Acoustic levitation for multimodal volumetric display

Hirayama, Ryuji, Plasencia, Diego, Masuda, Nobuyuki, Subramanian, Sriram

Ryuji Hirayama, Diego Martinez Plasencia, Nobuyuki Masuda, Sriram Subramanian, "Acoustic levitation for multimodal volumetric display," Proc. SPIE 11463, Optical Trapping and Optical Micromanipulation XVII, 114630Q (20 August 2020); doi: 10.1117/12.2569328

Event: SPIE Nanoscience + Engineering, 2020, Online Only

Acoustic Levitation for Multimodal Volumetric Display

Ryuji Hirayama*[a], Diego Martinez Plasencia[a], Nobuyuki Masuda[b], Sriram Subramanian[a]
[a]School of Engineering and Informatics, University of Sussex, Falmer, Brighton, BN1 9QJ, UK;
[b]Department of Applied Electronics, Tokyo University of Science, 6-3-1, Niijuku, Katsushika-ku, Tokyo 125-8585, Japan

ABSTRACT

Current display approaches, such as VR, allow us to get a glimpse of multimodal 3D experiences, but users need to wear headsets as well as other devices in order to trick our brains into believing that the content we are seeing, hearing or feeling is real. Light-field, holographic or volumetric displays avoid the use of headsets, but they constraint the user's ability to interact with them (e.g. content is not reachable to user's hands, user's constrained to specific locations) and, most importantly, still cannot simultaneously deliver sound and touch. In this talk, we will present the Multimodal Acoustic Trapping Display (MATD): a mid-air volumetric display that can simultaneously deliver visual, tactile and audio content, using phased arrays of ultrasound transducers. The MATD makes use of ultrasound to trap, quickly move and colour a small particle in mid-air, to create coloured volumetric shapes visible to our naked eyes. Making use of the pressure delivered by the ultrasound waves, the MATD can also create points of high pressure that our bare hands can feel and induce air vibrations that create audible sound. The system demonstrates particle speeds of up to 8.75 m/s and 3.75 m/s in the vertical and horizontal directions, respectively. In addition, our technique offers opportunities for non-contact, high-speed manipulation of matter, with applications in computational fabrication and biomedicine.

Keywords: Acoustic levitation, Volumetric displays, Haptics sensation, Parametric audio

1. INTRODUCTION

Science-fiction movies portray volumetric systems that provide not only visual but also tactile and audible 3D content. Current display approaches, such as VR, allow us to get a glimpse of such 3D experiences. However, users need to wear headsets as well as other devices (e.g. gloves and headphones) in order to trick our brains into believing that the content we are seeing, hearing or feeling is real.

To create 3D visual content without the need for such devices, researchers have been proposing variety of 3D displays, classified often into three families: ray displays (light-field), wave displays (holographic) and point displays (volumetric) [1]. Light-field and holographic displays rely on a 2D display modulator (screen), constraining the visibility of 3D content to the volume between the observer's eyes and the screen (that is, the direct line of sight). On the other hand, volumetric approaches are based on light-scattering, -emitting or -absorbing surfaces and offer unconstrained visibility anywhere around the display. Some of those volumetric displays operate in air with no barrier between user and image (thus referred to as free-space displays) and can be created using dusts [2], plasma [3, 4], mirage [5] and scanning particles using light [6] and electromagnetic field [7]. However, none of these approaches rely on operating principles that can produce tactile and audio content as well.

In this paper, we present a Multimodal Acoustic Trapping Display (MATD): a particle-scanning-based volumetric display using acoustic levitation that can simultaneously deliver tactile and audio content [8]. Acoustic levitation displays reported before the MATD [9-12] have only demonstrated control of a reduced number of points at reduced speeds and do not involve touch or audible sound. By contrast, our MATD makes use of ultrasound to trap, quickly move and colour a small particle in mid-air, to create coloured volumetric shapes visible to our naked eyes. Making use of the pressure delivered by the ultrasound waves, the MATD can also create points of high pressure that our bare hands can feel and induce air vibrations that create audible sound. In other words, the MATD can create multimodal 3D content that we can see, feel and hear using acoustophoresis as the single operating principle.

