumes the above categories.

Responsive media – digital information that does not require deliberate user input and that responds appropriately according to preference, alertness, state of mind and physical, digital, and social context.

Figure 1 illustrates the combination of technologies that feed into this new era of media experiences and the wide array of devices and applications to which responsive media adds value. This paper extrapolates across these trends to identify technological and user-experience challenges that the multi-media research community needs to resolve.

Media Technologies Lab Head, Huawei R&D, Santa Clara, CA, USA

Corresponding author:
B. Begole
Email: bo@begole.net

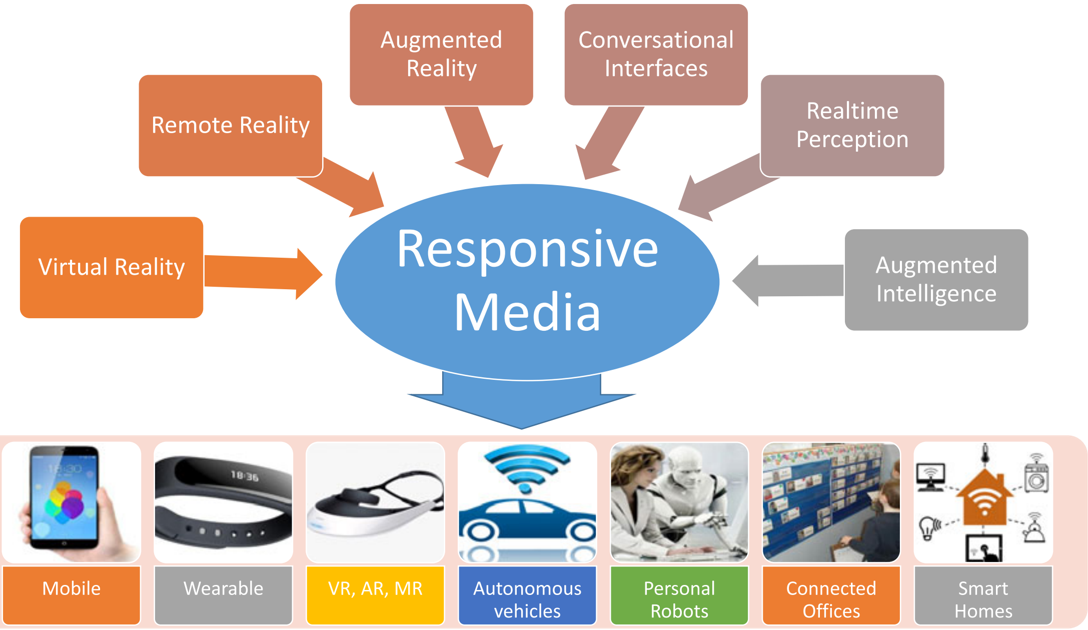

Fig. 1. Responsive media contains the core media technologies that enable a number of computing platforms, products, and services.

II. RESPONSIVE DIFFERS FROM INTERACTIVE MEDIA

Although there is some overlap, the notion of responsive media (described in [3]) is broader than that of "interactive media". Interactive media requires audience members to deliberately select branches of media content, whereas in responsive media, the system makes a selection based on an inference of the audience's preference. In the latter case, the audience is not required to break out of their experience in order to control the narrative flow, allowing them to continue to suspend disbelief while consuming the content.

Responsive media applies to more than just entertainment, however. The notion applies also to forms of human–computer interaction (HCI) that emulates human–human communication. Responsive media goes farther than speech recognition, which is basically still a human issuing commands to a computer. In a responsive system, the computer can sense unspoken human emotions, states of mind, or *intentions*. The system is capable of taking the initiative to respond to the human need, sometimes even before the human is aware of it herself.

