AUTD3: Scalable Airborne Ultrasound Tactile Display

Shun Suzuki, Seki Inoue, Masahiro Fujiwara, *Member, IEEE*, Yasutoshi Makino, and Hiroyuki Shinoda, *Member, IEEE*

***Abstract*—Through nonlinear effects, airborne ultrasound phased arrays enable mid-air tactile presentations, as well as auditory presentation and acoustic levitation. To create workplaces flexibly, we have developed a scalable phased array system in which multiple modules can be connected via Ethernet cables and controlled from a PC or other host device. Each module has 249 transducers and the software used can individually specify the phase and amplitude of each of the connected transducers. Using EtherCAT for communication, the system achieves high accuracy synchronization among the connected modules. In this article, we describe the details of the hardware and software architecture of the developed system and evaluate it. We experimentally confirmed the synchronization of 20 modules within an accuracy of $0.1\ \mu\mathrm{s}$ and the phase and amplitude can be specified at $8\ \mathrm{bits}$ resolution. In addition, using nine modules, we confirmed that we could make a focal point of the size consistent with the theory at $500\ \mathrm{mm}$ above the array surface.**

***Index Terms*—Mid-air ultrasound haptics, phased array.**

I. Introduction

In this paper, we describe and evaluate the hardware and software architecture of the scalable airborne ultrasound phased array system proposed in [1]. Fig. 1 shows the phased array system, referred to as an Airborne Ultrasound Tactile Display 3 (AUTD3). The AUTD3 modules can be connected and extended to each other using Ethernet cables. Furthermore, the phase and amplitude of each transducer can be individually controlled, enabling the production of arbitrary pressure patterns. Owing to these features, AUTD3 has already been used in various studies, where a multi-module aperture was created to realize a large workspace, sometimes surrounded by multiple modules with different orientations. For example, in [2], eighteen AUTD3 devices cooperate to generate a focal point at a distance of approximately $1\ \mathrm{m}$ from the array surface and it was shown that such focus could be felt by a chest skin through clothing.

Manuscript received September 7, 2020; revised December 21, 2020 and February 24, 2021; accepted March 26, 2021. Date of publication March 31, 2021; date of current version December 16, 2021. This article was recommended for publication by Associate Editor Dr. Matteo Bianchi and Editor-in-Chief Marcia K. O'Malley upon evaluation of the reviewers' comments. This work was supported in part by JSPS KAKENHI under Grant 16H06303 and in part by JST CREST under Grant JPMJCR18A2. *(Corresponding author: Shun Suzuki.)*

The authors are with the Graduate School of Frontier Sciences, The University of Tokyo13143 , Kashiwa-shi, Chiba 277-8561, Japan (e-mail: suzuki@hapis.k.u-tokyo.ac.jp; inoue@hapis.k.u-tokyo.ac.jp; Masahiro_Fujiwara@ipc.i.u-tokyo.ac.jp; yasutoshi_makino@k.u-tokyo.ac.jp; hiroyuki_shinoda@k.u-tokyo.ac.jp).

Digital Object Identifier 10.1109/TOH.2021.3069976

Acoustic phased arrays have mainly been applied to nondestructive diagnosis and inspection of human bodies and constructions, based on solid-borne and structure-borne sound. In addition to sensor applications, airborne ultrasonic phased arrays as wave sources have been studied in a parametric speaker [3], [4] and a focal point sound source [5]. Recently, acoustic levitation using acoustic radiation pressure has also attracted researchers' interest [6]–[8]. In addition to levitating particles smaller than its wavelength, Inoue *et al.* achieved non-rotating levitation of a solid object several times larger than the wavelength ($8.5\ \mathrm{mm}$), using a $40\ \mathrm{kHz}$ ultrasound [9]. Dynamic control of lightweight objects is also of current interest [10]–[12]. Furumoto *et al.* proposed a floating display by controlling floating balloons three-dimensionally [13], [14]. Ping-pong ball control using high-speed feedback has also been demonstrated [15].

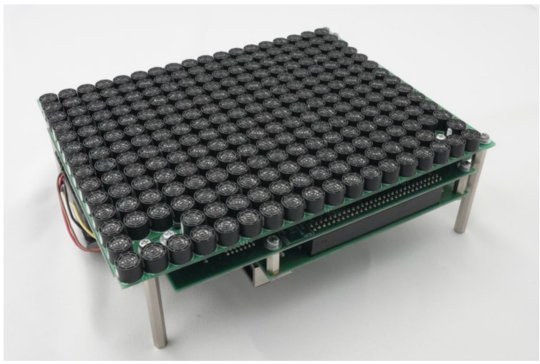
Ultrasonic tactile presentation originates from underwater stimulation. Dalecki *et al.* first showed that a tactile sensation could be generated through ultrasonic irradiation in water [16]. The first aerial tactile stimulation was demonstrated by Iwamoto et al. who demonstrated that airborne convergent ultrasound could apply a tactile sensation to the skin [17]. Subsequently, many systems have been proposed that provide a tactile sensation through an ultrasound phased array [18]–[20].

The advantage of mid-air haptics is attributed to the absence of physical contact. The user's movement is not restricted, the stimulation area is freely selected, and the system is hygienic. For basic-research purposes, the pressure pattern can be controlled with reproducibility while avoiding the bias of feeling that they are touching a tactile display device. In addition, it is sometimes preferable that the display area is not hidden by the devices.

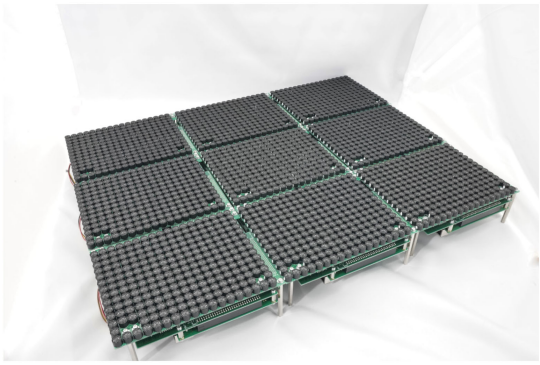
A problem with mid-air ultrasound haptics is the weak displayed stimulation. As a solution to this, stimulation techniques as lateral modulation [21] and spatiotemporal modulation [22] have been proposed that increase the perceived intensity.

Furthermore, recent studies have shown that not only vibratory sensations but also thermal sensations could be displayed. For example, in [23], a cold feeling is given remotely through the transport of cold air by the acoustic flow associated with a Bessel beam. In [24], a method for providing heat using the thermoacoustic phenomenon of ultrasonic waves passing through a narrow slit was demonstrated.

Table I presents an overview of the phased array systems for a haptic presentation. In the original study by Hoshi *et al.* [18] and the early research conducted thereafter, prototypes were



(a) AUTD3 module



(b) Nine connected modules

Fig. 1. Photograph of (a) a single module of the proposed array system and (b) nine connected modules. The connected modules can be controlled from a computer as a single array.

created for various purposes [19], [20]. These studies showed the possibility of achieving a tactile presentation using ultrasound phased arrays. Hasegawa *et al.* proposed an extensible phased array system that constructs a tree-topology network [25]. However, owing to the limitation of communication throughput (approximately $1\ \mathrm{Mbps}$), it was not possible to individually transmit the phases and amplitudes of all of the transducers. Thus the phases and amplitudes are calculated on each module; i.e., the displayed patterns were limited to a single focus or very simple patterns where the necessary information for each module is transmitted in a short code. Based on the studies in [19], [26], "Ultrahaptics" has made phased array devices available on the market. Although the application programming interface is currently public, the entire program is not disclosed. Besides, a multi-module connection is not assumed. Marzo *et al.* proposed Ultraino, an open-platform phased array project [27]. Both hardware and software are open-source and we can freely design the geometry of the transducers. Ultraino uses Arduino Mega as an interface to a personal computer (PC) and a signal generator, which makes it easy and inexpensive to use. Furthermore, it can be extended by connecting the devices and synchronizing them by sharing a global clock. However, owing to the limitation of the number of pins, only 64 transducers can be driven per device. In addition, USB to UART, which is relatively slow at $25\ \mathrm{KB/s}$, is used for communication with the PC, limiting the number of connections and the update rate.

