Holographic Whisper: Rendering audible sound spots in three-dimensional space by focusing ultrasonic waves

Yoichi Ochiai [1, 3] *
[1] University of Tsukuba
Ibaraki, Japan
wizard@slis.tsukuba.ac.jp,
1heisuzuki@gmail.com

Takayuki Hoshi [2, 3] *
[2] The University of Tokyo
Tokyo, Japan
star@star.rcast.u-tokyo.ac.jp

Ippei Suzuki [1]
[3] Pixie Dust Technologies, Inc.
Delaware, United States
{ochyai, star}@pixiedusttech.com

ABSTRACT
We propose a novel method of spatial audio rendering using ultrasound. An ultrasonic phased array generates one or more focal points in air, and they act as point sources of audible sound when the ultrasound waves are modulated. Our sound-point loudspeaker has two major advantages over conventional ultrasound-based sound-beam (super-directional) loudspeakers. The higher audience selectivity means that our sound-point loudspeaker can deliver sound to the ears of the target person, whereas a sound-beam loudspeaker delivers sound to not only the target person but also other persons standing in the same direction. The other advantage is lower exposure to ultrasound; while an audible sound beam travels along an ultrasonic beam in a sound-beam loudspeaker, audible sound can be heard along the direction perpendicular to the ultrasonic beam in our sound-point loudspeaker. This paper reports the principles of our sound-point loudspeaker, prototype construction, evaluation, and applications.

Author Keywords
Spatial sound control, Ultrasound, Phased-array focusing, Self-demodulation effect, Point source, Aerial interaction

ACM Classification Keywords
H.5.1. Information interfaces and presentation (e.g., HCI): Multimedia information systems: Audio input/output

INTRODUCTION
Audio information is one of the essential components of multimedia content and real-world interactions. In contrast to light rays, sound waves have lower directivity and are received by listeners who were not targeted by the speaker. Conventional loudspeakers are almost omnidirectional, which is usually an undesirable feature in public places. A technology to provide different audio information to different users is desired in ubiquitous computing (in other words, non-wearable computing). With such a technology, personalized sound delivery becomes possible and sound environments can be designed more freely. Continuous efforts have been made to this end, such as wave field synthesis, directional audible-sound loudspeakers, and ultrasound-based super-directional loudspeakers.

Wave field synthesis [1, 2] is a method to reproduce wave fronts from virtual sound sources with a large number of loudspeakers. The listening space has to be densely surrounded by the loudspeakers because the principle is the Kirchhoff-Helmholtz integral, and so this method is very costly to implement. Besides, this method aims to reproduce a sound environment and it is not suitable for generation of localized sound areas. Rather than this method, we discuss directional and super-directional loudspeakers for our purpose.

Conventional directional loudspeakers are arrays of sub-units (small audible-sound loudspeakers). All the sub-units are driven by an identical signal and, as a result, plane sound waves are radiated. The plane waves of audible sound form a narrow sound beam [3]. Steering of the beam is also possible [4]. Although these loudspeakers produce sound beams in a narrow angle, all the persons standing within that angle receive the audio information. An ultrasound-based super-directional loudspeaker [5-8] utilizes the plane waves of ultrasound as an end-fire array of audible-sound sources, based on the self-demodulation effect of air, to generate a narrower sound beam. In this case, listeners are exposed to high-intensity ultrasound.

Another category of sound manipulation is the production of sound spots in air. An array of loudspeakers generates focal points of audible sound waves by tuning time delays, or phase delays, between the loudspeakers (i.e., phased-array control) [9]. Another method uses multiple ultrasound-based super-directional loudspeakers to generate audible sound at the crossing points of ultrasonic beams [10]. These methods have spatial requirements: The former requires a large device consisting of hundreds of loudspeaker sub-units and the latter requires multiple devices surrounding the listening space.

We introduce a new loudspeaker that generates discrete sound producing spots in air with a single ultrasonic device, that we named "Holographic Whisper." Our proposition is

* Joint first authors.

CHI 2017, May 06 - 11, 2017, Denver, CO, USA

ACM 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025989

	Single channel		Phased array	
	Fixed sound beam	Fixed sound point	Steerable sound beam	Movable sound point
Audible sound	BOSE [3]		JBL [4]	Shinagawa [9]
Ultrasound	Yoneyama [5] Aoki [6] Pompei [7]	Matsui [10]	Shi [12]	This study

Figure 1. Position of this study among related work. Holographic Whisper focuses on ultrasound-based movable sound points (red box). It can produce steerable sound beams (red dotted box) when the focal length divided by the device size is larger than a threshold value.

to apply the phased-array control technique to an array of ultrasonic transducers. In other words, the transducers are driven by different phase-shifted signals individually, instead of being driven by an identical signal. The phase differences are adequately calculated for an intended spatial distribution (e.g., a single or more focal points) of ultrasound after interference. We intend to generate focal points of ultrasound that act as sound point sources.

The primary contribution of this paper is the production of aerial sound point sources with an ultrasonic phased array for three-dimensional (3D) audio interaction. The remainder of this article is organized as follows. First, we introduce related work on loudspeakers and ultrasonic phased arrays. Second, we theoretically describe the principles and control methods of the ultrasonic phased array. Third, we show the implementation and experimental results of the prototype of the proposed focusing loudspeaker. Fourth, we give examples of applications. Finally, we discuss the limitations and scalability. We believe that this study fills a gap between the design spaces of 3D visual images and sound images, and contributes to the progress of human-computer interaction (HCI) research and leads to various new applications.

RELATED WORK

Our ultrasonic loudspeaker renders point sources of audible sound using an ultrasonic phased array. Related technologies are audible-sound directional/spot loud-speakers and ultrasound-based super-directional loud-speakers. Figure 1 shows the positon of our study among these technologies. Phased-array control of ultrasonic waves for other purposes is also an important topic. We introduce these topics as follows.

Audible-sound directional/spot loudspeakers

Directional loudspeakers, especially for public spaces, already exist in the market [3]. Sub-units of a loudspeaker array are driven by an identical audible sound signal, and form a plane wave of audible sound (audible-sound beam). By controlling these sub-units individually, it is possible to steer the beam [4].

Generation of focal points of audible sound waves has been previously reported [9]. A 120-ch loudspeaker array generated focal points of audio signals such as voices and music. A single loudspeaker sub-unit had dimensions of 66 × 107 mm, and the array had dimensions of 1,650 × 1,350 mm. The diameter of a focal point depends on sound frequency, and it was approximately 400 mm for 1 kHz, which was comparable to the size of a human head. Because we aim to realize a similar effect without such a large array, we use ultrasound instead of audible sound.