A recent research prototype called #scanners from the University of Nottingham uses a consumer electroencephalogram (EEG) to detect eye blinks and to estimate a viewer's attention and meditative levels [4]. From those inputs, the system can estimate various aspects of the viewer's state of mind such as engagement, arousal, attention, relaxation levels, and more. The system switches among four different perspectives (layers) of a video depending on the viewer's state of mind to offer a dynamic viewing experience.

Rather than using sensors that require direct electrode contact, off-body sensors such as cameras, microphones, and infrared sensors can detect a number of clues that indicate our intentions. Head position and gaze tracking is a clear indication of what a person is paying attention to. Drivers often look over their shoulder, for example, before changing lanes. Expression and vocal tone can indicate concern, urgency, frustration, anger, amusement, and a variety of other emotional states that systems can use to prioritize their actions. Semi-autonomous vehicles should know when the driver or passengers are angry or afraid, for example. Pupil dilation is an indication of surprise and skin flush can indicate embarrassment or the adrenalin rush of excitement. We humans use these clues all the time when interacting with each other; we estimate what might be on another person's mind based on behavior cues and shared knowledge.

Similarly, computing systems are increasingly able to construct estimates of what a human intends via models and currently sensed environmental cues. Sensing technologies have advanced to be able to detect fine-grained human activity and behavior. Whereas we used to have to explicitly tell a computer what to do, increasingly they are inferring our intention and proactively making suggestions. For example, some smartphones have the ability to notify their owner of when to leave for an appointment based on traffic conditions. The owner does not need to instruct the phone that she intends to leave enough time for travel, but it knows that it is a common human intention.

Sensing and vision technologies will become more accurate, less expensive and more pervasive. They will see finer grained clues of human situations and intentions. Responsive media will accelerate as more sensors come online, not simply because more data are generated, but because machine learning and user-experience designers will be able to map the situation to human goals and intentions.

III. REMOTE REALITY

We live today in multiple parallel realities. There is of course the mundane physical reality in which we live. Then there are the multitudes of computer-mediated realities made ever-more real by the novel head-mounted displays making headlines every day. Such synthesized views of the world can be wholly computer generated or they can be captured by cameras, microphones and other sensors to bring other real-world places closer through the transmission of video, audio and other data in real-time from remote places where we are unable to travel – *remote reality* (RR). Spanning across all these categories is an orthogonal dimension in which any reality (remote or synthetic) is overlaid with supplemental information from the digital realm – *AR*[1].

Remote Reality is poised to supplant old notions of telepresence with new abilities that verge on a kind of *omnipresence* in which we can virtually teleport anywhere in the world at any time [7].

Remote reality, when fully realized, will be far more immersive than the *telepresence* of the past by rendering remote objects at full life size, full stereoscopic depth and full surround audio to create a strong sense of reality. We can refer to the ultimate realization as *full-field communication* when it reproduces a realistic subset of the light field and the sound field at the remote location.

Remote reality will enable a wide range of new business applications by providing full visual, audio, and (eventually) tactile fidelity of a remote environment. More thorough descriptions of such business cases can be found in [7, 8] and are briefly categorized here.

Tele-medicine – full diagnostic detail can be seen by physicians in remote places, allowing consultations with remote specialists and also training of rare conditions to medical students worldwide.

Tele-repair and tele-operation – similar to telemedicine, but for machinery, remote specialists can troubleshoot and guide on-site technicians through complex procedures.

Tele-tourism – people will be able to operate drones to tour exotic parts of the world, underwater and eventually out in space.

Tele-education – unlike today's online education systems which are basically recorded lectures, with RR students can visit hard-to-reach parts of the world or virtually examine microscopic and macroscopic phenomena.

Sports and music events – audience members can ride along with athletes and performers to become part of the event.

Commerce – hard-to-find items from large-scale construction to artisanal crafts can be examined in full detail, expanding markets.

IV. CHALLENGES

Consider the potential when media responsiveness intersects virtual, augmented and remote realities. At that point, we will see a computer-based interaction that approaches human abilities to infer another's intention. Achieving the full capacity of Remote Reality requires overcoming a number of technology challenges.