---

*the current email address: ryuji.hirayama@ucl.ac.uk

Optical Trapping and Optical Micromanipulation XVII, edited by Kishan Dholakia,
Gabriel C. Spalding, Proc. of SPIE Vol. 11463, 114630Q · © 2020 SPIE
CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2569328

2. OPERATING PRINCIPLES OF THE MATD

2.1 System overview

The MATD system is composed of two opposed arrays of 16 × 16 transducers, aligned on top of each other and with a separation of 24 cm (Fig. 1). We used Murata MA40S4S transducers (40 kHz, 1 cm diameter) for the two arrays and high-intensity RGB LEDs (OptoSupply, OSTCWBTHC1S) to illuminate a levitated particle. A Waveshare CoreEP4CE6 FPGA board was used to receive updates from the CPU (3D position, RGB colour, phase and amplitude) and controls the ultrasound transducers individually and the illumination LEDs.

The MATD (illustrated in Fig. 1) analytically computes single twin traps (for levitation) or focusing points (for haptics) at a hardware level using the FPGA. This enables position and amplitude updates of the trap in a volume of 10 × 10 × 10 cm^3, at a rate limited only by the transducer frequency (i.e. 40 kHz). The hardware-embedded computation of the twin trap provides controlled and fast levitation of our scanning particle (1-mm-radius, white expanded polystyrene bead) and is synchronized with the RGB LEDs, allowing the creation of a POV (persistence of vision) display with accurate control of the perceived colour. Mid-air tactile feedback at controlled locations (for example, the user's hand) is created by using a secondary focusing trap and custom multiplexing policy. The system can also produce audible sound from the ultrasound transducers using amplitude demodulation in mid-air.

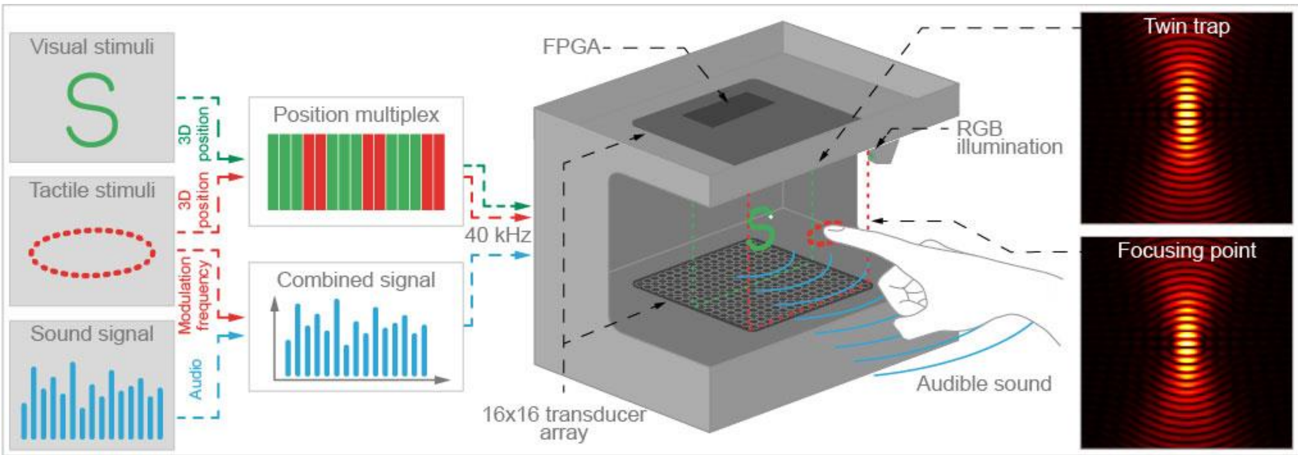

Figure 1. A geometrical description of the MATD, illustrating that visual and tactile stimuli, along with sound are used as input to create multimodal content (from [8]).