II. Modular Phased Array System

Typical acoustic phased array devices used in previous haptics studies, including commercially available ones, are approximately $30\ \mathrm{cm}$ square boards containing ultrasonic transducers with $1\ \mathrm{cm}$ spacing. If a single board is utilized, the workspace is within the $30\ \mathrm{cm}$ cube in front of the board. The limitation is due to the focusability of the acoustic wave.

Applying the Fresnel approximation, the focal size generated by the rectangular aperture array is given by

$$\frac{2\lambda r}{R}, \tag{1}$$

where r is the distance from the array surface, R is the aperture size, and λ is the wavelength [18]. This means that we must shorten the wavelength or increase the aperture size to create a small focal point at distance r. The shorter the wavelength is, i.e., the higher the frequency, the stronger the attenuation from air absorption [28]. Therefore, a practical solution is to widen the aperture.

The directivity of the ultrasound transducer is another critical factor. The sound pressure at focus located distant from the array is proportional to the number of transducers. However, when the focus is close to the array, the sound pressure becomes independent of the total number of transducers. The sound pressure p of an ultrasonic wave at point $\boldsymbol{x}$ emitted from a transducer placed at the origin is represented by the following equation:

$$p(\boldsymbol{x}, t) = p_0\ A \frac{D(\theta)}{\|\boldsymbol{x}\|} \mathrm{e}^{-\beta\|\boldsymbol{x}\|} \mathrm{e}^{\mathrm{j}(\phi + k\|\boldsymbol{x}\|)} \mathrm{e}^{\mathrm{j}\omega t}, \tag{2}$$

where t is time, p_0 is a constant value related to the output power of the transducer, A and ϕ are parameters related to the driving signal, $D(\theta)$ is the directivity, β is the attenuation coefficient, k is the wavenumber, and ω is the angular frequency. The directivity $D(\theta)$ usually takes the maximum value in the normal direction of the transducer. Thus, the best design to maximize the acoustic irradiation on a target is to surround the workspace with transducers oriented to the target. Therefore, the surrounding alignment of multiple modules is sometimes necessary.

Inoue *et al.* first demonstrated AUTD3, a modular phased array system, to meet the above necessity in 2016 [1]. Their modules can be daisy-chained using Ethernet cables and operated as a single phased array system. The aperture size can be widened by arranging the modular arrays and the workspace can be freely designed. Here, the critical problem is the synchronization of such modules. For example, in the case of a $4\ \mathrm{bits}$ phase quantization of a $40\ \mathrm{kHz}$ ultrasound, the timing margin between the transducers is approximately $1\ \mu\mathrm{s}$. In [1], the authors proposed a phased array system using EtherCAT as a communication system, whose distributed clock mechanism enables modules to synchronize at the sub-microsecond level.

AUTD3 is an extensible phased array system based on EtherCAT. Each array consists of 18×14 transducers (owing to the assembly, the total number of transducers is 249, with three

TABLE I
OVERVIEW OF A PHASED ARRAY SYSTEM FOR TACTILE PRESENTATION

	Hoshi et al., 2010 [18]	UltraHaptics, Carter et al., 2013 [19]	Haptgram, Korres et al., 2016 [20]	Hasegawa et al., 2018 [25]	Ultraino, Marzo et al., 2018 [27]	Ultrahaptics, (Commercial)	AUTD3 (Proposed)
Transducers	324	320	96	249	64	256	249
Extensible	No	No	No	Yes	Yes	No	Yes
Voltage	24 V	15 V	20 V	24V	17 V	24V	24 V
Phase res.	$2\pi/16$	$2\pi/50$	N/A	$2\pi/640$	$2\pi/10$	N/A	$2\pi/256$
Update rate	N/A	N/A	N/A	2 kHz	25 Hz	40 kHz	1 kHz
Transducer Individual control	No	Yes	No	No	Yes	N/A	Yes
Interface	Digital I/O	Ethernet	JTAG	via MCU	USB	USB	EtherCAT
Open source	No	No	No	No	Yes	No	Yes

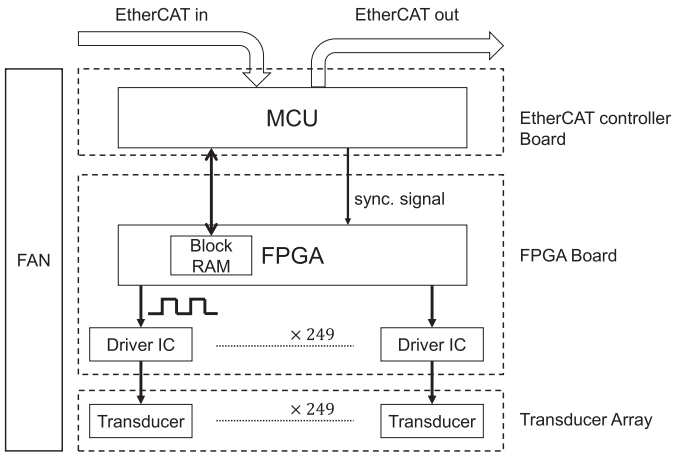


Fig. 2. Hardware configuration of AUTD3.

transducers omitted, see also Fig. 3). The size of the array is $192\ \mathrm{mm} \times 151.4\ \mathrm{mm}$, including the substrate. According to the EtherCAT specifications, a maximum of 65 535 devices can theoretically be connected and thus more than 16 million transducers can be driven at maximum. The phase and amplitude can be specified in 8 bits. These transducers can be specified individually from the software. The amplitude and phase update rate is 1 kHz at maximum, which depends on the number of connected devices as there is an upper limit to the transmission throughput, 100 Mbps. Therefore, the theoretical maximum number of transducers driven at 1 kHz is $100\ \mathrm{Mbps}/1\ \mathrm{kHz}/16\ \mathrm{bit} = 6250$. And, the update rate decreases linearly when the number of transducers exceeds the limit.

This paper is based on the conference paper that reported and demonstrated AUTD3 [1]. However, the details of its realization are not described in [1]. In addition, this device has yet to be objectively verified. In this paper, we examine the basic expandability of AUTD3 and the accuracy of its phase and amplitude resolution to verify whether it can be used for general purposes as a standard device. Throughout the experiment, we confirmed a synchronization of less than $0.1\ \mu$s in a system consisting of 20 devices. We also confirmed that the phase and amplitude of the transducer could be specified at an accuracy of 8 bits. Furthermore, by cooperating with nine devices, we observed a focus 500 mm above the array surface. In addition, we confirmed that the acoustic pressure of a focus increase as the number of modules increase.

III. HARDWARE

As described in [1], the requirements for a 40 kHz frequency phased array module are as follows: an accurate synchronization, a high refresh rate modulation mechanism of over 2 kHz, real-time logic for generating a waveform, and controllability from a non-real-time computer.

First, to achieve both scalability and accurate phase synchronization, we employ a distributed clock, compensating for the clock drift by measuring the communication delay rather than sharing a global clock. As already mentioned, the timing margin is approximately $1\ \mu$s when the phase is quantized in 4 bits. Furthermore, in recent applications where the focus is moved at a high speed, it is necessary to perform fine phase changes quickly. Therefore, high-precision synchronization is required. It is possible to share a global clock in the connection of several modules for synchronization. The advantage of the distributed clock is that no physical care is necessary for the electrical delay.