A) Challenges for networked reality

The aim of VR and RR is to generate a digital experience at the full fidelity of human perception – to recreate every photon your eyes would see, every small vibration your ears would hear, and eventually other details like touch, smell and temperature. Although the quality of video and audio media is improving in terms of resolution, dynamic range and color gamut [9], fooling human perception is a difficult challenge because humans can process an equivalent of nearly 6.2 terabits/s of visual information (uncompressed), as explained below and enumerated in Table 1.

The fovea of our eyes can detect dots as fine-grained as 0.3 arc-minutes of a degree [10], meaning we can differentiate approximately 200 distinct dots per degree in the fovea's field of view. Converting that to "pixels" on a screen depends on the size of the pixel and the distance between our eyes and the screen, but let us use 200 pixels per degree as a rough estimate. While the fovea's field of view is a somewhat narrow 5°, our eyes can mechanically shift (saccade) across a field of view of approximately 150° horizontally and 120° vertically [11] within an instant (less than 10 ms in some cases [12]). That's 30 000 horizontal by 24 000 vertical pixels. This means the ultimate display would need a region of 720 million pixels per eye (1.44 gigapixels for stereo) for full coverage because your eyes can saccade across a wide field of view within an instant.

Those are just for a *static* image; but the world does not sit still. For video, multiple static images are flashed in

Table 1. Theoretical maximum data rate of human visual perception. Boldface indicates maximum theoretical visual data rate of human vision (without allowing head or body motion).

	Value	Units
Horizontal field of view (FOV)	150	Degrees
Vertical FOV	120	Degrees
Human visual acuity	0.3	Arc-minute
Arc-minutes per degree	60	Arc-minutes
"Pixels" per degree	200	Pixels/degree
Horizontal pixels/FOV	30 000	Pixels
Vertical pixels/FOV	24 000	Pixels
Pixels in FOV	720 000 000	Pixels
Stereo? (1 = no, 2 = yes)	2	Eyes
Pixels in Stereo FOV	1 440 000 000	Pixels
Number of color channels	3	Per pixel
Bits per channel	12	Bits/channel
Bits/FOV	51 840 000 000	Bits per image
Frame rate	120	Frames/s
Bandwidth/s	**6 220 800 000 000**	**Bits/s**
Compression ratio	600	:1
Compressed	10 368 000 000	Bits/s

[1] The term *mixed reality* (MR) is sometimes considered synonymous to AR but Milgram and Kishino differentiated AR as a subset of MR [5]. Computer-mediated reality is an even more generic superset encompassing VR, MR, AR, RR, and more [6].

sequence, typically at a rate of roughly 30 frames per second (fps) today for film and television. But the human eye does not operate like a camera. Our eyes actually receive light constantly, not discretely, and while 30 fps is adequate for moderate-speed motion in movies and TV shows, the human eye can perceive *much* faster motion – some estimates are as high as 500 fps [13, 14]. For sports, games, physical science experiments and other high-speed immersive experiences, high video frame rates will be needed to avoid "motion blur" and disorientation.

Assuming 120 fps which is supported in the recent video standard ITU-R BT.2020-2 (Rec. 2020) [15], near instantaneous head and body rotation, 36 bits per pixel for full color gamut, stereo rendering, the total is 6.2 terabits/s. Today's video compression technologies can preserve consumer-grade quality using a compression ratio of roughly 300:1 [16, 17]. Even if future compression technologies are able to double that, reaching a factor of 600:1, systems still need 10.36 gigabits per second (Gbps) of throughput. Now add head and body rotation for 360 horizontal and 180 vertical degrees and we arrive at a total of approximately 18.2 Gbps, compressed. This represents only an estimate of the data rate that human vision system is capable of ingesting as our eyes sample the light fields from a three-dimensional (3D) space using 2D retinas; encoding the full 3D world would involve additional properties to represent the light rays characterized by a full plenoptic light field[2]. Of course, many optimizations can be performed to reduce this estimate to a more practical level such as perceptual coding, adaptive streaming, and foveated streaming.