2.2 Driving parameters

Transducer Operation: The transducers were driven using a 12 V peak-to-peak square wave signal at 40 kHz, producing a sinusoidal output owing to their narrowband response. Phase delays were implemented by temporal shifting of the 40-kHz square wave, whereas amplitude control was achieved by reducing the duty cycle of the square wave. The complex amplitude of the transducers do not vary linearly with duty cycle. The relationship between the duty cycle and the amplitude of the transducers (A_t) can be represented as:

$$A_t = \sqrt{\sin^2\left(\frac{duty}{100}\pi\right)} \tag{1}$$

We stored this function as a look-up table in the FPGA (to map the amplitude to the duty cycle) for efficient computation of the updates at the required rate (40 kHz). This resulted in a modulator providing 64 levels of phase (resolution of π/32 rad) and 33 levels of amplitude resolution.

Computation of traps: The computation of the focus points and twin levitation traps is embedded into the FPGA. For a focus point at position $\boldsymbol{p}$ and with phase ϕ_p, the phase of each transducer (ϕ_t) was discretized as follows:

$$\phi_t = \left(-\frac{32}{\pi}k|\boldsymbol{p} - \boldsymbol{p_t}| + \phi_p\right) \bmod 64. \tag{2}$$

where k represents the wave number for the frequency used, $\boldsymbol{p_t}$ represents the position of each transducer. The twin traps were computed by combining a high-intensity focus point (as in Eq. 2) and a levitation signature [13]. The levitation signature was implemented by adding a phase delay of π rad to the transducers in the top array, as used in [12], thus producing traps that maximize vertical forces. The transducer positions and discretized phase delays relative to the distance were stored in two look-up tables in the FPGA, simplifying the computation of the focus point and levitation twin traps.

Illumination control: We used one illumination module equipped with the high-intensity RGB LEDs, which was placed on the top right corner of our MATD prototype. Intensity of the LEDs were controlled by the FPGA using pulse width modulation. The resulting perceived luminance of the particle for an observer around the MATD can be analytically approximated using the definition of the bidirectional reflectance distribution function (BRDF). The white and diffuse surface of our particle enables us to approximate its BRDF as a Lambertian surface.

2.3 Content creation

Visual content: Human eyes can integrate different light stimuli under a single percept (that is, a single shape or geometry) during short periods of time (0.1 s is usually accepted as a conservative estimation, even in bright environments [14]), and thus our particle needs to scan the content in less than this time. Our parameters allow us to determine feasible paths (particle speed, acceleration and curvature within the limits identified), which can be revealed in less than 0.1 s by exploiting only a fraction of the display's capabilities.

Tactile content: Well differentiated tactile feedback was delivered using only a 25% duty cycle for tactile content. Thus, 75% of the cycles could still be used to position the primary trap, and the tactile content resulted in minimum loss of scanning speed. For our experiments, we chose a modulation frequency of 250 Hz, avoiding the primary range of human auditive perception [15] (2 kHz–5 kHz) to minimize parasitic noise, but remaining well within the optimum perceptual threshold of skin Lamellar corpuscles for vibration [16]. The 10-kHz update rate for tactile stimulation is sufficient for spatio-temporal multiplexing strategies to maximize the fidelity of mid-air tactile content [17].

Audio content: Audible sound is created by ultrasound demodulation using upper-sideband amplitude modulation [18] of the traps. Our sampling at 40 kHz encodes most of the auditive spectrum (44.1 kHz), and the high-power transducer array produces audible sound even for a relatively small modulation index (a = 0.2) while still modulating particle positions and tactile points at the 40-kHz rate. For simultaneous auditive and tactile stimulation, we combine the 40-kHz multifrequency audio signal with the tactile modulation signal (250 Hz), maintaining the sampling frequency of the individual signals and reducing losses in audio quality.

3. PERFORMANCE OF THE MATD

3.1 MATD's multimodal content

Figure 2 shows examples of 2D (Figs. 2a and 2b) and 3D (Fig. 2c) vector content that the MATD can create in mid-air. These simple shapes can be scanned in 0.1 s and thus can be seen as solid POV images for human eyes, even under conventional indoor illumination conditions. Our volumetric content showed no substantial flicker and good colour reproduction, independently of the viewer's location. Figure 2d shows three examples of colour tests performed with vector images (numbers, as in a seven-segment display), showing good colour saturation. Brighter images can be obtained by adding extra illumination modules or more powerful LEDs. Also the examples of visual content in Fig. 2d were created with simultaneous audible sounds of 60 dB (voice pronouncing the numbers).

