# Mixed Reality Social Prosthetic System

*Ariel M. Greenberg and Jason A. Spitaletta*

AR VR MR

## ABSTRACT

*A Johns Hopkins University Applied Physics Laboratory (APL) team conceived of and developed a first-of-its-kind mixed reality "social prosthetic" system aimed at improving emotion recognition training and performance by displaying information about nonverbal signals in a way that is easily interpretable by a user. Called IN:URfACE (for Investigating Non-verbals: Using xReality for the Augmented Consideration of Emotion), the proof-of-concept prototype system uses infrared sensors to measure facial movements, pupil size, blink rate, and gaze direction. These signals are synchronized in real time, registered in real space, and then overlaid on the face of an interaction partner, such as an interviewee, through a mixed reality headset. The result is dramatic accentuation of subtle changes in the face, including changes that people are not usually aware of, like pupil dilation or nostril flare. The ability to discern these changes has applications in fields such as law enforcement, intelligence collection, and health care. This article describes how the system works, the technical challenges and solutions in designing it, and possible areas of application.*

## INTRODUCTION

The study of emotion includes psychological,[1] neuropsychological,[2] psycholinguistic,[3] and psychophysiological[4] and psychopathological[5] perspectives and their concomitant research methods. The psychophysiological concept of emotion incorporates both the autonomic and somatic nervous systems[4] and is considered a scientific descendant of Darwinian theory[6]; however, recent research identifies limitations of Darwin's views on emotion.[7] As the scientific study of emotion progresses,[8] contemporary models of emotion incorporate affect (the subjective experience of emotion), cognition (learning, memory), sensation, perception, and the neural correlates thereof.[9] This theoretical and empirical evolution owes much to the technological innovations that enable more advanced methods of understanding both the objective measurement of emotion and the subjective experience of emotion.[10] Humans are inherently social[11] and much of that interaction relies on emotional awareness, or the knowledge of the feelings present in oneself and others and the ability to incorporate them into different aspects of cognition such as decision-making and/or problem-solving.[12]

Understanding emotional states through nonverbal signals is a key component to inferring the thoughts and feelings of others.[13] This inference is known as empathic accuracy,[14] and these inferences can be made by carefully observing the face. This line of research was pioneered by Paul Ekman, one of the most influential psychologists of the 20th century.[15] The human face communicates through several nonverbal channels, including facial expressions, jaw tension, blood flow in the cheeks and forehead, changes in the size of the pupils, and movements of the eyes.[14,16] Understanding social signals from faces is critical for social interaction, but people vary widely in their ability to perceive, interpret, and assess these signals, particularly their truthfulness.[16] While Ekman's research remains influential, the evolving models of emotion suggest far more complexity in both the expression and interpretation of emotion.[17] Nevertheless, the human face remains a rich source of emotional information, but it can be challenging to interpret for neurotypical individuals, much less those with specific deficits.[18]

The scientific understanding of emotion has benefited from advances in technology, but the ability to address deficits in and maladaptive regulation of emotion, and/or the inability to accurately interpret emotion, requires greater attention. Detecting subtle facial signals is difficult, but research suggests that it is trainable.[16] Current training is restricted to seminars and online training materials and does not always replicate real-world environments in which emotion detection is needed. Incorporating mixed reality (XR) technology in training, with real people in real environments in real time, allows for more complex and ecologically valid training scenarios. Further, XR enables the user to receive information in context and without interruption. Together, these advantages impart confidence that XR treatment may improve training and information transfer, particularly to new counterparts in new situations. Furthermore, XR technology in combination with other wearable sensors may afford the opportunity to not only accelerate training of specialty fields but also address the deficits of the impaired.[19]

The concept of a social-emotional prosthetic has been abstracted by the artificial intelligence research community for some time;[20] however, XR has helped concretize those discussions. In particular, the capability for highly sensorized XR systems to focus on both the wearer and the other allows for a better examination of cognitive components such as attention and memory of the user while additional sensors are focused outward on the other. The capability to measure cognitive abilities in XR[21] is particularly useful when employing the more theoretically sophisticated component models of emotion that incorporate attention and memory along with affect.[17]

With the goal of improving emotion recognition training and performance, a cross-disciplinary APL team developed IN:URfACE (for Investigating Non-verbals: Using xReality for the Augmented Consideration of Emotion), a proof-of-concept prototype headset-based XR system that displays information about nonverbal signals in a way that is easily interpretable by a user (see Figure 1 for an illustration of the system's goal and the video on APL's YouTube channel, https://youtu.be/oNi1pCj6tY4, for a quick demonstration). The system collects facial signals by using a range of sensor modalities, synchronizes the signals in real time, registers them in real space, and overlays changes on the face of an interaction partner, such as an interviewee. Changes in facial signals, including expression, blood flow, and muscle tension, may be communicated to the headset wearer through "glimmers," or ephemeral holographic overlays, placed within the real world. Such overlays allow the user to receive information from the XR system in real time without interrupting the social interaction. With peripheral equipment, highly sensorized XR systems may be capable of detecting changes in voice and body language, as well as psychophysiological changes (heart rate, skin conductance, and blood pressure) associated with emotion and arousal that individuals may not be able to detect unaided.

This glimmer approach departs from alternative systems (e.g., Autism Glass[22] and Brain Power[23]) that also support emotion recognition by means of XR. These systems are prescriptively oriented toward recognition: when worn, the headsets declare the emotion expressed by a counterpart's face so that the headset wearer can match their perception of the emotion to the headset's identification. Instead, the IN:URfACE system enhances visibility of and draws attention to the changing facial actions that compose an emotional display so that the user can attune to the components of an expression and make the determination themselves. The authors believe this approach to scaffolding whereby the "experienced partner"[24] is the XR system will ultimately be able to accommodate a wider range of skill sets.