Ultrasound-based super-directional/spot loudspeakers

Ultrasound enables the creation of loudspeakers that have narrower directivity than audible-sound directional loudspeakers [5-8]. An ultrasound-based super-directional loudspeaker consists of an array of ultrasonic transducers that are driven by an identical signal. The principle consists of two effects. One is the self-demodulation effect, which is the radiation of audible sound from modulated high-intensity ultrasonic waves in air. This means air filled with modulated ultrasonic waves acts as a virtual sound source. The other is the parametric array effect, which is the formation of an audible sound beam. The ultrasonic array radiates an ultrasonic plane wave (a beam of ultrasound). This means that virtual sound sources are aligned along this ultrasonic beam in air. The phases of these virtual sound sources are different because of the propagation of the ultrasonic plane wave. As the result of interference of audible sound waves, a narrow beam of audible sound is generated.

Studies on the steering of audible sound beams from super-directional loudspeakers have been published [11, 12]. To achieve this, the phase delays between the transducers are controlled based on the steering angle, θ [rad]: When the spacing between the transducers is d [m] in the case of a 1D transducer array, the phase delay of the i-th transducer, $\Delta\varphi_i$ [rad], is calculated as follows,

$$\Delta\varphi_i = i \times kd \sin\theta , \qquad (1)$$

where k [rad/m] is the wavenumber of the ultrasound.

Sound spots can be generated by using multiple super-directional loudspeakers [10]. The ultrasonic beams have slightly different frequencies. Audible sound is generated within the crossing point of these ultrasonic beams as a result of beat formation. In this configuration, the target person is necessarily exposed to high-intensity ultrasound. Because we aim to realize a similar effect with lower exposure to ultrasound, we generate a focal point instead of a beam (Fig. 2).

Phased-array control of ultrasonic waves

Recently, ultrasound has been used to produce haptic feedback in air, by using the principle of acoustic radiation pressure, which is a nonlinear effect where high-intensity ultrasound presses the object surface. The phased-array focusing technique is used to achieve high intensity at the

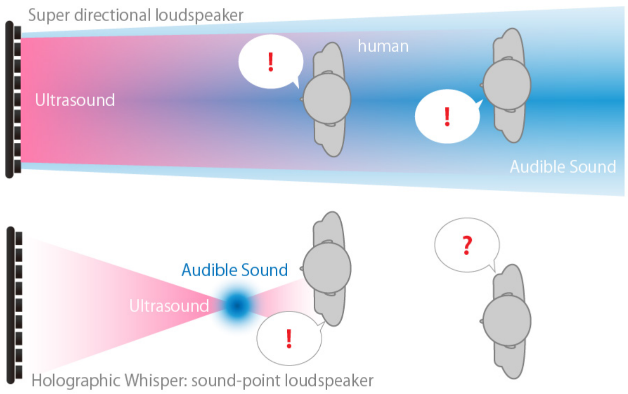


Figure 2. Super-directional loudspeaker and proposed sound-point loudspeaker. The latter loudspeaker can send audible sound to only the target person by focusing the ultrasound.

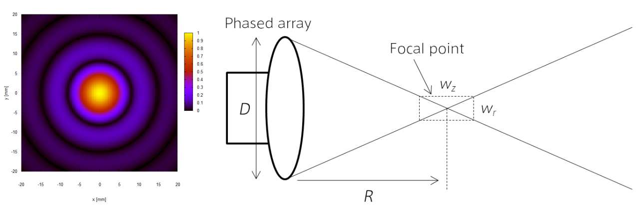


Figure 3. (Left) Spatial distribution of ultrasound on the focal plane, which is known as the Airy pattern. (Right) Focal length, device size, diameter of the focal point, and focal depth.

focal point [13, 14]. Generating multiple focal points simultaneously is possible [15, 16]. The same effect is also used to draw images on object surfaces [17, 18].

Acoustic levitation is another application of ultrasonic phased arrays. The principle of acoustic levitation is that potential wells in an ultrasonic spatial distribution suspend small objects. A line-like standing wave [19] is the simplest configuration, which is formed by two opposite phased arrays that generate focal points at the same position. By widening the ultrasonic beam in one direction and using four phased arrays, a grid-like standing wave is formed, to make a projection screen floating in air [20]. Multiple objects can be moved independently [21]. Generating potential wells with a single phased array has also been demonstrated [22, 23].

The trend that ultrasonic phased arrays are actively used in the HCI field motivated us to apply it to audible-sound field manipulation. Because audible sound radiation requires lower intensity than force generation for noncontact haptic feedback and acoustic levitation, we can distribute the acoustic power of ultrasound among more focal points.

PRINCIPLES

We utilize the self-demodulation effect to radiate audible sound from ultrasound. In order to achieve an effective ultrasonic intensity at a localized area (focal point), we employ a phased-array focusing technique. We use amplitude modulation (AM) to modulate ultrasound, which is implemented into digital control circuits as pulse-width modulation (PWM) of the driving rectangular waves. These topics are explained as follows.

Self-demodulation effect

The temporal change of ultrasonic (primary) waves, p_{u} [Pa], leads to the radiation of audible sound (secondary) waves, p_{a} [Pa] [5]. This effect is called self-demodulation, and is expressed by

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)p_{\mathrm{a}} = -\frac{\beta}{\rho c^4}\frac{\partial^2}{\partial t^2}p_{\mathrm{u}}^2, \qquad (2)$$

where c [m/s] is the speed of sound in air, β is the nonlinear parameter, and ρ [kg/m^3] is the mass density of air. The left side of this equation is the wave equation of the secondary wave, and the right side is the driving force arising from the second-order differential of the square of the primary wave. This effect becomes obvious when p_{u} is effectively high and its square value is not negligible. Here, we name this the audible-sound radiation threshold p_{TH} [Pa]. Both of AM and frequency modulation (FM) can be used to modulate ultrasound.

In the case of super-directional loudspeakers, ultrasonic beams are used and the self-demodulation effect is always accompanied by the parametric array effect. As a result, beams of audible sound are generated. We intend to separate the self-demodulation effect from the parametric array effect. For this purpose, we use the phased-array focusing technique to generate a focal point that is the only place from where audible sound is radiated.

Phased-array focusing

In order to generate a focal point, the distance from the i-th transducer to the focal point, l_i, is an important parameter in calculating the phase delay, differently from the beam steering technique expressed by Eq.(1). This is expressed as

$$\Delta\varphi_i = k\left(l_0 - l_i\right). \qquad (3)$$

Although here we consider the i-th transducer in a 1D array, the extension to the (i, j)-th transducer in a 2D array is straightforward. The position of the focal point can be freely moved by changing the phase delays. Note that the calculated phase delays become identical to those of a conventional super-directional loudspeaker when the position of the focal point is set at an extremely long distance. We can then generate and steer a sound beam, instead of generating a sound point.