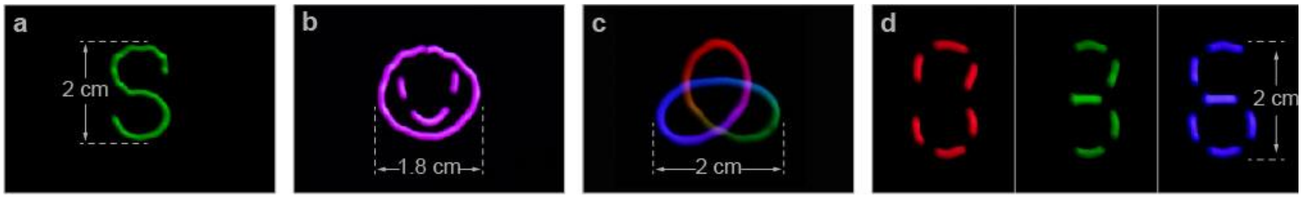

Figure 2. Example POV images. 2D vector content of (a) a letter "S" and (b) smiley face. (c) A multicolour 3D vector content of torus knot (3:2). (d) 2D vector content of 7-segment numbers ("0", "3" and "6") with audio sound (from [8]).

Figure 3 shows more complicated visual content: 2D rasterized content (Fig. 3a), 3D vector content (Fig. 3b) and 3D rasterized content (Fig. 3c). These shapes cannot be scanned in the POV time (0.1 s) by the current MATD system.

Therefore, these photos were taken with long exposure time (20 s). The 2D rasterized image in Fig. 3a were created with simultaneous tactile stimuli, which was strong enough to be felt by our hands.

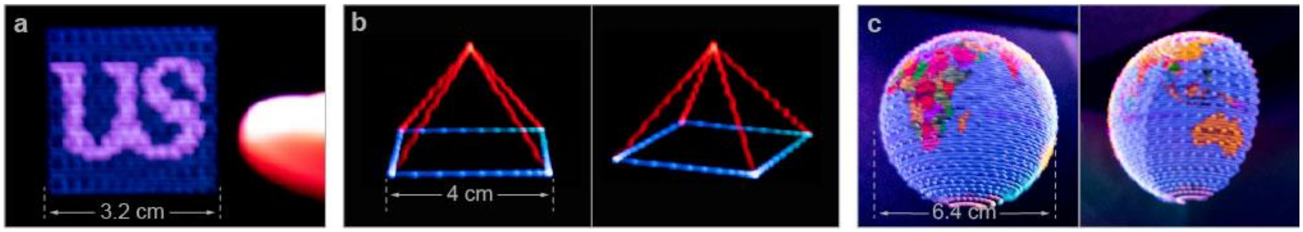

Figure 3. Long-exposure (20 s) shots of visual content: (a) 2D rasterized image of the "US" logo with simultaneous tactile stimuli, (b) 3D vector image of a "pyramid" and (c) 3D rasterized image of a "globe" (from [8]).

3.2 Quantitative evaluations of the MATD

Manipulation speed: We performed linear speed tests to characterize the maximum displacement speeds for each of our three experimental conditions for particles moving along horizontal and vertical directions. Linear paths of 10 cm were used for these tests, with the particles starting at 5 cm to the left and stopping at 5 cm to the right of the centre of the MATD (for the vertical tests, 5 cm above or below the centre). Particles started at rest and were constantly accelerated to reach the maximum speed at the centre of the array. They were then constantly decelerated until brought back to rest at a position 10 cm away from the starting position. When exploring the potential maximum linear speeds, we followed a bisection method. We performed 10 tests at each velocity, and considered a test to be successful only if 9/10 repetitions were successful and report only the highest successful speed observed. The same test procedure was used in all subsequent experiments (that is, acceleration, radius of curvature, and corner speeds).