Today, such throughput requires local rendering on a computer with a video cable directly connected to the display (e.g. HDMI 2.0 provides 18 Gbps and DisplayPort 1.2 can reach 32.4 Gbps). As more advanced network technologies are deployed to the home and wirelessly, this rendering could move to the cloud where more compute and memory resources can be brought to bear. A requirement for network-based rendering, however, is to achieve end-to-end latency (that is the time between when a user interacts with the system and when she sees the effect) of <100 ms [18]. That budget has to be met even as data packets cross-multiple links and servers in a network. Even lower latency will be required for VR streaming in order for users to feel that the responses to motion are "instantaneous" and to avoid motion sickness induced by mismatches between the human body's sense of physical motion (proprioception) and visual perception [19]. Zheng et al. find that even 2 ms of latency is perceivable by some users [20].

B) Challenges for intelligent interaction

Speech interaction will be increasingly necessary as we create more and more devices without keyboards such as wearables, robots, AR/VR displays, autonomous cars, and internet of things (IoT) devices. This will require something more sophisticated than the scripted pseudo-intelligence that digital assistants offer today. Like humans, digital attendants need to speak, listen, explain, adapt, and understand context. In addition, agents need to understand not only speech, but also the gestures of people as they use their bodies to add information such as diexus (pointing), action and emotion. All of these recognition technologies are increasing rapidly, but still have far to go to match human-level interactivity.

1) Speech recognition has come of age

Not long ago speech recognition was so bad that we were surprised when it worked at all; now it is so good that we are surprised when it does not work. Over the last several years, speech recognition has improved dramatically and is approaching the accuracy at which humans recognize speech [21]. There are three primary drivers at work here.

First, teaching a computer to understand speech requires sample data and the amount of sample data has increased dramatically as mined search engine data is increasingly the source [22].

Second, new algorithms have been developed using deep neural networks and other machine learning techniques that are particularly well-suited for recognizing patterns in ways that emulate the human brain [21].

Finally, recognition technologies have moved to the cloud where large data sets can be maintained, and computing cores and memory can be scaled easily. Though sending audio data over a network may delay responsiveness, latencies of mobile networks are decreasing to address that limit. The anticipated latency for 5 G networks is 1 ms (for the physical layer) [23].

The result is that many users are increasingly talking to their smartphones for a variety of tasks. We can expect speech input to be a dominant input mechanism in the currently emerging computing systems including intelligent vehicles, IOT, robots, and others.

2) Intelligent conversational interaction

Speech recognition is only the bottom layer of the intelligence stack, however. To make the interaction truly natural, the machines need to make sense of the speech. Today's agents seem amazingly intelligent with abilities to control the devices, retrieve complex information queries, and to coordinate services on the web.

Today's agents use techniques that are a step beyond yesterday's key word-based search, but fundamentally they are still based on matching and search, not true understanding. The agents today use a form of natural language "understanding" that detects what task the user is trying to accomplish (intent) and the properties of the task (entities). Keywords and entities are recognized in the utterance and matched against slots in a task template, which is then used to execute the task.

Basically, the system recognizes a command phrase (usually a verb) that identifies a task domain like *call*, *set an alarm for*, or *find*. Each of these task domains evokes a kind of "template" that the system needs to fill in with properties like the name of the person to call, the time for the alarm, the

[2] Minimally, this would add two more dimensions per pixel: vertical and horizontal angles of incidence.

name of a place to find or book a reservation. If it does not find all the necessary information in the user's statement, it can ask for more details in a kind of scripted dialog.

Today's agents can take and execute commands, but they do not approach the proactive service of a human concierge who intuitively understands your desires and can even suggest things before one would think to ask. Today's assistants cannot go off-script when recognition fails. They often cannot explain their own suggestions. They cannot anticipate problems and suggest alternatives. They rarely take the initiative.