The sound pressure distribution at the focal plane is the Airy pattern (Fig. 3 (left)) that is well known in the field of optics. The diameter of a focal point, w_{r} [m], generated by a circular phased array is theoretically given by

$$w_{\mathrm{r}} = 2.44\lambda\frac{R}{D}, \qquad (4)$$

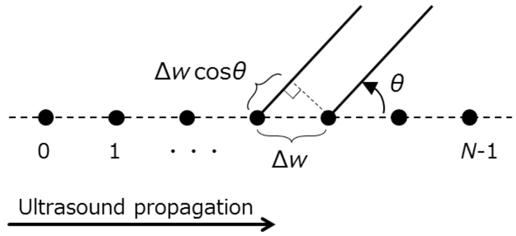


Figure 4. Discrete model of the focal point to estimate the directivity of audible sound. The virtual point sources are distributed along the direction of ultrasound propagation.

where λ [m] is the wavelength of ultrasound, R [m] is the focal length, and D [m] is the diameter of the array. This equation indicates that a higher frequency results in a smaller focal point, and the focal length and array size have a trade-off relationship. The focal depth, w_z [m], (Fig. 3 (right)) is derived from the geometrical relationship (i.e. $w_r : w_z = D/2 : R$), which is given by

$$w_z = 4.88\lambda \left(\frac{R}{D}\right)^2. \tag{5}$$

We tune the amplitude of ultrasound so that $p_u > p_{TH}$ only inside the focal point, to make a point source of audible sound.

Directivity of audible sound

The focal depth dominantly determines the directivity of radiated audible sound. We estimate it by assuming the focal point as an array of point sources with a length of the focal depth and neglecting the width of the focal point (Fig. 4). The number of point sources is N, i.e., $w_z = (N\text{-}1)\Delta w$. The amplitude of audible sound received at a faraway observation point from each point source is assumed as A/N [Pa]. Note that, in reality, the amplitude of each point source is not the same and gradually decreases with wave propagation owing to air absorption and secondary-wave radiation. The interval between the n-th and (n+1)-th point sources is Δw [m]. The phase delays originating from the wave propagation and observation angle, θ [rad], of the n-th point source are $nk\Delta w$ and $-nk\Delta w\cos\theta$, respectively. The amplitude of audible sound received at the observation point, $p(\theta)$ [Pa], is derived as follows:

$$\left|p(\theta)\right| = \left|\sum_{n=0}^{N-1} \frac{A}{N} e^{jnk\Delta w} e^{-jnk\Delta w \cos\theta}\right| = A\left|\frac{\text{sinc}[k(w_z + \Delta w)(1-\cos\theta)/2]}{\text{sinc}[k\Delta w(1-\cos\theta)/2]}\right| \tag{6}$$

where j is the imaginary unit. We obtain the following equation as $\Delta w \to 0$,

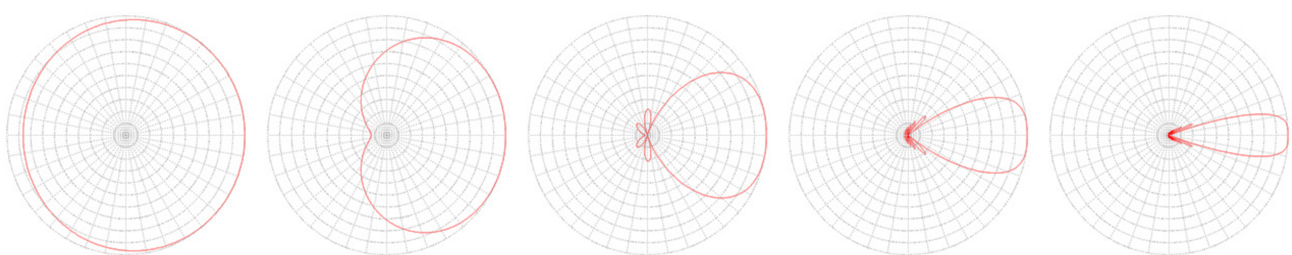
$$p(\theta) \to A\left|\text{sinc}\frac{kw_z(1-\cos\theta)}{2}\right| = A\Phi(\theta). \tag{7}$$

That is, a sinc function expresses the directivity, $\Phi(\theta)$, and the focal depth is a parameter when the wavenumber of audible sound is given.

Figure 5 shows examples of the directivity, $\Phi(\theta)$, of 1-kHz sound for various w_z. The focal point acts as a sound point source if the directivity is quasi-omnidirectional as we intend, and the focal point generates one or more beams if the directivity has zero points for some angles. It is useful to see the directivity as a function of R/D in designing it. For example, the amplitude at $\theta = \pi/2$ rad becomes -3dB when w_z = 150 mm and it is no longer quasi-omnidirectional. This situation occurs when $R/D \approx 1.9$ ($R \approx$ 320 mm with D = 170 mm) for λ = 8.5 mm. Therefore, we can use this phased array as a quasi-point-source loudspeaker at the range of $R/D < 1.9$ for frequencies less than 1 kHz. Note that it becomes similar to a super-directional speaker at the range of $R/D > 1.9$.

Pulse-width modulation

As mentioned above, ultrasonic waves are modulated with audible sound. We implement PWM into a digital control circuit to achieve AM of ultrasonic carrier waves. The driving signals are 40-kHz rectangular waves having



w_z = 50 mm ($R/D \approx 1.1$) | w_z = 150 mm ($R/D \approx 1.9$) | w_z = 500 mm ($R/D \approx 3.5$) | w_z = 1500 mm ($R/D \approx 6.0$) | w_z = 5000 mm ($R/D \approx 11.0$)

Figure 5. Simulated directivities of focal points of various lengths. *R*/*D* is also shown, which is calculated assuming λ = 8.5 mm. Ultrasonic waves propagate from left to right as shown in Fig. 4. It is quasi-omnidirectional up to *R*/*D* ≈ 1.9.

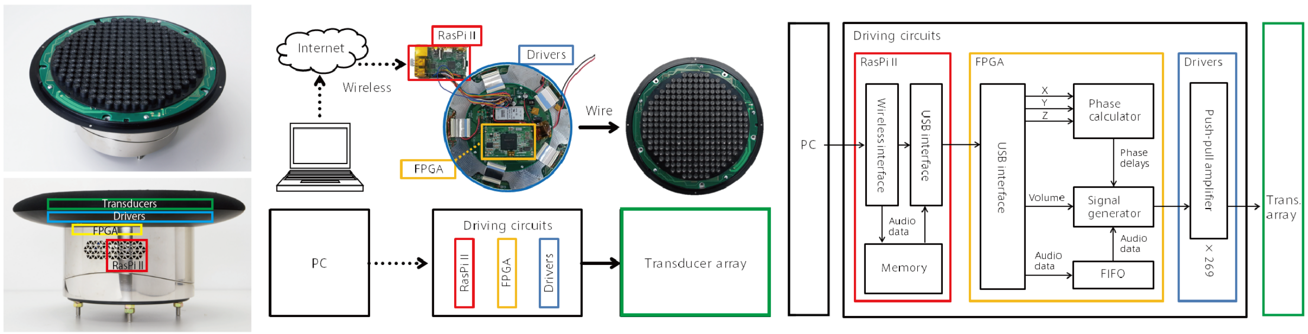


Figure 6. Circuit configuration of the developed prototype. It communicates with the control PC via a Wi-Fi network.