Our tests revealed high scanning speeds and accelerations, much higher than those of optical [6] or acoustic [12] setups demonstrated until now. The most critical display parameters are summarized in Table 1 according to the various modes of operation of the MATD: single particle with no amplitude modulation (visual content only), single particle with minimum amplitude (in the worst-case scenario, displaying visual and audio content) and time-multiplexed dual trap with minimum amplitude (in the worst-case scenario, delivering all visual, audio and tactile content). Table 1 provides the maximum displacement parameters along the horizontal direction (in the worst-case scenario, with weaker trapping forces) as conservative reference values that allow content reproduction independently of the particle direction. The system demonstrates particle speeds of up to 8.75 m/s in the vertical directions when providing only visual content.

Table 1. Main parameters of the MATD.

	Visual only	**Visual and audio**	**Visual, audio and tactile**
Highest speed recorded (m/s)	3.75	3.375	2.5
Highest acceleration recorded (m/s^2)	141	122	62
Highest speed for corner features (m/s)	0.75	0.5	0.375
Highest image frame rate until now (Hz)	12.5	10.0	10.0
Colour (bpp)	24	24	24

We used a static camera (CANON, EOS 750D) placed 12 cm in front of the MATD and removed all light. We used a long-exposure shot to record our trials and used our RGB illumination system to illuminate (that is, colour-code) the evolution of the particle along its path at steps of 1 ms. Figure 4a shows the results for particles travelling along the horizontal direction. In the top panel, the solid black lines represent the speed of the levitation trap, and the coloured lines show examples of actual particle velocities, as captured during the tests. The path observed in Fig. 4a shows the expected correlations between particle velocities (top), particle-to-trap distances (middle) and accelerations (bottom). Points of zero Δp (that is, no net force being applied to the particle) correspond to the maximum/minimum points in each velocity plot (that is, derivative equal to zero), and the sign of Δp is aligned with the monotonicity of velocity plots, increasing when Δp is negative and decreasing otherwise. Similar correlations can be observed between the Δp (middle) and acceleration plots (bottom). Accelerations remain positive when Δp is negative and vice versa (that is, the trap acts as a restorative force), and the prominent features in both plots match well. It is worth noting that the particle almost always remained within a few millimetres of the place where the actual levitation trap was placed (*Δp*), being subjected to high acceleration rates. This observation is important to understand the behavior of the MATD.

Tactile generation and quality: Our measuring setup comprised a modified 3D printer (OpenBuilds Sphinx 55), in which the extruder had been removed and replaced by a calibrated Brüel & Kjær 4138-A-015 microphone connected to a PicoScope 4262 oscilloscope. We measured the sound pressure level (SPL, in decibels) generated by our MATD system for a single tactile point at the centre of the array using the multiplexing schedule.

Figure 4d shows the measured SPL when all visual, tactile and auditive content is present. This condition used a primary trap for delivering visual content and audio with a 2-kHz signal combined with a 250-Hz tactile signal. The result shows accurate positioning and focusing of the tactile points and sound pressure levels greater than 150 dB, well above the threshold of 72 dB required for tactile stimulation [19].

Audio generation and quality: We explored the quality of the audio generated by the MATD. The audio signal used in the tests was a chirp signal with frequency increasing quadratically from 100 Hz to 20 kHz. We trapped one particle and used our chirp audio signal to modulate the amplitude of our transducers. We recorded the sound generated with an Audio-Technica PRO35 microphone (the spectrogram of the recorded sound is shown in Fig. 4c), revealing accurate representation of the input signal with some degradation due to harmonics.

The use of position multiplexing cannot be avoided if simultaneous tactile and audio-visual content is to be delivered. Position multiplexing introduces frequency aliasing at the 10-kHz multiplexing rate (as well as harmonic frequencies) as a result of acoustic pressure being focalized at different locations. Our tests show that our multiplexing approach reduces audible artefacts when compared to the use of both amplitude and position multiplexing, particularly for harmonics, and that our approach minimizes artefacts in the human primary auditory range [8].