Amid growing concerns about the application of artificial intelligence to emotion recognition,[25] this more conservative and constructive approach enables the user to make their own determination given accentuated pre-

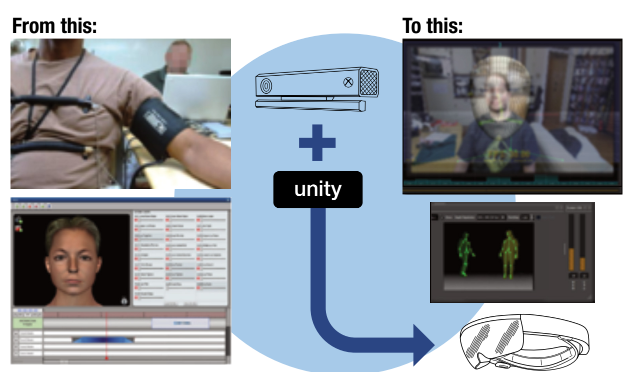


**Figure 1.** Goal of IN:URfACE MR social prosthetic system. The system aims to improve emotion recognition training and performance by displaying information about nonverbal signals in a way that is easily interpretable by a user.

sentation of the involved cues. By conveying landmarks instead of directions, this approach seeks to avoid a situation analogous to a GPS user blindly following guidance and driving their car into a lake. Furthermore, the approach affords the opportunity for the user to better develop self-efficacy during dyadic interaction, a component that more prescriptive approaches fail to cultivate.[26]

Depending on the context, the system may be used to train or also in practice. Studies demonstrating skill transfer from equipped training to unequipped practice will inform usage.

Many contexts stand to benefit from improved skills in empathic accuracy.[14] In contexts where it is critical to establish an affiliative stance, these skills have varying names: in intelligence, they are called operational accord[27]; in medicine, bedside manner[28]; and in psychotherapy, rapport.[29] These skills are also essential for the conflict resolution activities commonly required of law enforcement officers.

Those with deficits in this capacity, such as those on the autism spectrum, often find face reading challenging.[30] As a clinical intervention, this system might help compensate for communicative deficits.[31] It is this potential as an assistive technology that inspires its classification as a prosthetic.

## SYSTEM DEVELOPMENT OVERVIEW

Figure 2 is a high-level illustration of the system. The system includes an eye tracker, a depth camera for registration, a computer to process sensor readings and place them within a virtual environment registered to reality, a wireless access point, and an MR headset. Each component is described below.

### HoloLens and XR

The Microsoft HoloLens headset enables the system's XR component. The interviewer wearing the headset is able to view sprites overlaid on the interaction partner's face—for example, dots indicating changes in facial expressions or lines tracking eye movements (see Figure 3). The overlays allow the wearer to receive information from the MR system in real time and without computer mediation to avoid interrupting the social interaction.

In the proof of concept, the bulk of the processing was performed offboard and the processed output was streamed to the HoloLens, as shown in Figure 2. Ultimately, this capability would be embodied in an all-in-one solution for field deployment.

In addition to alerting the headset wearer to subtle changes in facial expression and eye movement, the system may be configured to measure other signals that

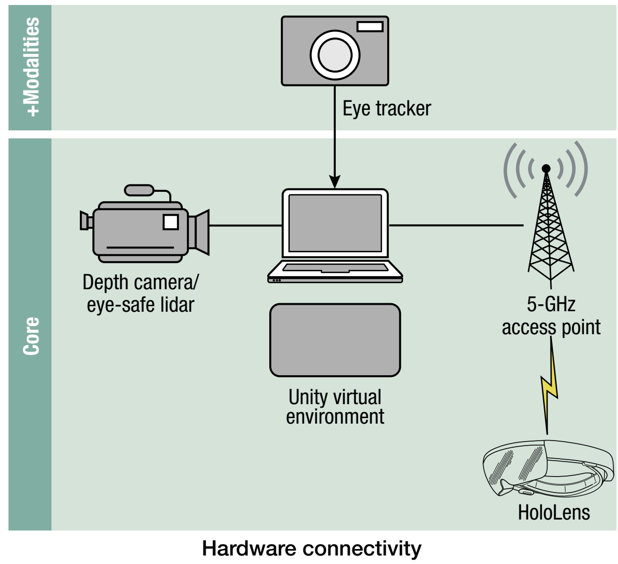


**Figure 2.** IN:URfACE system organization. The system includes an eye tracker, a depth camera for registration, a computer to process sensor readings and place them within a virtual environment registered to reality, a wireless access point, and an MR headset.


**Figure 3.** The MR social prosthesis system enables finer perception of nonverbal cues. The interviewer wearing the headset is able to view data points overlaid on the interaction partner's face, enabling the wearer to receive information in a way that is easily interpretable, in real time, and without interrupting the social interaction.

**Table 1.** System measurements, modalities, sensors, and MR displays

| Phenomenon | Modality | Sensor | MR Display |
|---|---|---|---|
| Musculature, microexpressions | Depth | Microsoft Kinect v2 TOF-IR | Action unit glimmer |
| Pupil diameter/saccadic activity | Eye-tracker, pupilometer | SensoMotoric Instruments (SMI) RED-m eye tracker (now a legacy device) | Colored blinking outlines |
| Hue and periodics: heart, blink, and respiration rates | Visible light | Microsoft Kinect v2 RGB | Eulerian magnification |
| Temperature, blushing/ blanching | High-frame-rate long-wave infrared (LWIR) | FLIR E60bx | False color |

people may have difficulty perceiving, including changes in blood flow or in muscle tension. See Table 1 for phenomena the system may measure, the modalities and sensors associated with each measurement, and the way the system displays each type of change to the headset wearer.

## Face Tracking

The system uses a depth camera (Microsoft Kinect v2 and its associated Software Development Kits [SDKs]) to capture facial muscle actions. From point clouds at a resolution of ~0.25 megapixels, ~100 facial points are tracked via the Microsoft Face Tracking SDK implementation of the CANDIDE-3 face model. These facial points are assembled as ~10 shape units (e.g., nose). These shape units, when associated with directions of movement, comprise ~20 action units (e.g., nostril flare). The coordination of those action units constitutes expressions, which correspond to ~6 basic emotions (e.g., disgust). This mapping to emotion by the Face SDK implements Ekman and Friesen's Facial Action Coding System (FACS).[13,32,33] To reconstruct a skinned, animated emoting face for overlay in the HoloLens, these data are fed into a Unity instance. Brekel Pro Face 2 wraps most of these features into a single product. This process is summarized in Figure 4.