Second, human beings can sense vibrations within a wide frequency range, from static pressure to nearly 1 kHz [29]. According to the sampling theorem, a refresh rate of 2 kHz would be sufficient to cover human tactile perception. In the previous developments, the typical stimulation was 200 Hz amplitude modulation (AM) and there have not yet been any reports on the necessity for frequency components higher than 200 Hz in mid-air haptics. This is because mid-air stimulation is weak and heightening perceptual strength has been prioritized. At the same time, we expect that high-frequency components will be required to produce a realistic tactile feeling in future systems.

Third, a square wave can drive the ultrasonic transducer; however, unlike pulse width modulation (PWM) motor control, it is necessary to control not only the duty ratio but also the signal timing (shift duration) at a sub-microsecond resolution. Therefore, we need to use a field-programmable gate array (FPGA). Here, the duty ratio determines the amplitude and the shift duration corresponds to the phase of the ultrasound.

Finally, for convenience, the host machine should not require a real-time OS. It is desirable to be able to use a standard PC with Windows, Linux, or macOS.

For the above reasons, the EtherCAT protocol was adopted for the network. EtherCAT is a Fieldbus system based on a 100 BASE Ethernet, with a transmission speed of 100 Mbps [30].

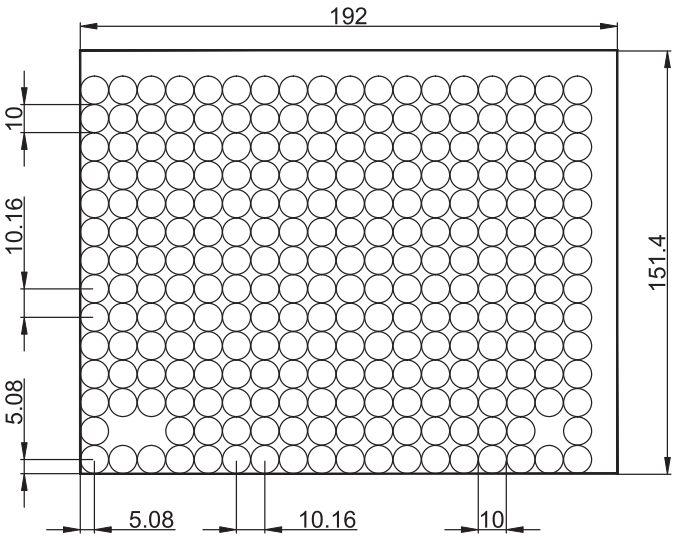


Fig. 3. Dimensional drawing of the transducer array. Dimensions are in mm.

The EtherCAT network consists of a master node and slave nodes. Only one frame is transferred periodically, which maximizes bandwidth utilization. The EtherCAT network has a built-in function called a distributed clock, enabling clock synchronization with a low jitter shorter than 1 μs.

Fig. 2 shows the hardware configuration of the AUTD3 system. The module consists of four main parts: an EtherCAT controller board, a transducer controller board, a transducer array, and a fan for cooling. The EtherCAT controller board is responsible for all communications with the PC and other modules. This board also communicates with the FPGA through block random access memory (RAM). In addition, the synchronization signal generated by EtherCAT is transmitted to the FPGA. The FPGA generates the drive signals, which are input into the transducers through the driver IC. Fig. 3 shows a dimensional drawing of the transducer array, including the substrate.

Below, we describe a method for driving a single transducer and operating multiple modules.

A. Driving Manner of a Transducer

We used a 40 kHz ultrasonic transducer (T4010A1, Nippon Ceramic Co., Ltd.). The directivity of the transducer is depicted in Fig. 4. The drive signal is generated using an FPGA driven by a 25.6 MHz clock. The signal generated by the FPGA is amplified to 24 Vpp by the driver IC (NJM2670, New Japan Radio Co., Ltd.) and applied to the transducer.

Fig. 5 shows the signal generation flow in the FPGA. The amplitude parameter D_i specified by the PC is first multiplied by the modulation data. It then passes through the low-pass filter (LPF) and is input to the PWM signal generator. The phase parameter S_i specified from the PC also passes through the LPF and is input to the PWM signal generation part.

1) Pulse Width Modulation: We control the transducer using PWM. The following equation describes the voltage applied to the transducer:

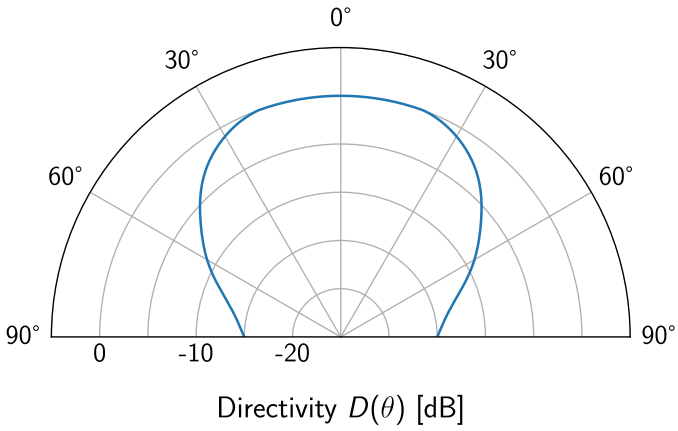


Fig. 4. Directivity of T4010A1 (reconfigured from the datasheet).

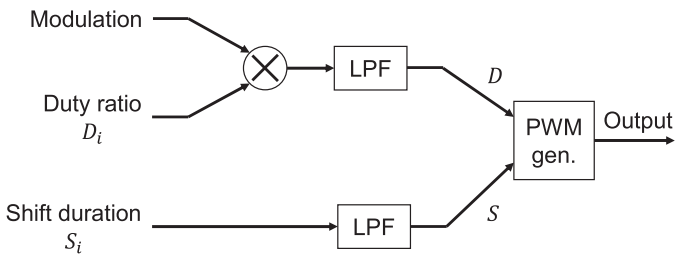


Fig. 5. Signal generation flow.

$$V(t) = \begin{cases} V_0 & (S + nT < t < S + D + nT) \\ 0 & (\text{otherwise}), \end{cases} \tag{3}$$

where $T = 25$ μs is the period, D is the pulse width, S is the shift duration of the signal, and n is an integer. The transducer has a sufficiently high Q-value and hence the output of the ultrasonic wave is proportional to the 40 kHz component, which is given as follows:

$$\frac{1}{T}\int_0^T \mathrm{d}t V(t) \mathrm{e}^{\mathrm{i}\frac{2\pi t}{T}} = \frac{V_0}{\pi} \sin\left(\frac{D}{T}\pi\right) \mathrm{e}^{-2\pi\mathrm{i}\frac{2S+D}{2\,T}}. \tag{4}$$

Therefore, the amplitude A and phase ϕ of the ultrasonic wave can be expressed as follows:

$$\phi = 2\pi \frac{2S + D}{2\,T}, \tag{5}$$

$$A \propto \sin\left(\frac{D}{T}\pi\right). \tag{6}$$

In AUTD3, the parameters D and S can be specified in 8 bits.

2) Silent Mode: As pointed out in [31], a sudden change in the amplitude of the transducer causes a loud noise. In [31], it was confirmed that the noise was suppressed by gradual phase change with linear interpolation. In our system, we implement the silent mode; the amplitude and phase parameters are passed through the LPF in the FPGA to suppress the sudden change in amplitude.