Nevertheless, today's digital assistants have raised our expectations of the intelligence that our devices are capable of. With this new standard of quality, users may soon come to expect even more truly supportive and conversant assistance.

3) Human-level conversation

What would it mean to have a truly conversational agent that is a revolutionary advance over today's digital assistants and worthy of our future IoT filled with wearables, autonomous vehicles, robots, and intelligent appliances?

An intelligent agent is already capable of making decisions based on task-domain ontologies, user utterances and preferences, and available services. To reach human-level ability, a responsive media system would need a few more qualities.

Semantically deep – First, a more fully responsive system would need to have language understanding less superficial than what we have today. Computers can easily miss intent or become confused and fall back to simple web search, because the system does not really *understand* what the user means. If it fails to recognize the type of task it is being asked to perform, it cannot retrieve a predefined script with which to ask for more details. A human would be able to remedy the specific misunderstanding by taking what he or she *did* understand and asking for more information in order to determine the task domain.

Explanatory – Unlike the "black box" recommendation systems today, a deeper language model will allow a conversational system to explain *why* it recommends a particular action or why it thinks something is true, just as a human can.

Resourceful – When humans detect a problem, we can plan around it and suggest alternatives. For example, although today's mapping systems can plan a route and even anticipate whether you will arrive before the store's scheduled closing time, a deeply intelligent agent should proactively notice exceptions to standard schedules and also to suggest alternatives. For example, if a restaurant I scheduled for lunch with a colleague is closed that day for a religious holiday, it should recommend something nearby that fits my general preferences. It could also suggest nearby parking alongside map directions, knowing that I will need to park after driving. Today's agents fall short of such resourceful problem solving.

Attentive – Responsive media systems should be constantly attentive. If one of my children tells me she just took the last yogurt out of the refrigerator, the system should notice and add it to our shopping cart without anyone having to tell it to. Of course, constant monitoring evokes concerns about invasions of privacy that must be addressed in the technology and experience design.

Socially intelligent – Intelligent agents should be aware of social situations including engagement with other people in the environment and follow common social norms when interacting with humans.

Context-aware – Social intelligence is actually a subset of the broader category of contextual intelligence [24, 25], which is the type of human intelligence that understands relationships between people, places, things, and actions. A context-aware system can predict the likely actions a person would take in a given situation based on the location, presence of other people and objects, and the knowledge of past actions in such situations.

Engaging – Perhaps most importantly, a responsive media system should engage people and express understanding of the importance of their requests. In human conversation, a tone of urgency is met with responsiveness. Humor is met with amusement. Worry is met with sympathy and suggestions. In all cases, humans use tone of voice to indicate an understanding of a desire for urgency, mirth, empathy, or resolution.

We do not need a mechanical personality to replace human companionship, just to create a more conversational interaction style that connects in a richer, more natural way. Recently, advances in speech recognition have been largely achieved by data-driven machine learning algorithms, but advanced conversational agents will likely need more knowledge-based techniques where experts instruct the system in order to achieve the semantic richness of human conversation.

4) Computers take the initiative

Speech input along with intention recognition extends the ways we interact with computers. No longer are we simply instructing the machines; now they will initiate interactions. This is a fundamental shift in our interaction paradigm from deliberate commands to implicit expectation that computers will know what to do for us.

This follows a natural progression in HCI where we add new ways to tell computers what we want them to do. In the early days of mainframe and mini-computing, the command language was very specific and technical. In the personal computing era, graphical user interfaces (GUI) allowed users to construct commands by pointing at icons and windows and by pressing menus and buttons. More recently, we have been able to use gestures and speech to show and tell our smartphones what to do.

HCI has marched through a progression of ways for humans to tell the machine what to do, illustrated in Table 2. Initially, instructing a computer required writing and loading programs and data, evolving to the command-line interfaces, through the more direct manipulation of GUIs and touch screens, and now to direct speech input. In all of those

Table 2. Progression of paradigms by which computers understand human users' intentions.