### *Eye Tracking*

Eye movement is tracked with the SMI RED-m, a now legacy device that exploits the red-eye effect in which infrared (IR) is differentially reflected off of the retina through the pupil. This effect makes the pupil appear brighter than the iris and serves the basis for gauging pupil size, gaze direction, and blink rate.

### Pupil Size

To measure changes in pupil size, the system first establishes a baseline pupil size from a user-triggered calibration. The system smooths the curve of noisy temporal measurements of pupil size to lend an average value over a time window, and then filters out any measurement that is outside typical values. Finally, it normalizes the pupil value by dividing the smoothed and filtered pupil size by the baseline pupil size to provide a more stable metric of size change and direction. Figure 5 illustrates this process.

### Gaze Direction

Eye rotation is measured from changes in corneal reflection. The cornea is bulbous, so the incident shape of the bright region contorts from round when viewed head-on to oblong when viewed obliquely. The SMI eye tracker interface provides an (x, y) screen coordinate for

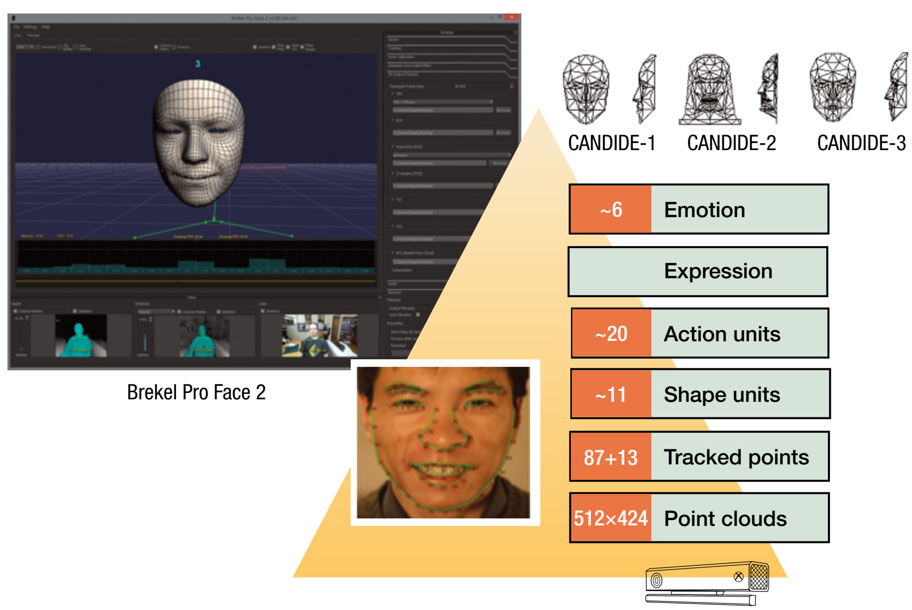


**Figure 4.** Face tracking. The system's depth camera captures facial muscle actions. From point clouds, facial points are tracked. These facial points are assembled as shape units (e.g., nose), which, when associated with directions of movement, comprise action units (e.g., nostril flare). The coordination of those action units constitutes expressions, which correspond to basic emotions (e.g., disgust).

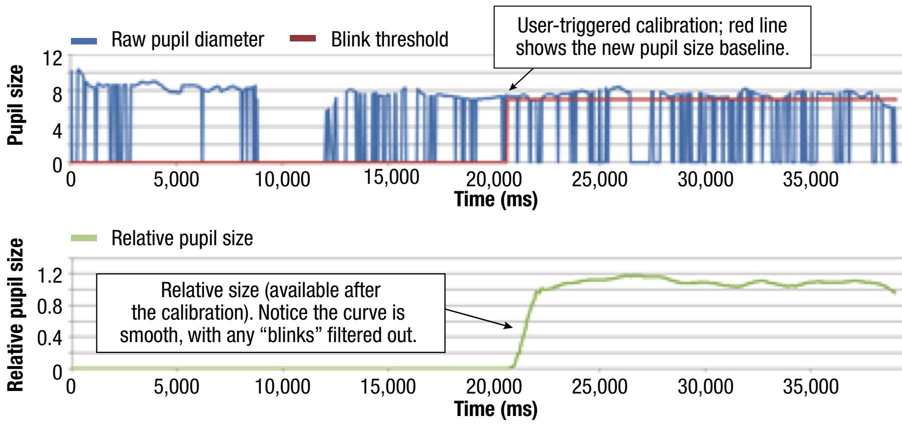


**Figure 5.** Pupil size measurement. The system first establishes a baseline pupil size from a user-triggered calibration (top graph). It then smooths the curve of noisy temporal measurements of pupil size to lend an average value over a time window and filters out any measurement that is outside typical values (bottom graph). Finally, it normalizes the pupil value by dividing the smoothed and filtered pupil size by the baseline pupil size to provide a more stable metric of size change and direction. (Note that the pupil data shown are typical.)

where each eye is "looking." These screen coordinates are converted to directional vectors by ray cast.[34] The direction and rotation are derived from this vector using Unity's built-in Quaternion library.[35] The typical signal coming out of the eye tracker is noisy and has value "gaps" caused by either eye blinking or the tracker's loss of eye registration.

### Blink Rate

Blinks are indirectly detected by the loss of registration of the bright region. This loss of registration is determined to be a blink versus another obscuration on the basis of its duration. As shown in Figure 6, a single "blink" is defined as occurring when the size of the pupil falls below the blink pupil size threshold parameter, remains below the blink threshold for at least the blink duration parameter, and finally returns to above the blink pupil size threshold parameter.