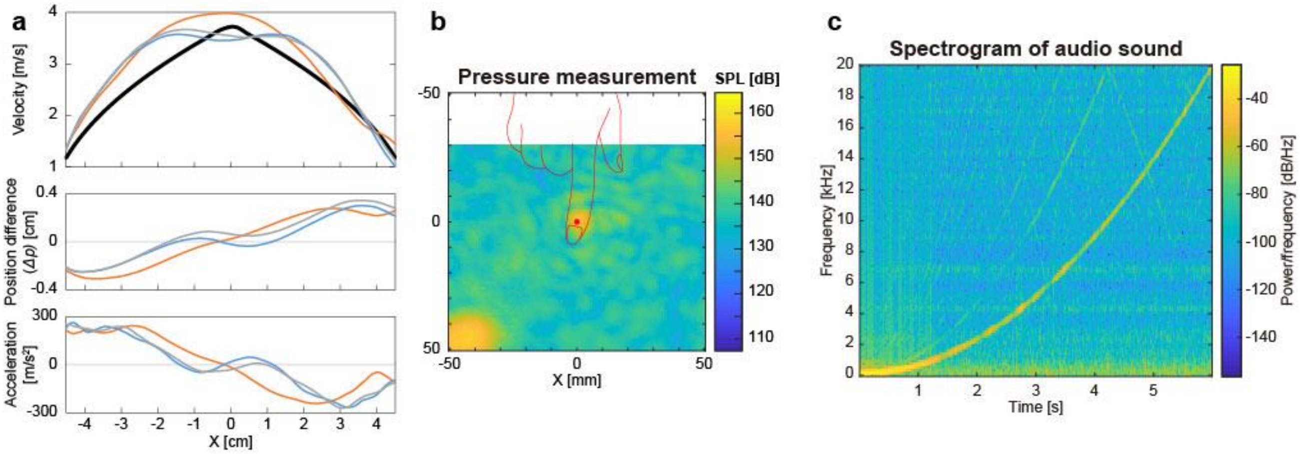

Figure 4. Quantitative evaluation of the MATD, summarized from [8]. (a) Plots of the linear speed test. (b) Measurement of the sound pressure level. (c) Spectrogram of the audio sound that the MATD creates.

4. DISCUSSION

Our instantiation of the MATD presented here was created using low-cost, commercially available components, making it easy to reproduce but also introducing limitations. Our tests were performed at a transducer voltage allowing continued usage (12 V peak to peak). Tests at higher voltages (15 V peak to peak, duration <1 h) indicate that increasing the transducer power can result in better performance parameters (for example, maximum horizontal speed of 4 m/s) and more complex content. Increased power would also allow operation of the MATD at a 50% duty cycle, further reducing audio artefacts. Similarly, transducers operating at higher frequencies (that is, 80 kHz) can also improve audio quality and, combined with a reduced transducer pitch, would improve the spatial resolution of the levitation traps (more accurate paths of the scanning particle). The MATD demonstrated the possibility to manipulate particles by retaining them in a dynamic equilibrium (rather than a static one, as most other levitation approaches), enabling the high accelerations and speeds observed. The use of models that accurately predict the dynamics of the particle (that is, in terms of acoustic forces, drag, gravity and centrifugal forces, but also considering interference from secondary traps and transient effects in the transducer phase updates) would allow better exploitation of the observed maximum speeds and accelerations, enabling larger and more complex visual content. Alternatively, they could enable a more efficient use of the acoustic pressure, providing similar speeds and accelerations to those of the MATD, but allocating a lower duty cycle to the primary trap. This power could then be dedicated to achieving stronger tactile content or supporting a greater number of simultaneous traps.

Our study demonstrates an approach to creating volumetric POV displays with simultaneous delivery of auditive and tactile feedback and with capabilities that exceed those of alternative optical approaches [6]. Polarization-based photophoretic approaches [20] could match the potential for particle manipulation (that is, speeds and accelerations) demonstrated in this study, but they would still be unable to include sound and touch. The demonstrated MATD prototype hence brings us closer to volumetric displays providing a full sensorial reproduction of virtual content. Beyond opening a new avenue for multimodal 3D displays, our device and techniques enable positioning and amplitude modulation of acoustic traps at the sound-field frequency rate (that is, 40 kHz), providing also an interesting experimental setup for chemistry or laboratory-on-a-chip applications (for example, multi-particle levitation and mode oscillations).