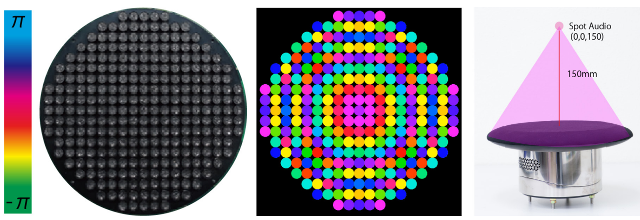


Figure 7. Phase delays to generate an ultrasonic focal point.

adequate phase delays. These signals are initially 3.3 $V_{0\text{-}p}$, and then amplified to V_0 = 24 $V_{0\text{-}p}$ by push-pull amplifiers. The direct current component is removed by high-pass filters before these signals drive the transducers. Because ultrasonic transducers act as narrow band-pass filters, the redundant frequency components in the rectangular waves are eliminated and only the 40-kHz component drives the transducers. The amplitude of the 40-kHz sinusoidal wave, V_1 [$V_{0\text{-}p}$], is not proportional to the pulse width, W [s], and is expressed as

$$V_1 = V_0 \frac{2}{\pi} \sin\left(2\pi \frac{W}{T}\right) \tag{8}$$

where T [s] is the period of the ultrasound (25 μs for 40-kHz ultrasound).

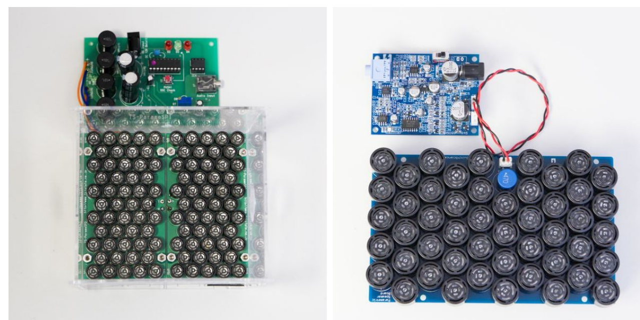
In the PWM method shown above, audio data sampled at 44.1 kHz is not exactly reproduced because the pulse width is updated at 40 kHz. This can be solved by sampling audio data at 40 kHz or using transducers of 44.1-kHz resonant frequency.

PROTOTYPE

Specifications

We developed a prototype of our audio focusing device (Fig. 6). It consists of 269 transducers (40 kHz resonant frequency, 10 mm diameter, T4010B4, Nippon Ceramic Co., Ltd.) arranged in a 170-mm-diameter circular area.

The maximum value of sound pressure at the center of the focal point is approximately 2,400 Pa RMS (estimated), and the diameter and depth of the focal point is calculated as w_r = 24.4 mm and w_z = 57.4, respectively, for a focal length R = 200 mm. This focal point is small enough to act as a quasi-point source of audible sound. The position of the focal point is digitally controlled with a resolution of 1/32 times the wavelength (approximately 0.25 mm for 40-kHz ultrasound) and can be refreshed at 1 kHz.



SDS A **SDS B**

Figure 8. Commercially-available super-directional loudspeakers used in the experiments for comparison.

The prototype phased array consists of three circuit boards. The first board contains the array of ultrasonic transducers, the second is a driving board including an FPGA and push-pull amplifier ICs, and the third is a computer board (RasberryPi 2) for Wi-Fi communication and control. The control application was developed in C++ and runs on Arch Linux.

Here, we explain how to output a single focal point. The RasberryPi 2 communicates with external computers via a Wi-Fi network to receive commands, and then sends the data (the coordinates of the focal point and pulse width) to the driving board. The driving board receives the data, calculates the adequate phase delays for all the transducers based on the distance to the focal point (Fig. 7), and generates the driving signals. The driving signals are sent to the transducers via the amplifiers and HPFs. Note that we can modify the phase-delay calculation algorithm to form spatial distributions other than a single focal point. PWM control is conducted digitally with a resolution of 624 steps. Although this resonant-type PWM is nonlinear as mentioned above, linear conversion from audio data to

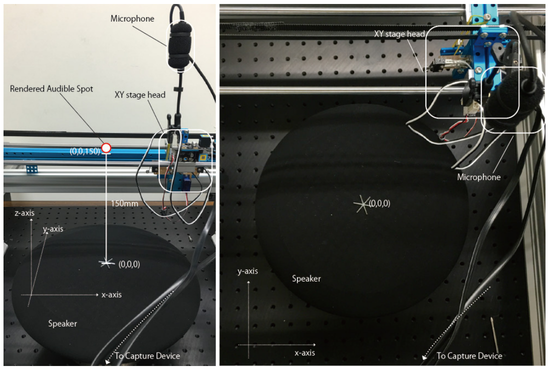


Figure 9. Experimental setup. A microphone is fixed on an XY stage to scan the sound pressure distribution above HW.

pulse width is used in the prototype within an approximately linear range (from around 100 to 500 of 624 steps) for simplicity. Precise nonlinear conversion using a sinusoidal function will improve the sound quality.

EXPERIMENTS

We conducted two experiments for evaluation of the spatial distribution and audible sound characteristics of our Holographic Whisper (HW) device. We compared our prototype device with two commercially-available super-directional speakers; SDS A (k2617 parametric speaker kit, Tri-State) and SDS B (super-directional speaker kit, Switch Science) shown in Figure 8. The WH generated a single sound point at 150 mm above (along the *z*-axis) the speakers. With this focal length, the focal depth is w_z = 16 mm and the focal point effectively acts as a point source. All sound signals for evaluation were recorded in 24 bit and 48 kHz WAV format.

Spatial acoustic distribution

The spatial distribution of a 1 kHz sine wave was measured. Figure 9 shows the experimental setup. The HW was directed upward, and each speaker was driven by a 1-kHz sine signal. An *x-y* stage was used to move the microphone (ATM-350, Audio-Technica) over the speakers. The sound signal was captured by an audio interface (QUAD CAPTURE, UA-55, Roland). The capture intervals in the *x*- and *y*- directions were 10 mm.