## Real-Time Sensor Streaming

Sensor data must be sent, synchronized, processed, and displayed to the user without introducing lag that will make the user experience awkward and unnatural. Unity on the HoloLens alone cannot connect to Brekel on the laptop computer because of incompatible networking libraries (UWP). Therefore, the system coordinates Unity instances on the laptop *and* on the HoloLens. Unity provides an application programming interface (API) that handles networking by sharing a game scene between "local players" and networking using a central server.

In this networked multiplayer game, the laptop acts as a server and a local player. The HoloLens acts as another local player. The shared game scene is the mask tracked on the Kinect. The mask is the local player's "game character." Two masks can be seen at one time, and movements of the masks are synced both ways. This solution eliminated the need for two masks and confined all processing to the laptop.

## Real-Time Spatial Registration

External sensor data must be transformed to the user's point of view, and locations and dimensions for the overlays must be calculated. Both people in a two-person (dyadic) social interaction move about, so overlays must be transformed to account for changes in position relative to static sensors and to one another. HoloLens and Kinect each have separate spaces, but to

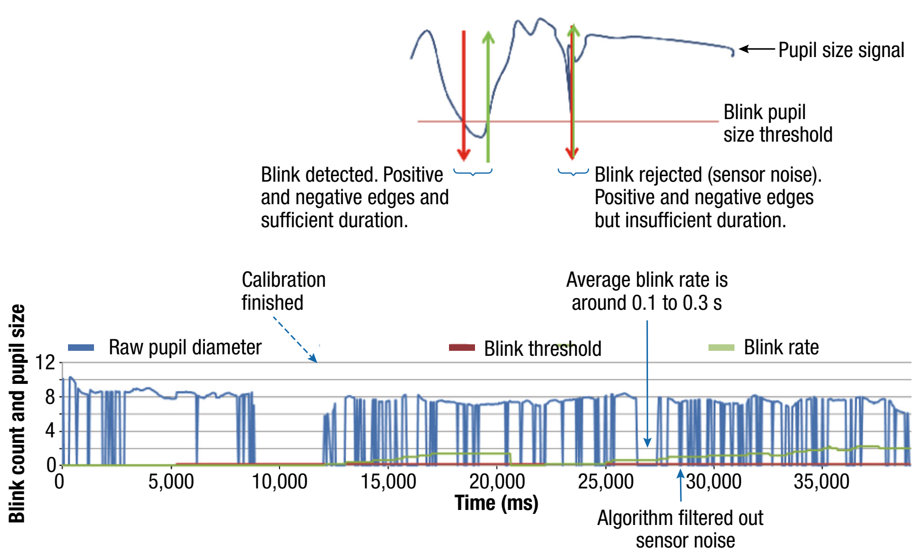


**Figure 6.** Blink rate measurement. A single "blink" is defined as occurring when the size of the pupil falls below the blink pupil size threshold parameter, remains below the blink threshold for at least the blink duration parameter, and finally returns to above the blink pupil size threshold parameter. (Note that the pupil data shown are typical.)

map a mask onto a real face, these spaces must match. An object being tracked by both devices is needed to sync these spaces. Figure 7 illustrates the process for arriving a registration solution.

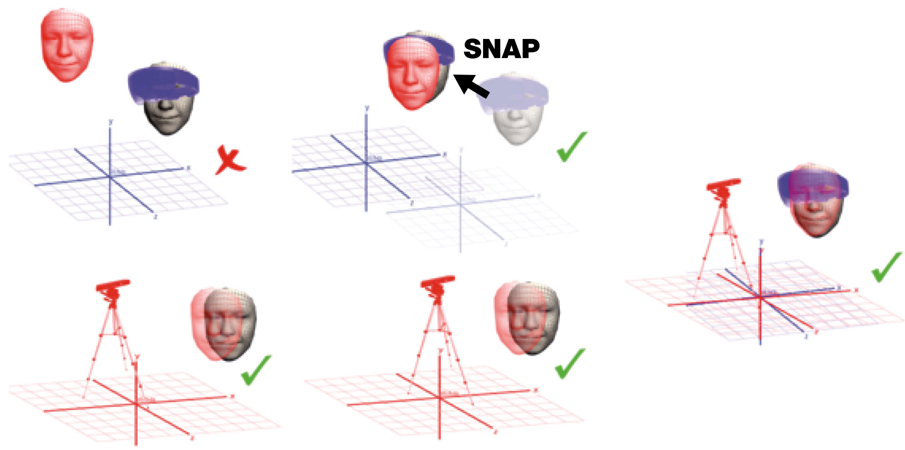


**Figure 7.** Real-time spatial registration. In the first attempt (left), the HoloLens was placed directly on the Kinect before the applications were started. With both devices looking from the same viewpoint, the HoloLens's space was roughly aligned, revealing that face alignment was off by a few feet. In the second attempt (middle), the HoloLens was placed on the user's head while the user's face was captured by the Kinect. Hitting the Enter key moved the HoloLens's space so that user's face "snapped" in front of the mask captured by the Kinect, but face alignment was still off by a few inches. The third attempt involved manually adjusting the virtual position by changing position values. This was a tedious and time-consuming process that could not be performed in real time. So instead, the Space key was held to enable the HoloLens's movements to nudge the mask toward better registration (right).

## Overlay Design

Overlays must be easily interpreted by the user while not obstructing the user's vision or interfering with natural interaction. To avoid overwhelming the viewer by displaying an abundance of information all at once, the system offers a selection of information presentation styles. One option displays the original mask with facial points provided by Unity. Another removes the opaque mask but retains the facial points for better visibility.