ACKNOWLEDGEMENTS

We acknowledge funding from the EPSRC project EP/N014197/1 ‘User interaction with self-supporting free-form physical objects’, the EU FET-Open project Levitate (grant agreement number 737087), the Royal Academy of Engineering Chairs in Emerging Technology Scheme (CiET1718/14), the Rutherford fellowship scheme, JSPS through Grant-in-Aid number 18J01002 and the Kenjiro Takayanagi Foundation.

REFERENCES

[1] Smalley, D., Poon, T. C., Gao, H., Kvavle, J. and Qaderi, K. “Volumetric displays: turning 3-D inside-out,” Opt. Photonics News 29(6), 26–33 (2018).

[2] Perlin, K. and Han, J. Y. “Volumetric display with dust as the participating medium,” US patent 6,997,558 (2006).

[3] Kimura, H., Uchiyama, T. and Yoshikawa, H. “Laser produced 3D display in the air,” In Proc. ACM SIGGRAPH 2006 Emerging Technologies (2006).

[4] Ochiai, Y. et al. “Fairy lights in femtoseconds: aerial and volumetric graphics rendered by focused femtosecond laser combined with computational holographic fields,” ACM Trans. Graph. 35, 17 (2016).

[5] Ruiz-Avila, J. “Holovect: Holographic Vector Display,” Kickstarter, https://www.kickstarter.com/projects/2029950924/holovect-holographic-vector-display (2016).

[6] Smalley, D. E. et al. “A photophoretic-trap volumetric display,” Nature 553, 486–490 (2018).

[7] Berthelot, J. and Bonod, N. “Free-space micro-graphics with electrically driven levitated light scatterers,” Opt. Lett. 44, 1476–1479 (2019).

[8] Hirayama, R., Martinez-Plasencia, D., Masuda, N. and Subramanian, S. “A volumetric display for visual, tactile and audio presentation using acoustic trapping,” Nature 575, 320–323 (2019).

[9] Ochiai, Y., Hoshi, T. and Rekimoto, J. “Pixie dust: graphics generated by levitated and animated objects in computational acoustic-potential field,” ACM Trans. Graph. 33, 85 (2014).

[10] Sahoo, D. R. et al. JOLED: a mid-air display based on electrostatic rotation of levitated Janus objects. In Proc. of UIST2016, 437–448 (ACM, 2016).

[11] Norasikin, M. A. et al. SoundBender: dynamic acoustic control behind obstacles. In Proc. of UIST 2018, 247–259 (ACM, 2018).

[12] Marzo, A. and Drinkwater, B. W. Holographic acoustic tweezers. Proc. Natl Acad. Sci. USA 116 84–89 (2018).

[13] Marzo, A. et al. Holographic acoustic elements for manipulation of levitated objects. Nat. Commun. 6, 8661 (2015).

[14] Bowen, R. W., Pola, J. and Matin, L. “Visual persistence: effects of flash luminance, duration and energy,” Vision Res. 14 295–303 (1974).

[15] Gelfand, S. A. Essentials of Audiology 3rd edn (Thieme, 2010).

[16] Makous, J. C., Friedman, R. M. & Vierck, C. J. “A critical band filter in touch,” J. Neurosci. 15, 2808–2818 (1995).

[17] Frier, W., Pittera, D., Ablart, D., Obrist, M. and Subramanian, S. “Sampling strategy for ultrasonic mid-air haptics,” In Proc. of the 2019 CHI Conference on Human Factors in Computing Systems 121 (ACM, 2019).

[18] Gan, W. S., Yang, J. and Kamakura, T. “A review of parametric acoustic array in air,” Appl. Acoust. 73, 1211–1219 (2012).

[19] Carter, T., Seah, S. A., Long, B., Drinkwater, B. and Subramanian, S. “UltraHaptics: multi-point mid-air haptic feedback for touch surfaces,” In Proc. of UIST 2013 505–514 (ACM, 2013).

[20] Shvedov, V., Davoyan, A. R., Hnatovsky, C., Engheta, N. and Krolikowski, W. “A long-range polarization-controlled optical tractor beam,” Nat. Photon. 8, 846–850 (2014).