Figure 10 shows the amplitude distribution of 1-kHz sound

Figure 11. Experimental setup. A microphone is set in front of loudspeakers to record sound signals.

radiated from the HW on the *x-y* planes at different distances along the *z*-axis. The amplitude is a relative value. It is shown that the HW could successfully generate a sound source only on the focal plane, which is within a 30 × 30 mm area. The sound near the surface is observed, however it does not affect the focal point. This comes from the electric circuit and can be reduced in future devices.

Signal characteristics

We conducted experiments in order to show that the HW can reproduce audio data. A song "Twinkle, twinkle, little star" starting from two different keys, C3 (130.81 Hz) and C4 (261.63 Hz) was used. The HW and two SDSs were set vertically to radiate ultrasound upwards. The audible sound was captured at a distance of 750 mm. We used a microphone (ECM-CG60, Sony) and an audio interface (DR-07MKII, TASCAM). The focal length of the HW was adequately long so that the HW radiated a plane wave in this experiment. That is, all the speakers generated plane waves toward the microphone.

Figures 11 and 12 show the experimental setup and results, respectively. The uppermost panel of Fig. 12 shows the original temporal waveform and the upper middle, lower middle, and bottom panels show the Fourier transforms of the original signal, and recorded signals from SDS A, SDS B, and WH, respectively. The 16 kHz and 8 kHz components included in all the data are subharmonics of the 40 kHz carrier frequency.

Figure 13 shows the magnified images of Fig. 12. The HW seemed to reproduce signals from 3 kHz to 8 kHz more accurately in comparison with SDS B, and approximately as well as SDS A. At a minimum, we can conclude the sound quality of HW is no less than SDS A and B.

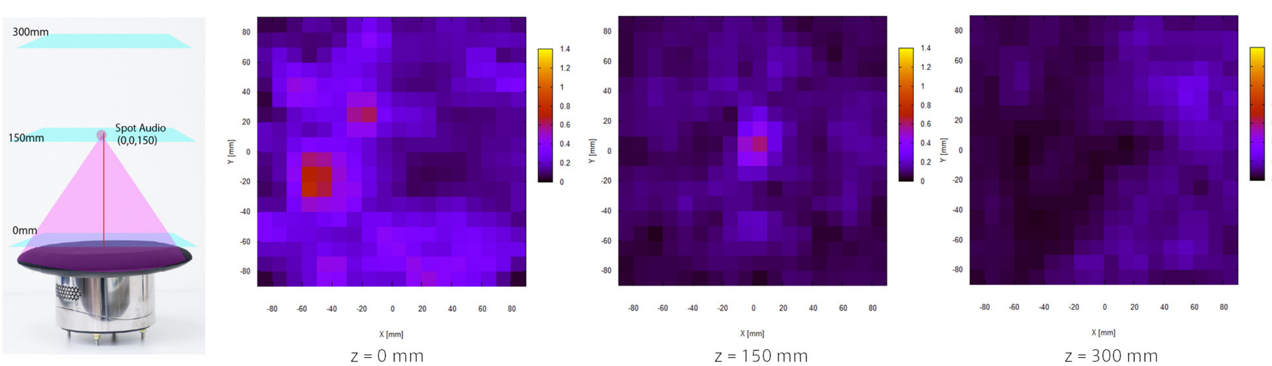


Figure 10. Measured spatial distribution of 1-kHz audible sound before, on, and after the focal point. *R* = 150 mm.

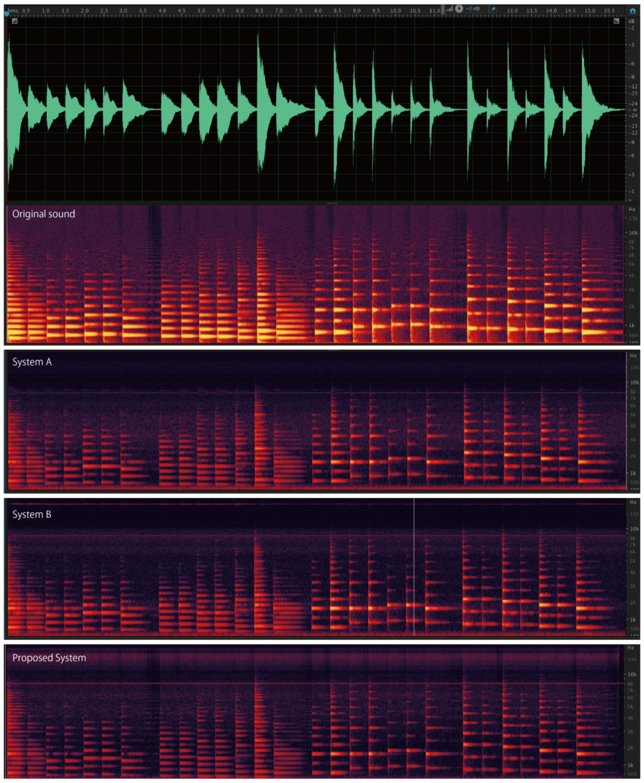


Figure 12. From the top, original signal and its spectrum, and spectrums of measured signals from SDS A, SDS B, and HW.

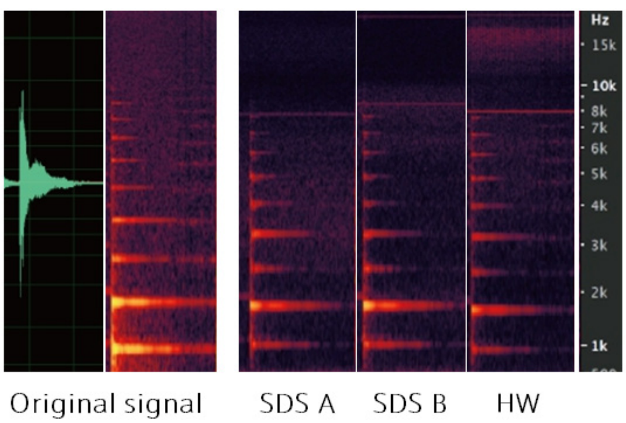


Figure 13. Magnified images of a single note from Fig. 12. There is no significant difference between these signals.

USER STUDIES

The perceived position of a sound image generated by Holographic Whisper was investigated. Holographic Whisper was facing upwards (Fig. 14). It generated a sound point source at various positions on the *x*-*y* plane and along the *z*-axis. Fourteen people (4 females and 10 males, age from 18 to 36, no auditory difficulties) participated in the following studies. Before each user study, they were instructed regarding the ability of Holographic Whisper and experienced all the trial conditions.

User study 1: Azimuth indication

Holographic Whisper can move the sound point source on the *x*-*y* plane. This user study was conducted to confirm whether the participants can perceive the position of the focal point. Each participant was asked to freely walk around the device, to hear sound, and to answer where the amplitude is the highest. Three different positions (the chairs A, B, and C, as shown in Fig. 14) were tested for each participant. As a result, 42 (= 14 × 3) trials were conducted in total. The coordinates (x, y, z) of the focal point for A, B, and C are (0 mm, 300 mm, 500 mm), (300 mm, 300 mm, 500 mm), and (300 mm, 0 mm, 500 mm), respectively.