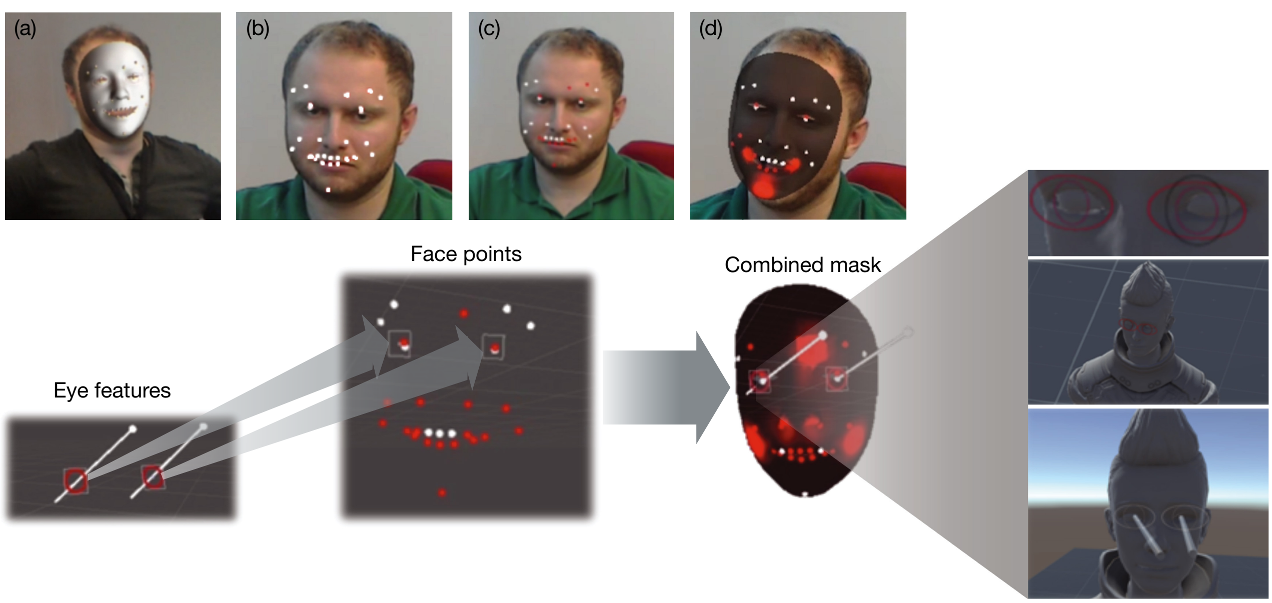


**Figure 8.** Overlay presentation options showing facial actions and eye features. One option displays the original mask with facial points provided by Unity (a). Another removes the opaque mask but retains the facial points for better visibility (b). Another style displays color-changing dots highlighting areas of movement (c). The most sophisticated style uses the color-changing dots as light sources that reflect on a transparent mask to illuminate wider regions of movement on the face (d, where the mask is shown in gray for emphasis; the dark regions of the mask are transparent through the HoloLens). Black circles represent the baseline pupil size, and colored circles represent the current pupil size. The size of the circles represents the current pupil size relative to the baseline (e.g., if the colored circle is smaller, the current pupil size is less than the baseline pupil size). Blink rates are represented by ellipses, with black ellipses representing the baseline blink rate and colored ellipses representing the blink rate relative to the baseline (e.g., if the colored ellipsis is bigger, the current blink rate is higher than the baseline blink rate). Gaze directions are displayed as cylinders, with each cylinder rotated to point to the direction of the gaze.

Another style displays color-changing dots highlighting areas of movement; in other words, the dots change colors if regions of interest move about the face faster than a threshold speed. The most sophisticated style uses the color-changing dots as light sources that reflect on a transparent mask (shown in gray for emphasis in Figure 8, the dark regions of the mask are transparent through the HoloLens) to illuminate wider regions of movement on the face. This approach to implement glimmer is intuitive and performant.

The eye features are anchored (eye features are set as children to the corresponding parent points on the face mask) to their respective left and right eye points of the face mask. As the overall mask moves, so do the points. Since the eye features are mounted to the points, they also move with the overall face.

## NEXT DIRECTIONS

With the goal of later differentiating the system toward the application areas previously mentioned—intelligence, law enforcement, and health care—next directions are concerned with advancing the system to support the interpretation of facial cues and to develop metrics to evaluate the system's performance.

Three classes of factors converge in the forging of an expression (emotional display): environmental, circumstantial, and mental (Figure 9). The goal is for the system to consider context to decompose observable state into contributing components. The system could reveal confounds: is the facial signal witnessed a reflection of emotion or something neutral? For example, mouth movement could be caused by smiling or by readying to speak. The system could assess congruence; understanding whether facial signals are conflicting is critical for detecting emotional leakage, or fleeting displays that reveal suppressed affect, like quick and tiny movements of the face called microexpressions. Environmental factors also bear on psychophysiological measurands and need to be ruled out to attribute a display to affect. For example, pupil flare may due to changes in ambient light or to changes in workload. Additionally, the system may process and integrate multiple cues in coordination to assist the user in assessing social signals. Such advancements support understanding, enabling the system to sense contextual information and then integrate it with facial cues to support more accurate interpretation of the social meaning of those cues.

Since the environment, including ambient temperature, light, and sound, can alone change facial display, the system needs environmental sensing to disambiguate the source of change. Toward this end, IN:URfACE would be extended with sensors that capture ambient properties from the environment for correspondence with the psychophysiology of the subject. For example, a change in skin temperature below the nose may indicate stress, but only under controlled humidity.[36] Pupil dilation indicates anger, but only under controlled lighting.[37] Because environmental context circumscribes the meaning of particular measurements, this fusion of directed and ambient sensing endows contextualization capacity in preparation for the introduction of additional sensing modalities.