Era	1950s	1970s	1980s	2000s	2010s		2020s and future
User intention types	"Do a computation."	"Do what I instruct."	"Do what I touch."	"Do what I show."	"Do what I say."	"Do what I usually do."	"Do what I mean."
Input technologies	Punch cards, paper tape, magnetic tape	Keyboards. Terminals	Mouse, trackpad, touch screen, stroke recognition	Depth cameras, gesture recognition	ASR, NLP, Service coordination	GPS, accelerometer, camera, other sensors	Emotion detection, activity recognition, computer vision, eye tracking, head and body position, pulse, skin flush, vocal tone, laughter, etc.
Device types and applications	Mainframes, batch processing	Workstations, Personal Computers, keyword search	Personal Computers, Smartphones	Smartphones, Personal Digital Assistants	Smartphones, Smart homes	Smartphones, Smart objects	VR headsets, AR environments, Smart objects, VR Avatars, Smart Environments (IoT), Semi-autonomous vehicles (ADAS, HMI), Responsive Media

cases, the user is formulating a set of instructions for a computer to execute. But that is when computers were blind and deaf – now they can see, hear, feel and sense more about the external world than we humans ourselves. With that, they can not only see the current state of humans and objects, but anticipate what comes next – what the humans would want to have happen in such circumstances.

Already our location-aware phones are beginning to anticipate our needs from sensors and to proactively *tell us* what to do, such as when to leave for an appointment in order to account for traffic conditions. Soon, semi-autonomous cars will know when traffic conditions are becoming more dangerous to the point of telling the human driver to pay more attention. Robots will know when a child needs additional hydration and offer beverages in advance. Many more services will be enabled as machines become more aware of human needs. To be safe, the robots, cars and appliances should ask for confirmation before taking an action. But even in asking for confirmation, the fundamental paradigm of HCI has shifted. Instead of the human commanding, or "using", the computer, the computer has initiated the interaction.

In order to gain human confirmation before acting, the systems of tomorrow will need to gain our attention. Like humans, they should do so politely yet effectively – not with buzzers or garish graphics. Systems will need to be considerate of what is going on, other people in the environment, human emotional state and attention. They should follow normal social rules and not interrupt when people are talking, not jolt us with alarming sounds, not obstruct our goals with unwanted ads and other distractions.

V. A RESEARCH AGENDA FOR RESPONSIVE MEDIA

The prior sections have laid out a number of technological challenges for achieving richer responsiveness. Whether embodied in a digital assistant, robot, autonomous vehicle, or smart building, the aim of responsive media is to approach the abilities of humans in interpreting, anticipating, and proactively responding to other humans. This clearly implies a deeper understanding of human-to-human interaction than we have today.

Component technologies exist to perceive the state of humans, places and things in the world, but we have little understanding of human behavior models. What are sequential structures of engaged interactions? Which physiological indicators are most predictive of engagement? Can we prevent disengagement before it happens?

Today's digital agent scripts are focused on concrete task-specific objectives, urging the human to "fill in the blanks" to complete the task. Human assistants work at a more abstract level. For example, Fig. 2 shows the abstract elements of a sales interaction identified by Robert Prus [26] and modeled as a finite state machine that is not well enough defined for a computer to execute, but that is well

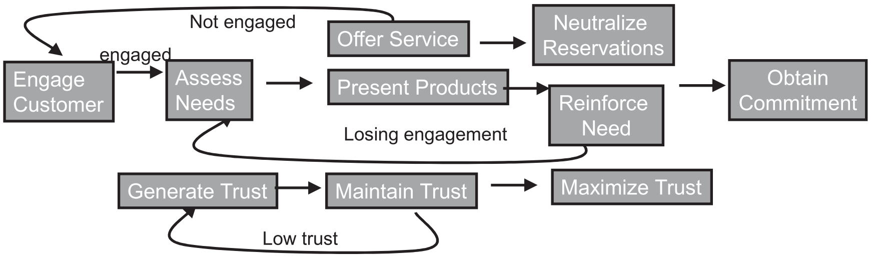

Fig. 2. Abstract conversation model of a human salesperson interacting with prospective customers.

enough defined for a human (after some explanation or training).