Figure 15 shows the merged results. There is just one error, and these results indicate that the participants perceived the position of the focal point with high accuracy.

User study 2: Height perception

Holographic Whisper can also move the sound point source along the *z*-axis. This user study was conducted to examine whether humans can perceive the height of the focal point. Each participant was asked to freely move his/her head up and down, to hear sound, and to answer the perceived height of the sound point source. Three different heights (the device surface, the air space above the device, and the ceiling, as shown in Fig. 14) were tested for each participant. As a result, 42 (= 14 × 3) trials were conducted in total. Note that, in the case of focusing at the ceiling, Holographic Whisper acted as a super-directional loudspeaker (i.e., the device generated a sound beam rather than a sound point).

Figure 15 shows the merged results. The diagonal components (i.e., right answers) have the highest scores, and these results indicate that the participants could discriminate the three heights.

User study 3: Height perception (low amplitude)

This user study was conducted to evaluate how correctly humans can perceive the height of the focal point. The amplitude of audible sound was lower than the other user studies in order to highlight the focal point as a sound source. Each participant was asked to freely move his/her head up and down, to hear sound, and to answer the perceived height of the sound point source. Four different heights (200, 400, 600, and 800 mm, as shown in Fig. 14) were tested for each participant. As a result, 56 (= 14 × 4) trials were conducted in total.

Figure 15 shows the merged results. Although the inner cases (400 and 600 mm) tend to be answered correctly, the both-end cases (200 and 800 mm) tend to be answered wrongly.

APPLICATIONS

Holographic Whisper enables new types of audio applications. Because its sound focal points are invisible and intangible, it is also applicable to augmented reality (AR) and virtual reality (VR) applications. It is especially

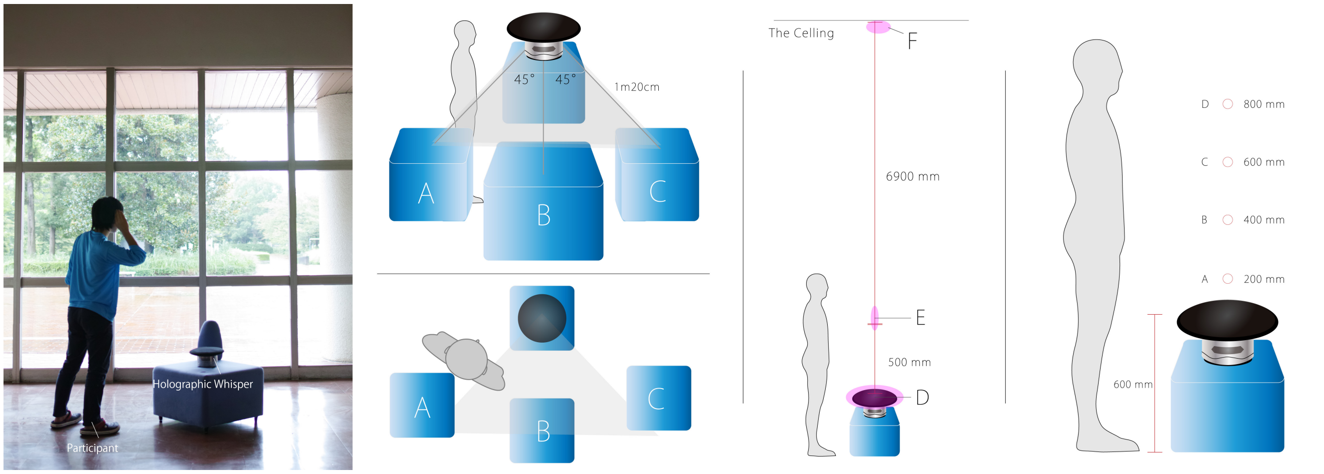


Figure 14. Experimental setups for user studies of azimuth indication and height perception (long and short distances).

User study 1: Azimuth indication		Answered		
		A	B	C
Presented	A	14		
	B		14	1
	C			13

User study 2: Height perception		Answered		
		D	E	F
Presented	D	9	3	
	E	5	9	3
	F		2	11

User study 3: Height perception		Answered [mm]			
		200	400	600	800
Presented [mm]	200	4	1	2	1
	400	6	5	2	3
	600	4	3	7	8
	800		5	3	2

Figure 15. Experimental results: Confusion matrices.

suitable if the position and/or movement of the sound source plays an important role. Here we show examples of possible applications.

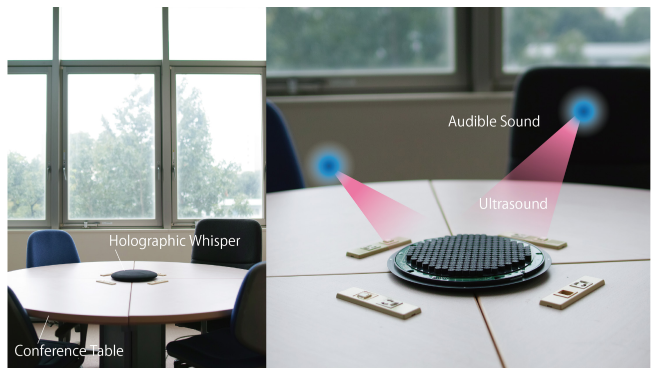


Figure 16: Audio application: Meeting table.

Audio applications

Sound sources floating over a table
In this application, Holographic Whisper is mounted in the center of a table facing upward (Fig. 16). A sound source is placed at the eye-level of audiences. In the case of teleconference, audiences cannot help but look at the sound source (i.e., loudspeaker) when a person who is at a remote place talks. However, they can maintain eye contact with each other using Holographic Whisper.

The most significant feature of this setup is that audible sound comes from an intermediate position between the table surface and ceiling where a sound point source is generated. If we place an ordinary loudspeaker instead of Holographic Whisper, audiences would perceive audible sound was originating from the table surface. If we place a super-directional loudspeaker, audiences would perceive that audible sound was originating from the ceiling as a result of reflection.

Multi-language simultaneous interpretation
Holographic Whisper can deliver different messages to the ears of different persons, in the same setup (Fig. 16), with smaller volume, if the positions of the ears are known or tracked by image processing. This ability is suitable for

Figure 17. Audio application: Targeting walking person.

Figure 18. AR application: Stuffed animals talking to a user.

simultaneous interpretation at international meetings. Each participant can receive his/her mother tongue from a translator without wearing an earphone.