Humans' internal states influence how they interpret actions and information—happy and relaxed people are more likely to perceive happiness in others, while stressed people are more likely to perceive aggression in others.[38] Hills appear steeper when viewers are fatigued.[39] Measuring users' psychophysiological responses as well as those of their subjects informs the system of how users visually search scenes and how they react to the meaning derived from scenes. This information may be fed back to the user to help them see any biases in their interpretation of a scene, and to the system itself to help it understand how people respond to varying types of scenes.[40]

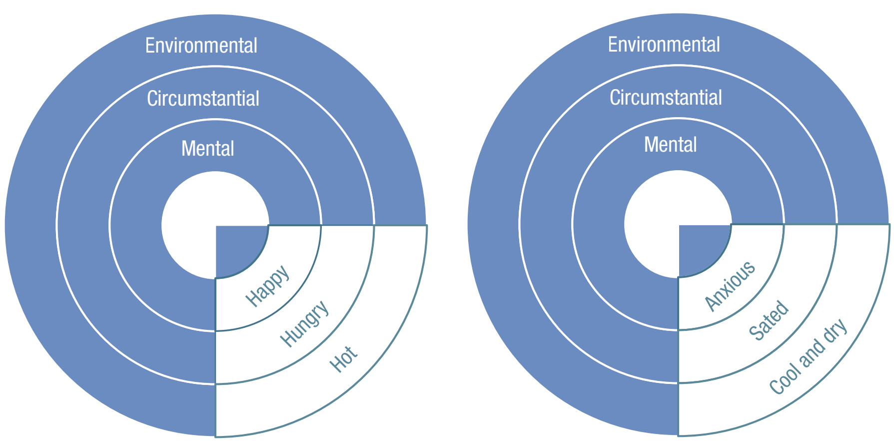


**Figure 9.** Classes of factors that converge in the forging of an emotional display. The goal is for the system to consider context to decompose observable state into contributing components.

Metrics to quantify how well the system performs in a naturalistic setting, like a conversation, must be defined in stride with development. Basic metrics gauge sensing and processing accuracy, such as the area under the receiver operator characteristic curve for detection of facial and ocular actions. Advanced metrics gauge the accuracy of the system at determining, for example, whether an apparent expression is due to an environmental change versus an emotional display.

The system was developed in the DISC-o, the Dyadic Interaction Studies Co-Operative, in APL's Intelligent Systems Center. This space is dedicated to the study of the social exchanges between agents engaged in one-on-one interaction. It is equipped with XR development hardware and software and various psychophysiological sensors. Contact sensor modalities in the DISC-o include functional near-infrared spectroscopy (fNIRS, now being upgraded to double the channels and enable hyperscanning, or the simultaneous collection from two participants), electromyography (EMG), electrocardiography (EKG), electroencephalography (EEG), photoplethysmography (PPG), electrooculography (EOG), electrodermal activity (EDA), and respiration sensors. Standoff sensor modalities include thermal, hyperspectral, radio frequency (RF), time-of-flight infrared (TOF-IR) depth sensors, eye-safe lidar, and pupillometry. It is possible to append each of these sensors onto the platform, some of which was accomplished with the Novel Perception project (see the article in this issue).

Assessing multiple faces simultaneously is very difficult, even for those most adept. See the video at https://www.jhuapl.edu/Content/techdigest/videos/multiplex-microexpressions.mp4 for a demonstration. For screening purposes, the system could be extended to aid the user in detecting anomalous reactions to a query broadcast to a group of individuals.

## CONCLUSION: APPLICATION OPPORTUNITIES

Success in many fields depends on skilled understanding of social signals. The XR social prosthesis system was initially developed for intelligence interviewers and police officers, who could use the system to better detect deception. Deception is detected through a combination of facial expressions, body language, voice, verbal content, and verbal style,[41] in addition to other psychophysiological signals.[42] Currently, nonmechanical deception detection is taught through in-person seminars and online interactive software, but how much this training generalizes to new situations is an open question. XR could increase the speed of skill acquisition, improve retention, and help students generalize what they have learned. Most people can detect deception only at chance levels,[43] and lengthy training is often required to improve this ability. Another application in this domain is to improve skills in threat assessment and conflict de-escalation and resolution. For example, the system could help train officers to recognize and overcome the impact of stress on perception of emotion. These skills would also benefit other fields, such as diplomacy and business, where negotiation is important. The system may also help inform the development of robotic autonomous systems that can adjust their behavior based on feedback from subtle changes in their human partners' psychophysiology.[44]

This system also has applications in health care. For example, practitioners could use it to improve the speed and accuracy of emotion assessment. With support in recognizing and responding to patients' nonverbal communication, practitioners may better develop bedside manner and rapport with patients to achieve favorable interaction outcomes. Building such alliances with patients is particularly important for psychotherapists and/or physicians, for whom XR shows promise as a training modality.[28]

The system might also be used to help restore function of patients experiencing social deficits as a result of a traumatic brain injury or cognitive decline, or of individuals along the autism spectrum. Deficits in any of the underlying components of emotional awareness can lead to a variety of dyadic and social challenges and diminished quality of life. Researchers have examined the feasibility of wearables to assess the functioning of individuals along the autism spectrum, and initial results are encouraging.[45] Future iteration of IN:URfACE may include means of objectifying the Social Responsiveness Scale, long considered a standard in the assessment of social functioning in individuals along the autism spectrum.[46] The system could also be adapted for clinicians or caretakers who interact with these specialized populations to assist them in interpreting social signals and emotional displays on an individualized basis. To explore these applications, the team has engaged with potential partners, including the Kennedy Krieger Institute and The Arc of the United States, and completed a survey on assistive technologies for individuals with intellectual and developmental disabilities.

The convergence of biology, psychology, and technology inevitably generates concerns about the ethics of human performance modification and augmentation. Exploring the efficacy of wearable technologies to both better quantify performance on emotion-laden cognitive tasks (either in a laboratory and/or naturalistic setting) and help qualify emotional experience for those with deficits can lead to a deeper understanding of affect across a broader range of capabilities. Few technologies are more promising than the highly sensorized variants of XR, like the system described herein, for such purposes, but far more research and technological development are required.

**ACKNOWLEDGMENTS:** Special thanks to the rest of the team, including Scott Sperling (now at Northrop Grumman), Brandon Scott, Kevin Chiou, and Alice Jackson. APL's Catalyst Program funded this effort through an FY17 Combustion Grant.

## REFERENCES


[1]K. Oatley, D. Keltner, and J. M. Jenkins, *Understanding Emotions*, 2nd ed., Hoboken, NJ: Wiley-Blackwell, 2006.