Fig. 6 shows the frequency response of the designed filter. We used an LPF with a cutoff frequency of approximately 150 Hz. Fig. 7 shows the temporal variation of the phase parameters,

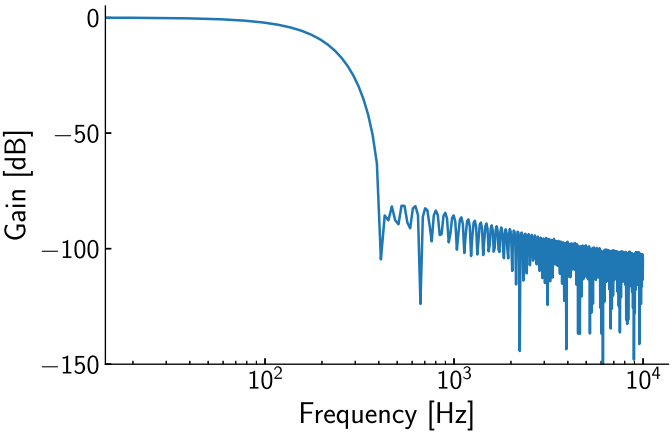


Fig. 6. Frequency response of LPF for eliminating noise.

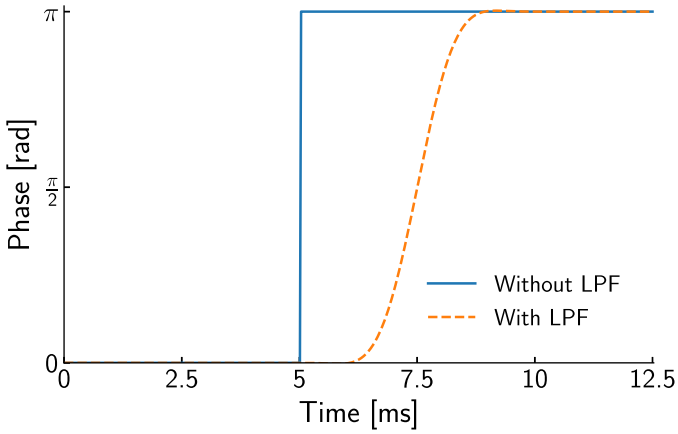


Fig. 7. Effect of LPF. The rapid phase change can be suppressed by passing through the LPF.

with and without a filter. The LPF clearly suppresses the abrupt change in phase. By gradually changing the amplitude, the amplitude fluctuation of the transducer can be suppressed. In our system, the same filter is used for the duty ratio.

Note that the amplitudes and phases may change undesirably as a result of this filtering. Silent mode causes delays in the phase change, as shown in Fig. 7. This leads, for example, to the focus generation being delayed by several ms. In addition, the focus may not even be generated in the case of fast-moving focus applications. Also, the amplitude of the AM may be reduced by the filter and the perception may be weakened. Thus, the enable flag in silent mode can be switched off from the software if deemed unnecessary.

3) Special On-Module Functions: AUTD3 has the function of applying periodic AM, apart from amplitude and phase control. We can specify a periodic 8 bits data array with a maximum size of 4000 from the host device. The data in this array are sampled sequentially at a sampling frequency of 4 kHz and multiplied by the duty ratio. Therefore, it is possible to apply AM at a frequency of 1 Hz to 2 kHz, which includes the human tactile frequency range.

In addition, AUTD3 has a function to provide periodic fast motion of a focal spot for lateral modulation or spatiotemporal modulation. In this function, the phases are calculated on the hardware based on the positional data array of the foci and updated at a refresh rate of 40 kHz at maximum.

B. AUTD3 Module

1) Ethercat: EtherCAT works differently to regular communication. The EtherCAT slaves (AUTD3 modules) read/write data on frames sent by the master (usually a PC in our system) while the frames pass through. This makes it appear that the master and slaves share the same memory space, with minimum delay.

Furthermore, a distributed clock allows the EtherCAT slaves to synchronize with a high accuracy of less than 1 μs. The first slave becomes the master clock, measuring the delay time between each slave and compensating for it accordingly. This compensation is conducted automatically by the hardware, resulting in fast and accurate synchronization.

2) Controller Board: EtherCAT requires a unique EtherCAT slave controller to achieve the characteristics mentioned previously. We used a CPU board (AP-RZT-0 A, AlphaProject Co., Ltd.) with two EtherCAT ports. The board can be daisy-chained with other boards and has 209 parallel I/O ports, some of which are connected to the FPGA. The role of this board is to transfer the phase, amplitude, and modulation data sent from the master to the FPGA.

An FPGA (XC7A200 T, XILINX, Inc.) was mounted onto the FPGA board to drive all transducers of the module. The board also has a circuit to boost the output of the FPGA from 3.3 V to 24 Vpp. Note that we have experimentally confirmed that the deviation in propagation delay among the signals input to the transducers from FPGA via driver ICs is much less than 1 μs and therefore ignorable.

The portions of the CPU board codes that relate to the logic of AUTD3 and the whole FPGA codes are available at https://github.com/shinolab/autd3-library-firmware.

3) Synchronization: The FPGA operates all of the transducers in the module; thus, they are driven with the same clock and synchronized. For synchronization between modules, we used the synchronization signals generated by the distributed clock of EtherCAT. The signals are fired at regular intervals, which can be specified as having a value of 32 bits in units of 1 ns (thus, approximately 4.2 s at maximum). The FPGA is operated using a 25.6 MHz crystal oscillator (SG-8002CE-25.6M-PCB-L2, Seiko Epson Co.). The frequency tolerance of this oscillator is ± 50 ppm; thus, a deviation of 100 μs in 1 s may occur among multiple modules. Therefore, it is necessary to periodically compensate for an interval of approximately 1 ms to maintain the synchronization of 0.1 μs.

IV. Software

The libraries used to control AUTD3 are publicly available at https://github.com/shinolab/autd3-library-software. The software supports cross-platform services as well as Windows, Linux, and macOS. In addition, we can use Raspberry Pi or microcontroller boards, such as Arduino, if an Ethernet port and sufficient memory are available. Although the library is

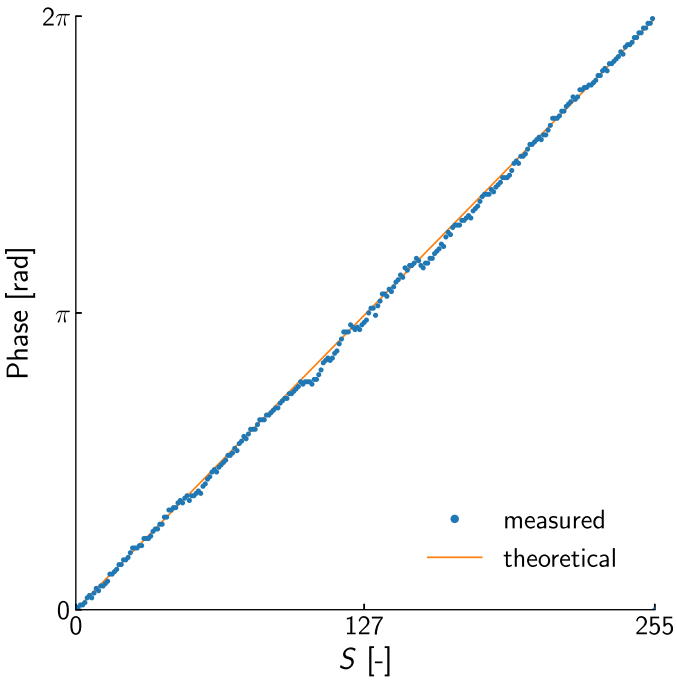


Fig. 8. Measured phase of ultrasound transmitted from a transducer.

