Development of Aerial-Input and Aerial-Tactile-Feedback System

Takayuki Hoshi

Kumamoto University

ABSTRACT

A system is proposed which enables users not only to operate computers by moving their hands in mid-air but also to feel touch feedback on their hands. The system consists of a PC, a marker-less hand-tracker for input, and a non-contact tactile display for feedback. The tactile display utilizes ultrasound to produce tactile stimulation from a distance. It is based on a nonlinear phenomenon of ultrasound (acoustic radiation pressure) and the phased array focusing technique. The principles are described and the prototype system is introduced.

KEYWORDS: Natural user interface, tactile display, airborne ultrasound, acoustic radiation pressure.

INDEX TERMS: K.5.2 [Information Interfaces and Presentation]: User Interfaces—Haptic I/O; K.5.2 [Information Interfaces and Presentation]: User Interfaces—Input Devices and Strategies

1 INTRODUCTION

Recently, various technologies are developed which enable users to operate computers by moving their hands in air. Combination of an IR-LED illuminator, a retro-reflector attached on a user's finger, and an IR camera is demonstrated in [1]. aeroTAP [2] extracts a user's finger from the images captured by a normal webcam. A high-speed camera is used to track a user's finger in [3]. Mgestyk [4] and Kinect [5] utilize depth cameras to sense the position of a user's hand and recognize gestures. These natural user interfaces (NUI) would be used as interfaces with PCs, games, portable devices, digital signage, and so on. They are also promising in hospitals, food factories, and public spaces because the interfaces keep users' hands away from getting dirty. For the technologies, together with conventional visual and/or audio feedback, tactile feedback is required in order to provide an enhanced and more intuitive user experience.

One of the requirements for a tactile display to be combined with NUI is allowing users to move their arms, hands, and fingers freely. It is most preferable that users are not required to touch on or hold any devices, unlike general tactile displays [6]. There are three types of conventional strategies for tactile feedback in a free space. The first is attaching tactile devices on user's fingers and/or palms. Employed devices are, for example, vibrotactile stimulators (CyberTouch [7]), motor-driven belts (GhostGlove [8]), or pin-array units (SaLT [9]). In this strategy, the skin and the device are always in contact and that leads to undesired touch feelings. The second is controlling the positions of tactile devices so that they contact with the skin only when tactile feedback is required. In the master-slave system shown in [10], the encounter-

Address: Rm. 613, Research Bldg. of Faculty of Engineering I, 2-39-1, Kurokami, Kumamoto-shi, Kumamoto-ken, Japan
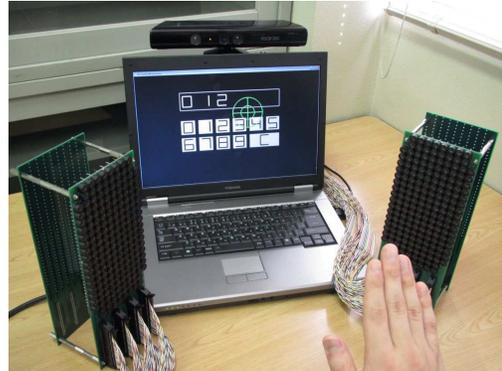E-mail: star@kumamoto-u.ac.jp

Figure 1. Developed aerial-input and aerial-tactile-feedback system. It consists of a PC, a depth camera, and an ultrasound-based tactile display.

type force feedback is realized by the exoskeleton master hand. The detailed tactile feedback for each finger is provided by the electrotactile display attached on the finger part of the master hand. The drawback of this strategy is that it requires bulky robot arms. The last is providing tactile feedback from a distance without any direct contact. For example, air jets are utilized in [11] to realize non-contact force feedback. Fans or air cannons are used in theme parks to amaze visitors. Although they are effective for rough "force" feedback, their spatial and temporal properties are quite limited and they cannot provide detailed "tactile" feedback. The reachable distance of an air jet is determined by the diameter and the velocity of the jet stream, which results in the trade-off relationship between the spatial resolution of the pressure on the skin and the distance from the device to the skin.

This paper proposes to combine a non-contact tactile display with NUI. The tactile display uses airborne ultrasound for tactile stimulation which makes people free from any bothering tactile devices to wear or touch. The usage of airborne ultrasound for a tactile display was first introduced in [12] and the feasibility was examined using the prototype consisting of 91 ultrasound transducers that generates a fixed focal point. After that, an upgraded version consisting of 324 transducers was developed and its output force was 16 mN at the maximum [13]. It was confirmed that the prototype could provide tactile cues effective for NUI.

As for hand-tracking, a depth camera is newly utilized in this paper, instead of a retro-reflective marker, infrared illuminators, and IR cameras used in the previous works of our research group. Owing to the collaboration of the two technologies, people can interact with the system with their *bare hands*.