[2]S. F. Taylor and I. Liberzon, "Neural correlates of emotion regulation in psychopathology," *Trends Cogn. Sci.*, vol. 11, no. 10, pp. 413–418, 2007, https://doi.org/10.1016/j.tics.2007.08.006.

[3]A. Majid, "Current emotion research in the language sciences," *Emotion Rev.*, vol. 4, no. 4, pp. 432–443, 2012, https://doi.org/10.1177/1754073912445827.

[4]J. T. Cacioppo, G. G. Berntson, J. T. Larsen, K. M. Poehlmann, and T. A. Ito, "The psychophysiology of emotion," in *Handbook of Emotions*, R. J. Lewis and J. M. Haviland-Jones, Eds., 2nd ed. New York: Guilford, 2000, pp. 173–191.

[5]H. Jazaieri, H. L. Urry, and J. J. Gross, "Affective disturbance and psychopathology: An emotion regulation perspective," *J. Exp. Psychopathol.*, vol. 4, no. 5, pp. 584–599, 2013, https://doi.org/10.5127/jep.030312.

[6]P. Ekman, Ed., *Darwin and Facial Expression: A Century of Research in Review*, Cambridge, MA: Ishk, 2006.

[7]L. F. Barrett, "Was Darwin wrong about emotional expressions?," *Curr. Dir. Psycholog. Sci.*, vol. 20, no. 6), pp. 400–406, 2011, https://doi.org/10.1177/0963721411429125.

[8]L. F. Barrett, "Are emotions natural kinds?," *Perspect. Psycholog. Sci.*, vol. 1, no. 1, pp. 28–58, https://doi.org/10.1111/j.1745-6916.2006.00003.x

[9]J. E. LeDoux, "Thoughtful feelings," *Curr. Biol.*, vol. 30, no. 11, pp. R619–R623, 2020, https://doi.org/10.1016/j.cub.2020.04.012.

[10]V. K. Kool and R. Agrawal, *Psychology of Technology*, Cham, Switzerland: Springer International Publishing, 2016.

[11]E. O. Wilson, *Sociobiology: The new synthesis*, Cambridge, MA: Harvard University Press, 2000.

[12]D. Lama and P. Ekman, *Emotional Awareness: Overcoming the Obstacles to Psychological Balance and Compassion*, New York: Macmillan, 2008.

[13]P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Palo Alto: Consulting Psychologists Press, 1978.

[14]W. J. Ickes, Ed., *Empathic Accuracy*, New York: Guilford Press, 1997.

[15]S. J. Haggbloom, R. Warnick, J. E. Warnick, V. K. Jones, G. L. Yarbrough, et al., "The 100 most eminent psychologists of the 20th century," *Rev. Gen. Psychol.*, vol. 6, no. 2, pp. 139–152, 2002, https://doi.org/10.1037/1089-2680.6.2.139.

[16]D. Matsumoto and H. S. Hwang, "Reading facial expressions of emotion," *Psychol. Sci. Agenda*, May 2011, https://www.apa.org/science/about/psa/2011/05/facial-expressions.

[17]L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychol. Sci. Publ. Inter.*, vol. 20, pp. 1–68, 2019, https://doi.org/10.1177/1529100619832930.

[18]M. B. Harms, A. Martin, and G. L. Wallace, "Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies," *Neuropsychol. Rev.*, vol. 20, no. 3, pp. 290–322, 2010, https://doi.org/10.1007/s11065-010-9138-6.

[19]H. H. Ip, S. W. Wong, D. F. Chan, J. Byrne, C. Li, et al., "Enhance emotional and social adaptation skills for children with autism spectrum disorder: A virtual reality enabled approach," *Comput. Educat.*, vol. 117, pp. 1–15, 2018, https://doi.org/10.1016/j.compedu.2017.09.010.

[20]K. M. Ford and P. J. Hayes, "On computation wings: Rethinking the goals of artificial intelligence," *Scientific Amer.*, vol. 9, no. 4, pp. 78–83, 1998.

[21]E. Park, B. J. Yun, Y. S. Min, Y. S. Lee, S. J. Moon, et al., "Effects of a mixed reality-based cognitive training system compared to a conventional computer-assisted cognitive training system on mild cognitive impairment: a pilot study," *Cogn. Behav. Neurol.*, vol. 32, no. 3, pp. 172–178, 2019, https://doi.org/10.1097/WNN.0000000000000197.

[22]Stanford Medicine. The Autism Glass Project. http://autismglass.stanford.edu/ (accessed Jul. 1, 2020).

[23]Brain Power. https://brain-power.com/ (accessed Jul. 1, 2020).

[24]J. Quarles, S. Lampotang, I. Fischler, P. Fishwick, and B. Lok, "Scaffolded learning with mixed reality," *Comput. Graphics*, vol. 33, no. 1, pp. 34–46, 2009, https://doi.org/10.1016/j.cag.2008.11.005.

[25]K. Crawford, R. Dobbe, T. Dryer, G. Fried, B. Green, E. Kaziunas, et al., "AI Now 2019 report," New York: AI Now Institute, 2019, https://ainowinstitute.org/AI_Now_2019_Report.pdf.

[26] A. C. Samson, O. Huber, and J. J. Gross, "Emotion regulation in Asperger's syndrome and high-functioning autism," *Emotion*, vol. 12, no. 4, pp. 659–665, 2012, https://doi.org/10.1037/a0027975.

[27]A. Abbe and S. E. Brandon, "The role of rapport in investigative interviewing: A review," *J. Investigat. Psychol. Offend. Profil.*, vol. 10, no. 3, pp. 237–249, 2013, https://doi.org/10.1002/jip.1386.

[28]K. Johnsen, R. Dickerson, A. Raij, C. Harrison, B. Lok, et al., "Evolving an immersive medical communication skills trainer," *Presence*, vol. 15, no. 1, pp. 33–46, 2006, https://doi.org/10.1162/pres.2006.15.1.33.