Targeting an individual pedestrian
Holographic Whisper can move the position of a sound point source, and this feature could be useful for following a target person for explanation in a museum and announcements in a café if their positions are tracked by image processing. Moreover, talking to a person who is behaving suspiciously, using Holographic Whisper, may contribute to the prevention of crimes. A sound point source is always generated near the ear even if the target person is walking (Fig. 17).

A similar effect is produced by a conventional super-directional loudspeaker with steering function. However, reflection from the walls and ceiling is not desired in crowded places, and so the installation must be designed carefully. Holographic Whisper does not produce a sound beam but a sound point, and so it has a higher degree of freedom of installation than super-directional loudspeakers.

AR/VR applications

Superimposition on real objects
Holographic Whisper can also add audio to real objects. For example, sound sources are placed in front of mouths of persons in displays or stuffed animals (Fig. 18). Then, users feel as if these persons or animals emit sounds from their mouths. It is also possible to for inanimate objects to be superimposed with indications or messages such as “do not touch me.”

A similar effect is produced by a conventional super-directional loudspeaker with reflection on a real object. However, the reflection angle is determined by the incident angle and the object’s geometry. In the case of Holographic Whisper, these conditions do not affect sound radiation.

Superposition on real environment
Holographic Whisper can enrich AR games such as Ingress [25] and Pokémon GO [26]. When players come to check points, called portal/Pokestops, messages and sound effects can be provided without bothering other people.

Superimposition on mid-air graphics
Recently, various technologies have been developed to render mid-air graphics, such as 3D displays and head-mounted displays. Ordinary methods to provide 3D audio are binaural headphones and 5.1-ch surround sound systems. However, Holographic Whisper can superimpose sound sources on mid-air graphics, e.g., virtual birds singing as they fly. Users feel as if these graphics emit sounds without wearing devices or using multiple loudspeakers.

Representation of invisible objects
Even if there is no visual information, people can feel the movement of objects by their sound localization ability. The moving sound sources generated by Holographic Whisper can represent invisible ghosts, virtual birds, or tiny insects flying around users.

DISCUSSION

Limitation

Ultrasound exposure
Intense ultrasound is required to radiate audible sound. It can potentially damage human ears, and so careful management is important for safe use of ultrasonic devices. The issue of ultrasound exposure has previously been studied and reported [24]. It is generally recommended to avoid unnecessary ultrasound exposure by shortening listening time, limiting ultrasound amplitude, and keep a proper distance from ultrasonic devices so that ultrasonic waves are well absorbed in air before reaching the ears.

Holographic Whisper has the advantage of providing less ultrasound exposure to a listener in comparison with super-directional loudspeakers. A super-directional loudspeaker radiates an audible-sound beam in the same direction as an ultrasonic beam. This leads to ultrasound exposure when delivering audio to audiences. However, Holographic Whisper generates an audible-sound point source from focused ultrasound. The radiated audible sound is quasi-omnidirectional, i.e., not only parallel to the propagation direction of ultrasonic waves but also perpendicular to that. Hence, audiences can receive audible sound whilst standing out of the area of intense ultrasound.

Here we estimate the ultrasonic sound pressure that audiences receive. Suppose that we have 130 dB SPL (typical value for self-demodulation effect [5]) at 1 cm from the center of the focal point. Because the sound pressure distribution on the focal plane is the Airy pattern (the first-order Bessel function of the first kind divided by the distance) as shown in Fig. 3, the sound pressure decreases to 96 dB SPL at 50 cm from the focal point. This is lower than the exposure limit, 110 dB SPL, recommended in [27].

Effective range

Holographic Whisper's directivity gets narrower to be similar to that of super-directional loudspeakers when the focal length R is set at a long distance. The device size D is a parameter to compensate this effect because R/D determines the directivity as shown in Fig. 5. The current prototype (D = 170 mm) can generate a quasi-point source up to $R \approx 320$ mm ($R/D \approx 1.9$). If the focal length up to 1 m is desired, the device size should be larger than 526 mm. This is a trade-off relationship between the effective range and the device size.

Grating lobes

Phased arrays generally suffer from grating lobes (especially large side lobes), which are unnecessary beams originating from the discreteness of transducer arrays. This effect is explained as aliasing owing to the transducer interval being longer than the wavelength. In the current prototype, grating lobes are generated at approximately 40° from the main lobe (d = 10 mm and λ = 8.5 mm).

These grating lobes radiate sound at a volume lower than the main focal point owing to the directivity of the transducers and they are not noisy. We can also avoid these grating lobes to affect unrelated people by controlling the position and the timing of the sound point source adequately, based on vision sensing, under the condition that people are moving and sparse.

In order to eliminate grating lobes, smaller diameter transducers such as [28] (d = 5.2 mm and λ = 8.5 mm) can be used so that they can be arranged at decreased intervals. Then, the angle of the grating lobes is increased, and finally goes beyond the directivity of transducers.

Amplitude of audible sound

The amplitude of audible sound is determined by the amplitude of ultrasound and modulation depth (fluctuation of amplitude for AM and frequency for FM) according to Eq.(2). It is desirable that the amplitude of ultrasound is small, and so a reasonable method is AM with ultrasonic amplitude as small as possible with a maximum modulation depth (i.e., from zero to maximum amplitude).

The amplitude of audible sound determines not only the directivity of sound point as discussed after Eq.(7) but also the audible range of sound point source. These parameters are in trade-off relationships, and designed carefully according to the purpose.

Arrangement of multiple sound sources

Holographic Whisper is not limited to a single sound point source. Various arrangements of multiple sound point sources are possible by further sophisticated phase control. The simplest method is generating two spot sources by dividing the phased array into two sections. A computer-generated hologram (CGH) is another method, which is usually used in the field of laser optics [29] and also applicable to acoustics [30]. Note that all focal points radiate the same sound because they are synchronized.

There is a possibility to produce the directivity of radiated audible sound by arranging point sources in air. In other words, the ultrasonic phased array generates an array of point sources of audible sound in air. The control method is tuning the distance between the point sources, instead of the phase delays, so that the radiated sound waves interfere with each other. The directivity of arranged point sources can also be altered by slightly changes a CGH of ultrasound to rearrange the positions of point sources.

Future work

The prototype of Holographic Whisper is currently at the stage of feasibility study. The future work includes the following issues.

The PWM control of the driving signals should be optimized. Equations (2) and (8) express the nonlinear characteristics of the self-demodulation effect and the resonant phenomenon, respectively. However, linear conversion from audio data to pulse width is used in the prototype for simplicity, which causes a distortion of sound.

Any high-frequency background noise should be eliminated for comfortable user experiences. Electric engineers say that this may come from the mechanical vibration of surface-mounted ceramic capacitors and it is an issue in commercial products such as PC. We plan to replace these capacitors with low-noise ones.