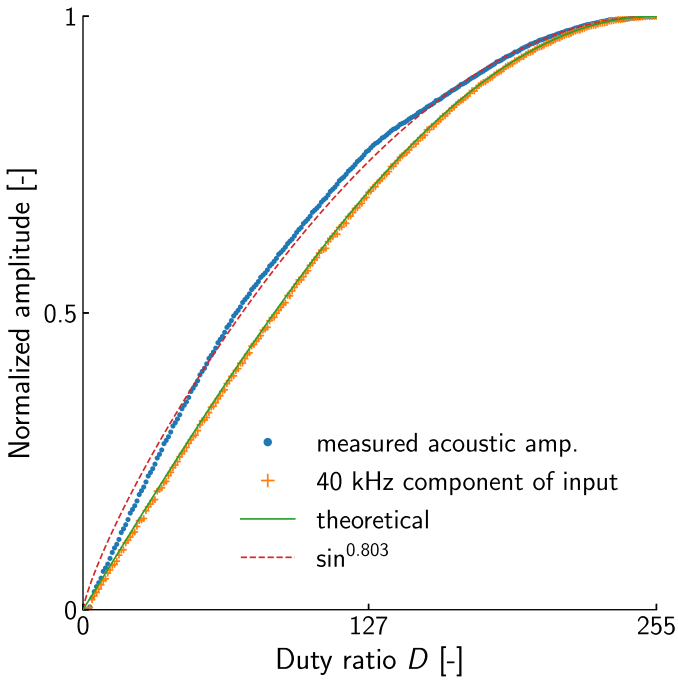


Fig. 9. Measured amplitude of ultrasound transmitted from a transducer and 40 kHz component of input voltage.

written in C++, it can also be written in other languages that have a foreign function interface with C (e.g., C# or Python) as the library also has a C language interface.

The user first specifies the geometry of the modules, and location and orientation of each module. Here, the positions and directions of all transducers are calculated. The user then calculates the phase and amplitude of all transducers using this information, as well as the focal position, as an example. In addition to this phase and amplitude data, header data, which consist of a flag to control the hardware (e.g., an enable flag for silent mode) and modulation data if necessary, are transferred to the device.

The library currently includes utilities to generate a single focus, Bessel beam [32], multi-foci [33], and plane waves. The library can be flexibly extended to suit the needs of the user because the phase and amplitude of all transducers can be controlled by the software on the host side. This allows the user to rapidly conduct a wide range of studies.

V. Evaluation

A. Resolution of Amplitude and Phase

First, we measure whether the phase and amplitude parameters can be specified in 8 bits. In this experiment, we drive only one transducer because driving all of them saturates the microphone. The sound pressure was measured using a standard microphone (Type 4138-A-015, Brüel & Kjær) at 200 mm above the transducer.

Fig. 8 shows the measured phases of the ultrasound. The horizontal axis of Fig. 8 is an 8 bits discretized phase parameter S, where $S = 0$ corresponds to $\phi = 0$ and $S = 255$ corresponds to $\phi = 2\pi \times 255/256$. The vertical axis is the phase of the measured sound. The phase was calculated using the cross-correlation function with respect to the observed sound at $S = 0$. We observed a linear response to S, as described in eq. (6).

Fig. 9 shows the normalized sound pressure. The 40 kHz component of the input voltage applied to the transducer is also shown in Fig. 9. The horizontal axis of Fig. 9 shows the 8 bits discretized amplitude parameter D, where $D = 0$ corresponds to a duty ratio of zero, and $D = 255$ corresponds to a duty ratio of 50%. The maximum value was 26.6 Pa at $D = 255$.

The 40 kHz component of the voltage was proportional to $\sin(\pi/2 \times D/255)$, according to eq. (6). The transducer extracts this component and emits it into space. Although we expected the amplitude of the ultrasound to be proportional to this component, it was not, as shown in Fig. 9. If we assume that the amplitude is expressed by $\sin^a(\pi/2 \times D/255)$, we obtain $a = 0.803$ through a fitting (the red dashed line). The same phenomenon was also reported in [25]. The applied voltage of the transducer followed the theoretical value; thus, this was not a problem of the drive circuit but rather a characteristic of the transducer. This implies that the response of the transducer is saturated with a high duty ratio.

Although the measured amplitude differed from the theoretical value and was not linear with the input parameters, it was confirmed that both the phase and amplitude could be specified at a resolution of 8 bits.

B. Synchronization

We subsequently investigated the accuracy of the synchronization. We connected 20 modules and measured the drive signal when the ultrasound was output. Fig. 10 shows the drive signals for the 1st, 8th, 14th, and 20th of the 20 connected modules. All transducers were driven in the same phase. As

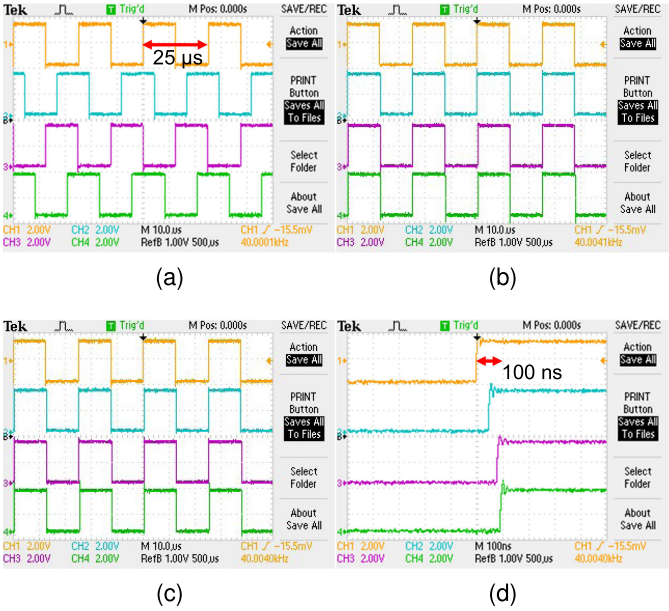


Fig. 10. Transducer driving signal of the first, eigth, 14th, and 20th among the 20 connected modules. These figures are taken (a) immediately after start-up, (b) immediately after the start of the synchronization, and (c) one hour after the start of the synchronization. (d) Enlarged view during synchronization.

shown in the figure, we confirmed that these modules were synchronized with an accuracy of approximately 100 ns.

C. Focus Generation With Nine Modules

We also measured the sound pressure field around the focal point. Nine AUTD3 modules were densely placed on the grid; three each in the row and column directions. Thus the aperture size of the array was 576×454.2. The sound pressure was measured with the standard microphone attached to the tip of the robot (M710-iC 20 L, FANUC Corp.). The xy-plane was horizontal to the surface of the array, the z-axis was perpendicular to it, and the origin was the center of the array surface. A focal point was generated at $(x, y, z) = (0, 0500)$ mm.

Fig. 11 shows the root mean square (RMS) values of the sound pressure field in the xy plane at $z = 500$ mm, where the duty ratio of the input square wave was $10/255 \times 50\%$. As expected from eq. 1, focal points of approximately twice the size of the wavelength were generated. The maximum RMS of acoustic pressure observed was 839 Pa.

D. Effect of the Number of Modules

Next, we investigated the effect of the number of modules on the focal sound pressure. The number of modules was one, four (two by two), or nine (three by three), and we placed them densely on a grid. The focus was generated at $(x, y, z) = (0, 0150), (0, 0300), (0, 0500)$ mm. The coordinate system was the same as Sec. V-C. Note that we decreased the sensitivity of the microphone using a silicon cover and then calibrated it at low sound pressure because the dynamic range of the microphone is 52.2 dB to 168 dB$(5$ kPa$)$.