Despite advances in AI through data-driven models and machine learning, we must also call on social and cognitive sciences to reach our goal.

VI. CONCLUSION

Media technology research is no longer simply about the digitization of video, audio and text. As digital media have become more pervasive, our research field has become more complicated to now include computational photography and videography, light field capture and display, computer vision and hearing, volumetric rendering, AR overlays, and more. Furthermore, the application domains of media technologies are no longer limited to the consumption of video and audio, but now include interaction with wearables, autonomous devices (cars, robots, and drones), interactive art and entertainment, and smart environments. The fields of multi-media technologies will continue to grow and bifurcate, but the one unifying problem that pervades all of the new sub-fields is that media increasingly needs to be *responsive* to the user's situation, preferences and objectives.

The idea of responsive media serves as a container for a collection of related technologies. Also, in contrast to purely technological labels that do not speak to what benefit a technology provides (e.g. "virtual reality" only tells us the reality is "virtual" not the benefits of such virtuality), the term responsive media reminds us of the utility of these technologies and *why* they are important – they are media that *respond* to stimuli, providing information at the point of need. Responsive media is a clear trend as computing systems become increasingly intelligent, perhaps exceeding human abilities in some tasks. For now, computing systems are still just tools and they will be increasingly useful as they allow humans to focus on higher-level and creative thinking.

ACKNOWLEDGEMENTS

I would like to express appreciation to current and former members of Huawei's Media Technologies Lab for contributing insight and technology toward this vision. Special thanks to Jose Alvarez for encouraging this strategic perspective on media technology research.

REFERENCES

[1] Zhang, W.; Begole, B.; Chu, M.: Asynchronous reflections: theory and practice in the design of multimedia mirror systems. Multimed. Syst., 16 (4–5) (2010), 293–307.

[2] Creating a new media business opportunity and technology platform – PARC, a Xerox company. [Online]. Available: https://www.parc.com/services/case-studies/2201/creating-a-new-business-opportunity-and-technology-platform.html [Accessed 9 November 2016].

[3] Begole, B.: The dawn of the age of responsive media, forbes. [Online]. Available: http://www.forbes.com/sites/valleyvoices/2016/01/12/the-dawn-of-the-age-of-responsive-media/ [Accessed 9 November 2016].

[4] Pike, M.; Wilson, M.L.; Benford, S.; Ramchurn, R.: # Scanners: a BCI enhanced cinematic experience, in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016, 293–296.

[5] Milgram, P.; Kishino, F.: A taxonomy of mixed reality visual displays. IEICE Trans. Inf. Syst., 77 (12) (1994), 1321–1329.

[6] Billinghurst, M.; Clark, A.; Lee, G. "A Survey of Augmented Reality." Foundations and Trends® in Human–Computer Interaction, Now Publishers, Boston, 8.2-3 (2014): 73–272.

[7] Begole, B.: Omnipresence and the coming age of 'remote reality, VentureBeat. [Online]. Available: http://venturebeat.com/2015/07/13/omnipresence-and-the-coming-age-of-remote-reality/ [Accessed 24 January 2017].

[8] Begole, B.: Parallel realities: full-field communications – Huawei Publications. [Online]. Available: http://www.huawei.com/en/publications/winwin-magazine/vr-or-nothing/parallel-realities-full-field-communications [Accessed 9 November 2016].

[9] Lee, H.-C.: Introduction to Color Imaging Science, *Cambridge University Press*, Cambridge, UK, 2005.