The positioning and tracking of the user is important in most of Holographic Whisper's applications. Vision sensing is suitable for our purpose because of its wide sensing area, easy installation, and wide variety of available post-processing methods. We intend to develop systems and services using Holographic Whisper.

CONCLUSION

We proposed a sound-point loudspeaker to render sound point sources in air by using an ultrasonic phased array. We explained the principles: Self-demodulation effect, phased-array focusing, directivity of audible sound, and pulse-width modulation. We confirmed a point source of audible sound is successfully generated. We conducted the experiments on sound quality and spatial distribution, and a user study on perceived position of a sound source in a room. We listed the examples of possible applications. The proposed sound-point loudspeaker has several advantages in comparison with conventional super-directional loudspeakers, such as audience selectivity and low ultrasound exposure.

REFERENCES

1. A.J. Berkhout. 1988. A holographic approach to acoustic control, *Journal of the Audio Engineering Society* 36, 977-995.
2. A.J. Berkhout, D.D. Vries, and P. Vogel. 1993. Acoustic control by wave field synthesis, *The Journal of the Acoustical Society of America* 93, 2764-2778.
3. BOSE. F1 flexible array loudspeaker system. from http://www.bose.com/prc.jsp?url=/promotions/entry_pages/f1/index_en.jsp
4. JBL. IntelliDisc-DS90. from http://www.jblpro.com/www/products/installed-sound/intellidisc-ds90
5. M. Yoneyama, J. Fujimoto, Y. Kawamo, and S. Sasabe. 1983. The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design. *The Journal of the Acoustical Society of America* 73, 5, 1532-1536.
6. K. Aoki, T. Kamakura, and Y. Kumamoto. 1991. Parametric loudspeaker - Characteristics of acoustic field and suitable modulation of carrier ultrasound, *Electronics and Communications in Japan* 74, 9, 76-82.
7. F.J. Pompei. 1999. The use of airborne ultrasonics for generating audible sound beams, *Journal of the Audio Engineering Society* 47, 9, 726-731.
8. W.-S. Gan, J. Yang, T. Kamakura. 2012. A review of parametric acoustic array in air. *Applied Acoustics* 73, 1211-1219.
9. K. Shinagawa, Y. Amemiya, H. Takemura, S. Kagami, and H. Mizoguchi. 2007. Three dimensional simulation and measurement of sound pressure distribution generated by 120 ch plane loudspeaker array. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, 278-283.
10. T. Matsui, D. Ikefuji, M. Nakayama, and T. Nishiura. 2014. Multiple audio spots design based on separating emission of carrier and sideband waves. In *Proceedings of International Congress and Exposition on Noise Control Engineering*, 6313-6317.
11. W.-S. Gan, J. Yang, K. S. Tan, and M.-H. Er. 2006. A digital beamsteerer for difference frequency in a parametric array. *IEEE Transactions on Audio, Speech, and Language Processing* 14, 3, 1018-1025.
12. C. Shi, Y. Kajikawa, and W.-S. Gan. 2014. An overview of directivity control methods of the parametric array loudspeaker. *APSIPA Transactions on Signal and Information Processing* 3, e20.
13. T. Hoshi, M. Takahashi, T. Iwamoto, and H. Shinoda. 2010. Noncontact tactile display based on radiation pressure of airborne ultrasound. *IEEE Transactions on Haptics* 3, 3, 155-165.
14. K. Hasegawa and H. Shinoda. 2013. Aerial display of vibrotactile sensation with high spatial-temporal resolution using large-aperture airborne ultrasound phased array, In *Proceedings of IEEE World Haptics Conference*, 31-36.
15. T. Carter, S. A. Seah, B. Long, B. Drinkwater, and S. Subramanian. 2013. Ultrahaptics: Multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, 505-514.
16. S. Inoue, Y. Makino, and H. Shinoda. 2015. Active touch perception produced by airborne ultrasonic haptic hologram, In *Proceedings of IEEE World Haptics Conference*, 362-367.
17. Y. Sugiura, K. Toda, T. Hoshi, Y. Kamiyama, T. Igarashi, and M. Inami. 2014. Graffiti fur: Turning your carpet into a computer display, In *Proceedings of Annual ACM Symposium on User Interface Software and Technology*, 149-156.
18. A Marzo, R. McGeehan, J. McIntosh, S.A. Seah, and S. Subramanian. 2015. Ghost touch: Turning surfaces into interactive tangible canvases with focused ultrasound, In *Proceedings of International Conference on Interactive Tabletops & Surfaces*, 137-140.
19. T. Hoshi, Y. Ochiai, and J. Rekimoto. 2014. Three-dimensional noncontact manipulation by opposite ultrasonic phased arrays, *Japanese Journal of Applied Physics* 53, 07KE07.
20. Y. Ochiai, T. Hoshi, and J. Rekimoto. 2014. Pixie dust: Graphics generated by levitated and animated objects in computational acoustic-potential field. *ACM Transactions on Graphics* 33, 4, article no. 85.
21. T. Omirou, A. Marzo, S.A. Seah, and S. Subramanian. 2015. LeviPath: Modular acoustic levitation for 3D path visualisations, In *Proceedings of Annual ACM Conference on Human Factors in Computing Systems*, 309-312.
22. A. Marzo, S. A. Seah, B. W. Drinkwater, D. R. Sahoo, B. Long, and S. Subramanian. 2015. Holographic acoustic elements for manipulation of levitated objects. *Nature Communications* 6, article no. 8661.
23. A. Marzo. 2016. GauntLev: A wearable to manipulate free-floating objects, In *Proceedings of Annual ACM Conference on Human Factors in Computing Systems*, 3277-3281.
24. C. Q. Howard, C. H. Hansen, and A. C. Zander. 2005. A review of current ultrasound exposure limits. *Journal Occupational Health and Safety - Australia and New Zealand* 21, 3, 253-257.
25. Niantic Inc. Ingress. from https://www.ingress.com/
26. Niantic Inc. Pokémon GO. from http://pokemongo.nianticlabs.com/

27. C.Q. Howard, C.H. Hansen, and A.C. Zander. 2005. A review of current ultrasound exposure limits, *The Journal of Occupational Health and Safety of Australia and New Zealand* 21, 3, 253-257.

28. Murata Manufacturing Co., Ltd. MA40H1S-R. from http://www.murata.com/en-global/products/productdetail.aspx?cate=cgsubUltrasonicSensors&partno=MA40H1S-R

29. J. Bengtsson. 1994. Kinoform design with an optimal-rotationangle method. *Applied Optics* 33, 29, 6879-6884.

30. Y. Xie, C. Shen, W. Wang, J. Li, D. Suo, B.-I. Popa, Y. Jing, and S.A. Cummer. 2016. Acoustic holographic rendering with two-dimensional metamaterial-based passive phased array, *Scientific Reports* 6, article no. 35437.