Fig. 12 shows the RMS of the acoustic pressure at the focus against the duty ratio. We confirmed that the maximum sound

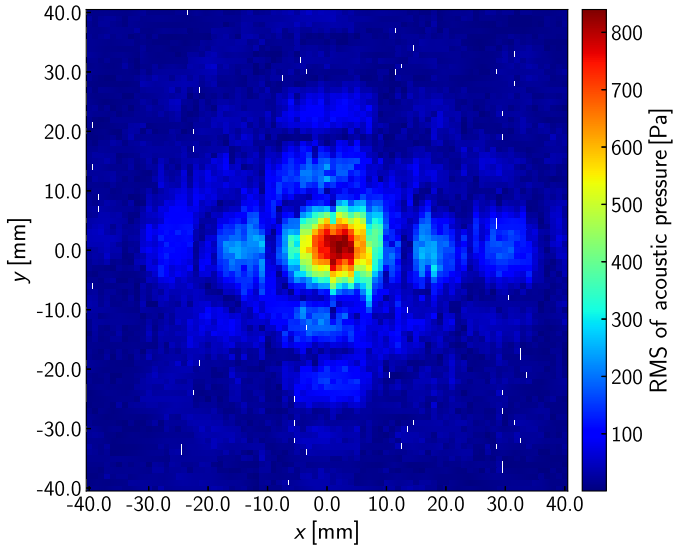


Fig. 11. Measured acoustic pressure field around a focus generated using nine modules. The duty ratio of the input square wave was $10/255 \times 50\%$.

pressure increases as the number of modules increases. We also confirmed that the sound pressure decreases as the focus becomes distant.

As shown in Fig. 12 (c), the sound pressure seems to be saturated at a certain level. This implies that the sound pressure generated by many transducers cannot be expressed by a linear superposition of that of a single transducer.

VI. Discussion

A. Phase and Amplitude Resolution

Our system specifies the phase and amplitude parameters of all transducers; hence, excessive resolutions of these parameters decrease the update rate. In addition, the resolution of the phase is also related to the synchronization margin. In the following, we discuss these necessary resolutions.

We consider the case of generating a focus. In the following, we simulate 54×42 transducers based on eq. (2). The focal point was created at 500 mm above the center of the array. Fig. 13 (a) shows the variation in sound pressure of the focus for different phase resolutions. Here, we set the ultrasound phase ϕ as follows,

$$\phi = 2\pi \frac{1}{x} \left\lfloor \frac{r}{\lambda} x \right\rfloor \tag{7}$$

where $\lfloor \cdot \rfloor$ is a floor function, x is the number of phase divisions, r is the distance between the focus and transducer, and λ is the wavelength. As shown in Fig. 13 (a), for the sound pressure, $2\pi/16$, i.e., a resolution of 4 bits is sufficient. Moreover, Fig. 13 (b) shows the calculated error between the produced maximum pressure point and the target point. Each target point was placed at $(0.1n, 0500)$ mm, $n = 0, 1, \ldots, 85$, where the origin was the center of the array and the coordinate system was set to be the as same as Sec. V-C. Fig. 13 (b) shows that we can control the focal position with a resolution of 0.1 mm if we

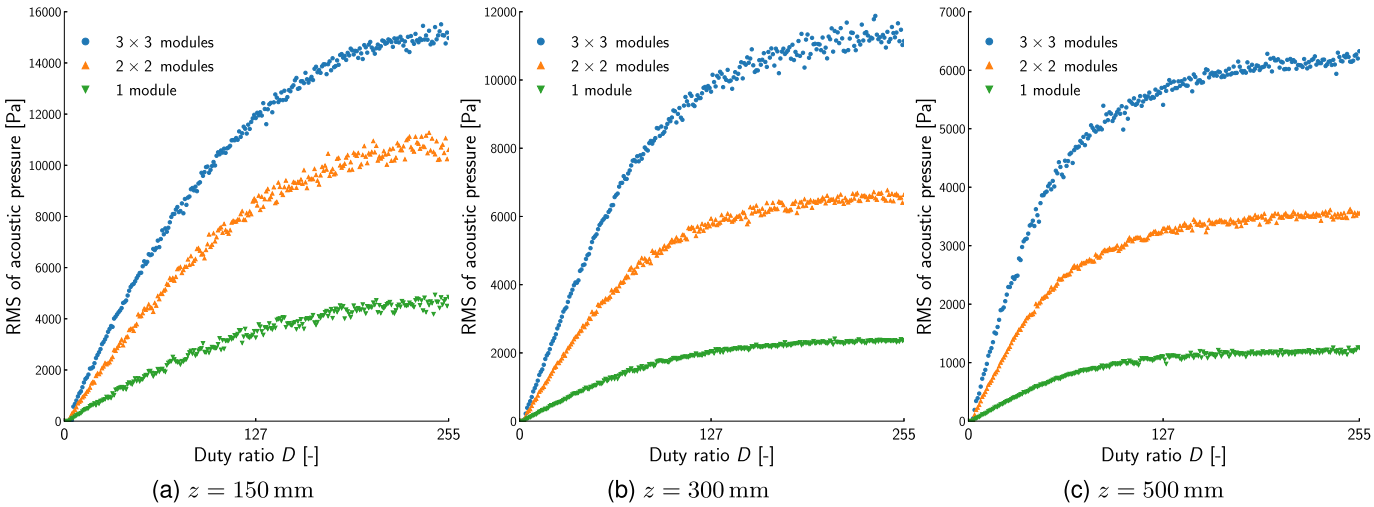


Fig. 12. Measured RMS of acoustic pressure at a focal point generated at (a) (0, 0150)mm, (b) (0, 0300)mm, and (c) (0, 0500)mm.

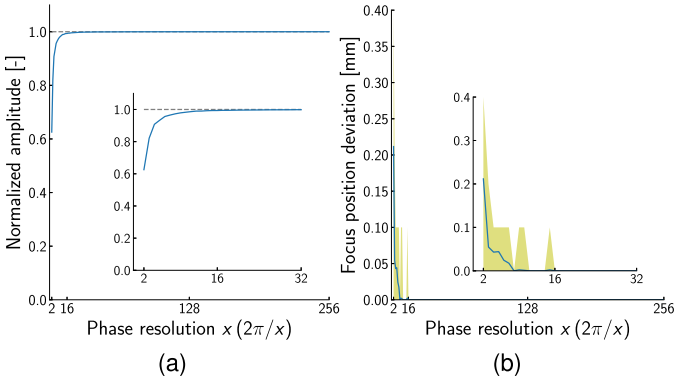


Fig. 13. (a) Amplitude of ultrasound at a focal point generated $500\ mm$ above an array with 54×42 transducers, where the horizontal axis represents a phase resolution. (b) Position error when generating focus. The solid line represents the mean and the colored area represents the existence range of an error. The subgraphs show enlarged views of from $x = 2$ to 32.

have an accuracy of 4 bits. This value is smaller than the two-point discrimination threshold and is comparable to that of the shift in threshold of a finger, which is the most sensitive part of the body. It is also much smaller the displacement amplitude of Lateral Modulation [21]. Therefore, a phase resolution of 4 bits is sufficient to control a single focus.

In the proposed system, we set the amplitude resolution to 8 bits though we have no solid ground, taking into account the margin and the convenience of implementation. In addition, we decided to set the phase resolution to 8 bits because this value is convenient for allowing the data to be in multiples of 1 byte.

B. Update Rate

The phase and amplitude refresh rate of our system depends on the master device and the number of modules connected to it. The system has no buffer for the phase or amplitude data on the hardware, while the master device specifies the phase and amplitude of all transducers. The refresh rate is limited to 1 kHz because the accuracy of the timer of a typical master, i.e., a standard PC, is approximately 1 ms.

However, this limitation is not a problem in most cases because there is a certain time constant in the phase shift. For example, when the driving phase is suddenly reversed, the amplitude of a resonant-type transducer largely decreases once and takes approximately 1 ms to recover [31]. As a result, it was reported that the maximum refresh rate for the standard 40 kHz transducer is approximately 1 kHz, except when the displacement is sufficiently smaller than the wavelength.

C. Individual Difference of Transducers

Another problem with an increase in the number of transducers is the individual differences in the transducers used. Even if the same drive signal is applied to the transducers, the phase and amplitude of the ultrasonic waves may differ depending on the individual. Herein, we discuss how this affects the focal point generation and the linear scalability of a phased array for the number of transducers.