[29]J. E. Charny, "Psychosomatic manifestations of rapport in psychotherapy," *Psychosomat. Med.*, vol. 28, no. 4, pp. 305–315, 1966, https://doi.org/10.1097/00006842-196607000-00002.

[30]R. B. Grossman and H. Tager-Flusberg, "Reading faces for information about words and emotions in adolescents with autism," *Res. Autism Spectr. Disord.*, vol. 2, no. 4, pp. 681–695, 2008, https://doi.org/10.1016/j.rasd.2008.02.004.

[31]C. Crowell, B. Sayis, J. P. Benitez, and N. Pares, "Mixed reality, full-body interactive experience to encourage social initiation for autism: Comparison with a control nondigital intervention," *Cyberpsychol. Behav. Soc. Network.*, vol. 23, no. 1, pp. 5–9, 2020, https://doi.org/10.1089/cyber.2019.0115.

[32]C. H. Hjortsjö, *Man's Face and Mimic Language*, Lund, Sweden: Studentlitteratur, 1969.

[33]P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System: The Manual on CD ROM*, Salt Lake City: A Human Face, 2002.

[34]"9. Raycasting," in *3D Physics*, Unity3D tutorial, last updated Feb. 26, 2020, https://unity3d.com/learn/tutorials/topics/physics/raycasting.

[35]Quaternions, Unity3D tutorial, last updated Jul. 1, 2019, https://unity3d.com/learn/tutorials/topics/scripting/quaternions.

[36]S. Ioannou, V. Gallese, and A. Merla, "Thermal infrared imaging in psychophysiology: Potentialities and limits," *Psychophysiol.*, vol. 51, no. 10, pp. 951–963, 2014, https://doi.org/10.1111/psyp.12243.

[37]J. L. Kleberg, C. Hanqvist, E. Serlachius, and J. Högström, "Pupil dilation to emotional expressions in adolescent social anxiety disorder is related to treatment outcome," *J. Anxiety Disord.*, vol. 65, pp. 26–33, 2019, https://doi.org/10.1016/j.janxdis.2019.04.006.

[38]G. L. Clore and J. R. Huntsinger, "How emotions inform judgment and regulate thought," *Trends Cogn. Sci.*, vol. 11, no. 9, pp. 393–399, 2007, https://doi.org/10.1016/j.tics.2007.08.005.

[39]D. R. Proffitt, M. Bhalla, R. Gossweiler, and J. Midgett, "Perceiving geographical slant," *Psychon. Bull. Rev.*, vol. 2, no. 4, pp. 409–428, 1995, https://doi.org/10.3758/BF03210980.

[40]M. E. Kret, "Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception," *Front. Psychol.*, vol. 6, article 711, pp. 1–10, 2015, https://doi.org/10.3389/fpsyg.2015.00711.

[41]Paul Ekman Group. "Deception detection." https://www.paulekman.com/deception/deception-detection/ (accessed Jul. 1, 2020).

[42]M. D. Happel, J. A. Spitaletta, E. A. Pohlmeyer, G. M. Hwang, A. M. Greenberg, et al., "National security and the assessment of individual credibility: Current challenges, future opportunities," *Johns Hopkins APL Tech. Dig.*, vol. 33, no. 4, pp. 303–313, 2017.

[43]P. Ekman and M. O'Sullivan, "Who can catch a liar?" *Amer. Psychologist*, vol. 46, no. 9, pp. 913–920, 1991, https://doi.org/10.1037/0003-066X.46.9.913.

[44]C. Neubauer, G. Gremillion, B. S. Perelman, C. La Fleur, J. S. Metcalfe, and K. E. Schaefer, "Analysis of facial expressions explain affective state and trust-based decisions during interaction with autonomy," in *Proc. Int. Conf. Intell. Hum. Syst. Integr.*, February 2020, Springer, Cham, pp. 999–1006.

[45]J. Daniels, J. N. Schwartz, C. Voss, N. Haber, A. Fazel, et al., "Exploratory study examining the at-home feasibility of a wearable tool for social-affective learning in children with autism," *NPJ Digit. Med.*, vol. 1, no. 1, pp. 1–10, 2018, https://doi.org/10.1038/s41746-018-0035-3.

[46]T. W. Frazier, K. R. Ratliff, C. Gruber, Y. Zhang, P. A. Law, and J. N. Constantino, "Confirmatory factor analytic structure and measurement invariance of quantitative autistic traits measured by the Social Responsiveness Scale-2," *Autism*, vol. 18, no. 1, pp. 31–44, 2014, https://doi.org/10.1177/1362361313500382.

**Ariel M. Greenberg,** Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Ariel M. Greenberg is a senior research scientist in APL's Intelligent Systems Center; assistant editor-in-chief of the *Johns Hopkins APL Technical Digest*; steering committee co-chair of the International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation (http://sbp-brims.org/); and an instructor in APL's Strategic Education Program. He received degrees in biology and electrical engineering from University of Maryland, College Park. In the spirit of cells-to-societies transdisciplinary research, Ariel's work addresses topics in psychophysiology and computational social science, toward applications in defense, intelligence, and health. His e-mail address is ariel.greenberg@jhuapl.edu.

**Jason A. Spitaletta,** Asymmetric Operations Sector, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

Jason A. Spitaletta is a major in the US Marine Corps Reserve and a researcher in APL's Asymmetric Operations Sector. He holds a bachelor's degree in biochemistry from Franklin & Marshall College, a master's degree in human factors from Embry-Riddle Aeronautical University, and a master's degree and a PhD in applied experimental psychology from the Catholic University of America. He also holds a graduate certificate from Stanford University's Summer Institute for Political Psychology. His primary research experience is in applied, experimental, and political psychology and cognitive neuroscience. Jason has operational experience in psychological operations (PSYOP) and intelligence assignments in the US Marine Corps as well as the joint and special operations communities. He has deployed to the Western Pacific, Iraq, and Uganda. He is an adjunct faculty member at the National Intelligence University and the Catholic University of America. His email address is jason.spitaletta@jhuapl.edu.