Researchers at the Advanced Digital Sciences Center (ADSC) have demonstrated the ability to determine the direction of origin of a sound source in three dimensions, in real time, using a miniature microphone array (Figure 1). Whereas state-of-the-art techniques that use large microphone arrays for good accuracy have difficulty achieving real-time performance, ADSC’s approach determines the direction of audio to within a $2^\circ\times2^\circ$ spherical surface area in real time, which is a leap forward in the state of the art in acoustic localization. A video showing the system in operation is available on the ADSC web site, http://adsc.illinois.edu.

This work is part of ADSC’s broader program to provide low-cost, high-quality, real-time telepresence for activities such as virtual meetings. Realistic teleimmersion must consider human hearing perceptual issues. For example, suppose that Alice sits in a coffee shop while she attends a virtual meeting. The restaurant noise and other patrons’ voices will overlap with Alice’s speech, which will significantly affect the other attendees’ hearing perception. Further, human hearing is extremely sensitive to the perceived direction of arrival of sounds. If there are small differences between the expected and perceived directions of arrival, so that the voice of the colleague on Alice’s (virtual) left seems to be coming from another direction, the illusion of co-presence will be broken. Thus, clean capture of acoustic sources, accurate estimation of their direction of arrival, and appropriate reconstruction of their virtual direction for each virtual listener are critical needs for immersive communications.

Multichannel speech enhancement and acoustic sound direction finding are classic problems in array signal processing, and have been extensively studied with arrays of physically separated microphones. The practical uses of these systems are limited because of the need for many microphones, which occupy a lot of space. Further, these systems require a large separation between microphones in order to detect low frequency sources accurately. Overall, state-of-the-art algorithms have limited ability to perform speech enhancement or 3D audio direction finding and reconstruction for small microphone arrays and for complicated situations such as a busy restaurant with multiple acoustic sources and highly reverberant spaces. Even with GPU assistance, there are no real-time algorithms for combined processing of speech enhancement, 3D audio direction finding, and reconstruction. Thus realistic teleimmersion requires a major advance in the development and implementation of microphone-array-based speech enhancement, 3D audio direction finding and reconstruction. Given all the audiovisual challenges, it is not surprising that virtual reality is one of the US National Academy of Engineering’s 14 grand challenges for engineering in the 21st century [NAE11].

Details of ADSC’s Contribution

Realistic teleimmersion will never be widely deployed if each user must have a large physical array of microphones. Thus ADSC’s audio work employs a biologically inspired miniature microphone array [MLKJ08], shown in Figures 1 and 2. This ‘zero-aperture’ microphone array uses four precisely placed microphones, and is the smallest microphone array sensor in the world. Each microphone measures only a few millimeters across, and is placed about a centimeter apart from the others. This XYZO array combines three directional gradient response microphones, each pointing in a different direction (X, Y, Z), with an omni-directional response microphone (O); together the array captures more acoustic information than traditional arrays composed only of omni-directional microphones.

The algorithms developed for the XYZO microphone array leverage the increased acoustic information obtainable from a miniature directional array, and employ a wide variety of signal processing techniques not feasible in traditional physically separated microphone arrays. Unlike a conventional spatially separated microphone array that uses the time difference of each microphone...
pair to determine the direction of a sound, the XYZO array employs the amplitude difference between the microphones. This makes the XYZO array less likely to be affected by frequency variations of target sources. But the localization mechanism behind the XYZO array is similar to the spatially separated microphone array in the sense that the array responses are highly dependent on the direction of arrival of sources. With understanding of this dependence, obtained either through experimental measurements or computation, conventional localizers such as Steered Response Power (SRP) and Multiple Signal Classification (MUSIC) can then be adopted for searching for the directional peaks.

For determining the direction of speakers in realistic teleimmersive scenarios, ADSC has demonstrated real-time 2D and 3D direction finding algorithms for the XYZO miniature microphone array, using an ordinary laptop, a commodity graphics card, and a general-purpose graphics processing unit (GPGPU). In addition to running in real time, the new hardware and software is simultaneously smaller in size, lower power, less expensive, more accurate, and more robust than conventional approaches based on a much larger microphone array. The algorithms developed for the XYZO microphone array leverage the increased acoustic information obtainable from a miniature directional array, and employ a wide variety of signal processing techniques not feasible in traditional physically separated microphone arrays. This work will be submitted for publication later this year.

To further improve the user experience and the user’s sense of being physically in the same location during an immersive teleconference, ADSC has developed a novel 3D audio capture and reconstruction algorithm for stereo headphones, based on the XYZO array. The approach produces the left and right ear signals from the microphone array outputs by multiplying a set of optimal time-invariant gain vectors. These vectors are derived based on the minimum-variance-distortionless-response (MVDR) beamformer and minimum-mean-squared-error (MMSE) estimation. These gain vectors integrate the two stages of beamforming and head related transfer function (HRTF) filtering and can be easily computed offline. This approach is independent of the number of virtual sound sources and is flexible for working with different sets of head related transfer functions. In subjective user tests, a wide variety of subjects reported good localization accuracy [SRJJ12].
Figure 3 shows the miniature microphone array system at work. Figure 4 shows the results from directional tracking of a military vehicle in an open field trial, using the XYZO array.

About Us
The Advanced Digital Sciences Center (ADSC), established in 2009 and located in Singapore, is a wholly-owned subsidiary of the University of Illinois at Urbana-Champaign. ADSC’s research program in interactive digital media is funded by the Agency for Science, Technology, and Research (A*STAR).

Douglas L. Jones is a professor in the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. He is a Fellow of the IEEE and served as a Member-at-Large of the IEEE Signal Processing Society Board of Governors during 2002-2005. He is an author of the laboratory textbooks *A Digital Signal Processing Laboratory Using the TMS32010* and *DSP Laboratory with TI TMS320C54x*. Doug received his Ph.D. from Rice University in 1987.

Shengkui Zhao is a postdoctoral fellow at ADSC. Before joining ADSC, he worked as a software developer at Radix Pte Ltd, Singapore, and as a project officer and then a research fellow in the School of Computer Engineering at Nanyang Technical University, Singapore. Shengkui received his Ph.D. from the Nanyang Technological University in 2009.

References