We assume that the amplitudes A of i-th transducers follow a normal distribution independently and identically for $i = 1, 2, \ldots, n$, i.e.,

$$A_i \overset{iid}{\sim} \mathcal{N}(A, \sigma_A^2). \tag{8}$$

The same applies to the phase ϕ,

$$\phi_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\phi^2). \tag{9}$$

Furthermore, we assume that these distributions are independent. Assuming that the focal point is far and the attenuation and directivity are the same for all transducers, the linear superposition of sound pressure from all transducers can be expressed as follows:

$$p = \left\| \sum_i A_i e^{j\phi_i} \right\| \tag{10}$$

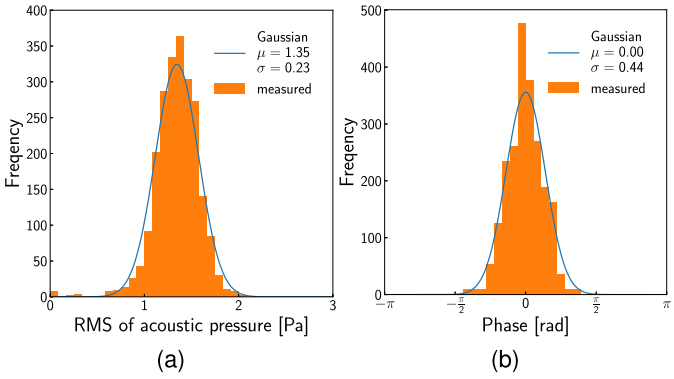


Fig. 14. Individual differences of (a) RMS of acoustic pressure, and (b) phase measured at 200 mm above the transducers. The duty ratio of the driving signal was $10/255 \times 50\%$.

Assuming that the phase variance is sufficiently small, the expected value of the sound pressure at the focal point is expressed as follows:

$$E[p] = E[A_i]E\left[\left\|\sum_i \mathrm{e}^{\mathrm{j}\phi_i}\right\|\right] \tag{11}$$

$$= AE\left[\left\|\sum_i 1 - \frac{\phi_i^2}{2}\right\|\right]. \tag{12}$$

Here, ϕ_i/σ_ϕ^2 follows a χ^2 distribution of n degrees of freedom; therefore,

$$E[p] = A\left\|n - n\frac{\sigma_\phi^2}{2}\right\| \tag{13}$$

$$= nA\left\|1 - \frac{\sigma_\phi^2}{2}\right\|. \tag{14}$$

That is, even if the transducers have an individual difference, the sound pressure at the focal point is proportional to n and we can consider that n transducers with amplitude characteristics $A\|1 - \sigma_\phi^2/2\|$.

Fig. 14 shows the measured individual difference of 2241 transducers (T4010A1) used in this paper. We measured RMS of acoustic pressure and phase at 200 mm above the transducers. The duty ratio of the driving signal was $10/255 \times 50\%$. As shown in Fig. 14. We confirmed that the individual differences in amplitude and phase of the transducer follow a normal distribution. The average RMS of acoustic pressure was 1.35 Pa, and the standard deviation of the phase was 0.44 rad. Therefore, from eq. (14), the amplitude of a focus distant from an array is expected to decrease by 9.7% even when linear superposition holds.

VII. Conclusion

In this study, we verified the characteristics of a scalable airborne ultrasound tactile display, AUTD3, which can be extended by connecting a module of 249 transducers through an Ethernet cable. Experimentally, we confirmed that the amplitude and phase of each transducer could be specified with a resolution of 8 bits. We also observed that 20 connected modules were synchronized within 0.1 μs for an hour experiment. Furthermore, by synchronizing the nine modules, a focal point of the theoretical size was observed. In addition, we confirmed that we could make acoustic pressure high by increasing the number of modules.

We can freely design a large workspace using these extensible AUTD3 modules. All modules can be synchronized at the sub-microsecond level using distributed clocks. In addition, multiple connected modules operate as a single array system and can be controlled by software on various devices with an Ethernet port, including a PC. In the case of a single module, the phase and amplitude of all transducers can be individually specified with a resolution of 8 bits under the limitation of the total throughput of 100 Mbps.

We evaluated the properties of AUTD3 in this paper. We aim to conduct various physiological investigations in haptics, which are enabled by the device. Because our system can drive a large number of transducers, it is also important to evaluate the health risks of being exposed to strong ultrasounds in practical applications. In addition, applications that generate complex sound fields at high refresh-rate (e.g. $>$ 10 kHz) have recently emerged [34]. For such applications, the communication speed (100 Mbps) is insufficient when using many AUTD3s. Because communication systems with speeds of Gbps are already spread through the market, our next challenge is to create a device that can update all transducers' phase/amplitude independently in large-scale integrated arrays.

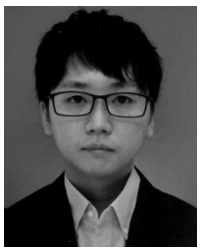
Finally, all the data and programs used in this paper are available at Open Science Framework (URL: https://osf.io/sqdg4/, DOI: 10.17 605/OSF.IO/SQDG4).

References

[1] S. Inoue, Y. Makino, and H. Shinoda, "Scalable architecture for airborne ultrasound tactile display," in *Proc. Int. AsiaHaptics Conf.*, 2016, pp. 99–103.
[2] S. Suzuki, R. Takahashi, M. Nakajima, K. Hasegawa, Y. Makino, and H. Shinoda, "Midair haptic display to human upper body," in *Proc. 57th Annu. Conf. Soc. Instrum. Control Engineers Jpn.*, 2018, pp. 848–853.
[3] P. J. Westervelt, "Parametric acoustic array," *J. Acoustical Soc. Amer.*, vol. 35, no. 4, pp. 535–537, 1963.
[4] M. Yoneyama, J.-i. Fujimoto, Y. Kawamo, and S. Sasabe, "The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design," *J. Acoustical Soc. Amer.*, vol. 73, no. 5, pp. 1532–1536, 1983.
[5] Y. Ochiai, T. Hoshi, and I. Suzuki, "Holographic whisper: Rendering audible sound spots in three-dimensional space by focusing ultrasonic waves," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2017, pp. 4314–4325.
[6] S. A. Seah, B. W. Drinkwater, T. Carter, R. Malkin, and S. Subramanian, "Correspondence: Dexterous ultrasonic levitation of millimeter-sized objects in air," *IEEE Trans. Ultrasonics, Ferroelect., Freq. Control*, vol. 61, no. 7, pp. 1233–1236, Jul. 2014.
[7] A. Marzo, M. Caleap, and B. W. Drinkwater, "Acoustic virtual vortices with tunable orbital angular momentum for trapping of mie particles," *Phys. Rev. Lett.*, vol. 120, no. 4, 2018, Art. no. 044301.
[8] Y. Ochiai, T. Hoshi, and J. Rekimoto, "Pixie dust: Graphics generated by levitated and animated objects in computational acoustic-potential field," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–13, 2014.
[9] S. Inoue, S. Mogami, T. Ichiyama, A. Noda, Y. Makino, and H. Shinoda, "Acoustical boundary hologram for macroscopic rigid-body levitation," *J. Acoustical Soc. Amer.*, vol. 145, no. 1, pp. 328–337, 2019.
[10] A. Marzo, S. A. Seah, B. W. Drinkwater, D. R. Sahoo, B. Long, and S. Subramanian, "Holographic acoustic elements for manipulation of levitated objects," *Nature Commun.*, vol. 6, 2015, Art. no. 8661.