[10] Curcio, C.A.; Sloan, K.R.; Kalina, R.E.; Hendrickson, A.E.: Human photoreceptor topography. J. Comp. Neurol., 292 (4) (1990), 497–523.

[11] Wandell, B.A.: Useful quantities in vision science, in Foundations of Vision, *Sinauer Associates Inc*, Sunderland, Massachusetts, 1995. inner cover pages.

[12] Fischer, B.; Ramsperger, E.: Human express saccades: extremely short reaction times of goal directed eye movements. Exp. Brain Res., 57 (1) (1984), 191–195.

[13] Armstrong, M.G.; Flynn, D.J.; Hammond, M.E.; Jolly, S.J.E.; Salmon, R.A.: High frame-rate television. SMPTE Motion Imag. J., 118 (7) (2009), 54–59.

[14] Davis, J.; Hsieh, Y.-H.; Lee, H.-C.: Humans perceive flicker artifacts at 500 Hz. Sci. Rep., 5 (2014), 7861–7861.

[15] Rec, I.: BT. 2020. Parameter Values for UHDTV Systems for Production and International Programme Exchange, *International Telecommunication Union*, Geneva, 2012.

[16] Akramullah, S.: Digital Video Concepts, Methods, and Metrics: Quality, Compression, Performance, and Power Trade-off Analysis, Apress, 2014.

[17] Sullivan, G.J.; Ohm, J.-R.: Meeting Report of the First Meeting of the Joint Collaborative Team on Video Coding (JCT-VC), Dresd. DE, 2010, 15–23. [Online] Available: http://ftp3.itu.int/av-arch/jctvc-site/2010_04_A_Dresden/JCTVC-A200.doc. Accessed 30 Apr 2017

[18] Miller, R.B.: Response time in user-system conversational transactions, in *Proc. AFIPS Fall Joint Computer Conference*, 1968, 267–277.

[19] John Carmack's Delivers Some Home Truths On Latency, Oculus Rift Blog, 26-Feb-2013.

[20] Zheng, F. et al.: Minimizing latency for augmented reality displays: Frames considered harmful, in *2014 IEEE Int. Symp. Mixed and Augmented Reality (ISMAR),* , 2014, 195–200.

[21] Amodei, D. et al.: Deep speech 2: End-to-end speech recognition in English and mandarin, *ArXiv Prepr. ArXiv151202595*, 2015.

[22] Chelba, C.; Bikel, D.; Shugrina, M.; Nguyen, P.; Kumar, S.: Large scale language modeling in automatic speech recognition, *ArXiv Prepr. ArXiv12108440*, 2012.

[23] Andrews, J.G. et al.: What will 5 G be? IEEE J. Sel. Areas Commun., 32 (6) (2014), 1065–1082.

[24] Sternberg, R.J.: Beyond IQ: A Triarchic Theory of Human Intelligence, Cambridge University Press, New York, NY, 1985.

[25] Begole, B.: Ubiquitous Computing for Business: Find new Markets, Create Better Businesses and Reach Customers Around the World 24-7-365, *Ft Press*, Upper Saddle River, NJ, 2011.

[26] Prus, R.C.: Making Sales: Influence as Interpersonal Accomplishment, vol. **172**, *Sage Publications, Inc*, New York, 1989.

Bo Begole is the VP and Global Head of Huawei's Media Technologies Lab, which creates technologies spanning ultra-high-efficiency compression, computer vision/hearing, augmented/virtual reality, and full field immersive communications. Previously, he was Sr. Director at Samsung Research's User Experience Center and a Principal Scientist at Xerox PARC where he directed the Ubiquitous Computing research program. He is the author of Ubiquitous Computing for Business, dozens of peer-reviewed research papers and has been granted more than 30 US patents. Dr. Begole is an ACM Distinguished Scientist and is active in many research conferences in the field of Human Computer Interaction. Dr. Begole received a Ph.D. in computer science from Virginia Tech in 1998 and prior to that was enlisted in the US Army as an Arabic language interpreter.