[11] R. Morales, A. Marzo, S. Subramanian, and D. Martínez, "Leviprops: Animating levitated optimized fabric structures using holographic acoustic tweezers," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, 2019, pp. 651–661.
[12] R. Hirayama, D. M. Plasencia, N. Masuda, and S. Subramanian, "A volumetric display for visual, tactile and audio presentation using acoustic trapping," *Nature*, vol. 575, no. 7782, pp. 320–323, 2019.
[13] T. Furumoto, K. Hasegawa, Y. Makino, and H. Shinoda, "Three-dimensional manipulation of a spherical object using ultrasound plane waves," *IEEE Robot. Automat. Lett.*, vol. 4, no. 1, pp. 81–88, Jan. 2019.
[14] T. Furumoto, M. Fujiwara, Y. Makino, and H. Shinodas, "Baluna: Floating balloon screen manipulated using ultrasound," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR).* , 2019, pp. 937–938.
[15] T. Morisaki *et al.*, "Hopping-pong: Changing trajectory of moving object using computational ultrasound force," in *Proc. ACM Int. Conf. Interactive Surfaces Spaces*, 2019, pp. 123–133.
[16] D. Dalecki, S. Z. Child, C. H. Raeman, and E. L. Carstensen, "Tactile perception of ultrasound," *J. Acoustical Soc. Amer.*, vol. 97, no. 5, pp. 3165–3170, 1995.
[17] T. Iwamoto, M. Tatezono, and H. Shinoda, "Non-contact method for producing tactile sensation using airborne ultrasound," in *Proc. Int. Conf. Hum. Haptic Sens. Touch Enabled Comput. Appl.*, 2008, pp. 504–513.
[18] T. Hoshi, M. Takahashi, T. Iwamoto, and H. Shinoda, "Noncontact tactile display based on radiation pressure of airborne ultrasound," *IEEE Trans. Haptics*, vol. 3, no. 3, pp. 155–165, Jul.-Sep. 2010.
[19] T. Carter, S. A. Seah, B. Long, B. Drinkwater, and S. Subramanian, "Ultrahaptics: Multi-point mid-air haptic feedback for touch surfaces," in *Proc. 26th Annu. ACM Symp. User Interface Softw. Technol.*, 2013, pp. 505–514.
[20] G. Korres and M. Eid, "Haptogram: Ultrasonic point-cloud tactile stimulation," *IEEE Access*, vol. 4, pp. 7758–7769, 2016.
[21] R. Takahashi, K. Hasegawa, and H. Shinoda, "Tactile stimulation by repetitive lateral movement of midair ultrasound focus," *IEEE Trans. Haptics*, vol. 13, no. 2, pp. 334–342, Apr.–Jun. 2020.
[22] W. Frier *et al.*, "Using spatiotemporal modulation to draw tactile patterns in mid-air," in *Proc. Int. Conf. Hum. Haptic Sens. Touch Enabled Comput. Appl.* , 2018, pp. 270–281.
[23] M. Nakajima, K. Hasegawa, Y. Makino, and H. Shinoda, "Remotely displaying cooling sensation via ultrasound-driven air flow," in *Proc. IEEE Haptics Symp.*, 2018, pp. 340–343.
[24] T. Kamigaki, S. Suzuki, and H. Shinoda, "Noncontact thermal and vibrotactile display using focused airborne ultrasound," in *Proc. Int. Conf. Hum. Haptic Sens. Touch Enabled Comput. Appl.*, 2020, pp. 271–278.
[25] K. Hasegawa and H. Shinoda, "Aerial vibrotactile display based on multiunit ultrasound phased array," *IEEE Trans. Haptics*, vol. 11, no. 3, pp. 367–377, Jul.–Sep. 2018.
[26] B. Long, S. A. Seah, T. Carter, and S. Subramanian, "Rendering volumetric haptic shapes in mid-air using ultrasound," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–10, 2014.
[27] A. Marzo, T. Corkett, and B. W. Drinkwater, "Ultraino: An open phased-array system for narrowband airborne ultrasound transmission," *IEEE Trans. Ultrasonics, Ferroelect., Freq. Control*, vol. 65, no. 1, pp. 102–111, Jan. 2018.
[28] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. Hester, "Atmospheric absorption of sound: Further developments," *J. Acoustical Soc. America*, vol. 97, no. 1, pp. 680–683, 1995.
[29] S. Bolanowski, G. Gescheider, R. Verrillo, and C. Checkosky, "Four channels mediate the mechanical aspects of touch," *J. Acoust. Soc. Amer.*, vol. 84, pp. 1680–1694, 12 1988.
[30] *Beckhoff Automat. GmbH. EtherCAT - Ethernet Fieldbus.* Accessed on: Jun. 1, 2020. [Online]. Available: https://www.ethercat.org/en/technology.html
[31] S. Suzuki, M. Fujiwara, Y. Makino, and H. Shinoda, "Reducing amplitude fluctuation by gradual phase shift in midair ultrasound haptics," *IEEE Trans. Haptics*, vol. 13, no. 1, pp. 87–93, Jan.–Mar. 2020.
[32] K. Hasegawa, L. Qiu, A. Noda, S. Inoue, and H. Shinoda, "Electronically steerable ultrasound-driven long narrow air stream," *Appl. Phys. Lett.*, vol. 111, no. 6, 2017, Art. no. 064104.
[33] S. Inoue, Y. Makino, and H. Shinoda, "Active touch perception produced by airborne ultrasonic haptic hologram," in *Proc. IEEE World Haptics Conf.*, 2015, pp. 362–367.
[34] D. M. Plasencia, R. Hirayama, R. Montano-Murillo, and S. Subramanian, "Gs-pat: High-speed multi-point sound-fields for phased arrays of transducers," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 138–1, 2020.

Shun Suzuki received the B.S. degree in physics from Waseda University, Tokyo, Japan, in 2017, and the M.S. degree, in 2019, in complexity science and engineering from The University of Tokyo, Tokyo, Japan, where he is currently working toward the Ph.D. degree. His research interests include mid-air haptics.



Seki Inoue received the B.S. degree in engineering and the Ph.D. degree in information science and technology from the University of Tokyo, Tokyo, Japan, in 2013 and 2018, respectively. He is currently a Project Researcher with the Graduate School of Frontier Sciences, The University of Tokyo, and a Top Data Scientist with Adways Inc., Tokyo, Japan. His research interests include acoustic levitation, machine learning, haptics, and optics. From 2015 to 2016, he was a Research Fellow with the Japan Society for the Promotion of Science.

Masahiro Fujiwara (Member, IEEE) received the B.S. degree in engineering, the M.S. and Ph.D. degrees in information science and technology from The University of Tokyo, Tokyo, Japan, in 2010, 2012, and 2015, respectively. He is currently a Project Assistant Professor with the Graduate School of Frontier Sciences, The University of Tokyo. His research interests include information physics, haptics, non-contact sensing, and application systems related to them.

Yasutoshi Makino received the Ph.D. degree in information science and technology from The University of Tokyo, Tokyo, Japan, in 2007. He is currently an Associate Professor with the Department of Complexity Science and Engineering, The University of Tokyo. He was a Researcher for two years with The University of Tokyo and from 2009 to 2013, he was an Assistant Professor with Keio University, Tokyo, Japan. From 2013, he moved to The University of Tokyo as a Lecturer, and he is an Associate Professor from 2017. His research interests include haptic interactive systems.

Hiroyuki Shinoda (Member, IEEE) received the B.S. degree in applied physics, the M.S. degree in information physics, and the Ph.D. degree in engineering from The University of Tokyo, Tokyo, Japan, in 1988, 1990, and 1995, respectively. He is currently a Professor with the Graduate School of Frontier Sciences, The University of Tokyo. He was an Associate Professor from 1995 with the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, Fuchu, Japan. After a period at UC Berkeley as a visiting scholar in 1999, he was an Associate Professor with the University of Tokyo from 2000 to 2012. His research interests include information physics, haptics, mid-air haptics, two-dimensional communication, and their applications. He is also a Member of the IEEJ, RSJ, JSME, and ACM.