The following paper outlines, first, the principles and the current prototype of the ultrasound-based tactile display. Next, the developed aerial-input and aerial-tactile-feedback system (Fig. 1) is introduced with the experimental results in Section 3. After Section 4 discusses the points of concern in the proposed system, finally, Section 5 concludes this paper.
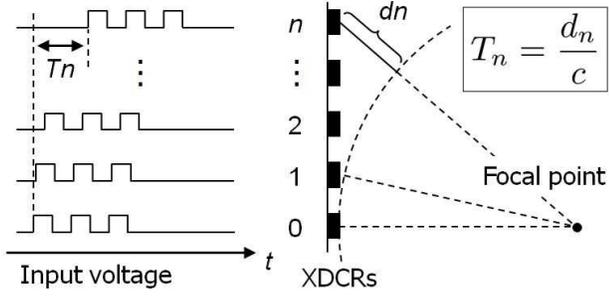
Figure 2. Phased array focusing. The time offset T_n [s] of the n-th transducer is determined based on the extra distance d_n [m] and the sound speed c [m/s].
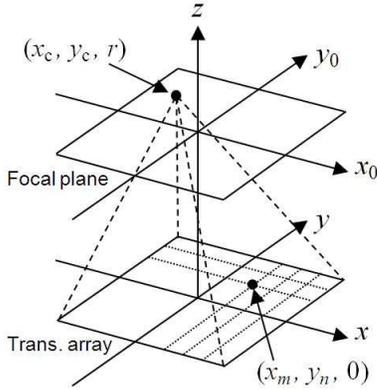

Figure 3. Coordinate system. The ultrasound is radiated from the transducer array and focused at (x_c, y_c) on the focal plane.

2 AIRBORNE ULTRASOUND TACTILE DISPLAY

2.1 Acoustic Radiation Pressure

The tactile display is based on a nonlinear phenomenon of ultrasound: Acoustic radiation pressure [14][15]. Assuming a plane wave, the acoustic radiation pressure P [Pa] is described as

$$P = \alpha \frac{I}{c} = \alpha \frac{p^2}{\rho c^2} \qquad (1)$$

where I [W/m^2] is the sound intensity of ultrasound, c [m/s] is the sound speed, p [Pa] is the RMS sound pressure of ultrasound, and ρ [kg/m^3] is the density of medium. α is the constant ranging from 1 to 2 depending on the amplitude reflection coefficient R at an object surface; $\alpha \equiv 1+R^2$. If the object surface reflects the incident ultrasound perfectly, $\alpha = 2$, while if it absorbs the entire incident ultrasound, $\alpha = 1$. In case that the ultrasound beam is reflected vertically at the object surface, the surface is subjected to the constant vertical force in the direction of the incident beam. Equation (1) suggests that the spatial distribution of the radiation pressure P can be controlled by synthesizing the spatial distribution of the ultrasound p.

2.2 Phased Array Focusing

To produce the radiation pressure perceivable by human skins, the Phased Array Focusing technique (Fig. 2) is used. The focal point of ultrasound is generated by controlling the phase delays of multiple transducers. That technology is widely used in sonography, optics, astronomy, and various radar systems. It is also employed to generate a localized audible sound spot in a 3D space [16]. The radiation pressure at the focal point is proportional to the square of the number of transducers according to (1). For example, the RMS sound pressure $p_{0.3}$ at the distance of 300 mm radiated from a single typical transducer (T4010A1 [17], Nippon Ceramic Co. Ltd.) is 121.5 dB SPL (= 24 Pa) at the normal-rated voltage (10 Vrms), and the resulting radiation pressure is 8.3×10^{-3} Pa. If we drive 100 transducers so that the phases of all the ultrasound coincide at a point, the radiation pressure as large as 83 Pa is generated there.

Here we theoretically derive the resulting sound pressure field on the focal plane generated by an ultrasound transducer array. While the same theory was applied to an $N \times N$ array (i.e. quadrate) in [13], it is applied to an $M \times N$ array in this paper. First, the specifications of the array are listed. The diameter of the transducer housing and the radius of the diaphragm are $d = 10$ mm and $a = 4$ mm, respectively. The resonant frequency is 40 kHz. While the transducer has directivity (its half-amplitude full angle is 100 deg), we assume a spherical sound wave for simplicity. We use $M \times N$ pieces of the transducers and arrange them into an $Md \times Nd$ [m^2] rectangle.

Next, the sound pressure field is formulated. We take the coordinate system shown in Fig. 3. Let r [m] be the focal length. The RMS sound pressure p_r from each transducer on the focal plane is in inverse proportion to r. The phase delay is assumed to be adequately determined so that the focal point is generated at the position (x_c, y_c, r). Then, assuming that the RMS sound pressure from each transducer is constant p_r anywhere on the focal plane, the resulting sound pressure field $p(x_0, y_0)$ is written as

$$p(x_0, y_0) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \sqrt{2}\, p_r e^{-jkr'} e^{j(kr''-\omega t)}$$

$$\approx \sqrt{2}\, p_r MN \frac{\mathrm{sinc}\left(\dfrac{Mdv_x}{2}, \dfrac{Ndv_y}{2}\right)}{\mathrm{sinc}\left(\dfrac{dv_x}{2}, \dfrac{dv_y}{2}\right)} e^{j\{\varphi(x_0,y_0)-\omega t\}} \quad (2)$$

where

$$r' \equiv \sqrt{(x_m-x_c)^2+(y_n-y_c)^2+r^2} \approx r + \frac{(x_m-x_c)^2+(y_n-y_c)^2}{2r}, \quad (3)$$

$$r'' \equiv \sqrt{(x_m-x_0)^2+(y_n-y_0)^2+r^2} \approx r + \frac{(x_m-x_0)^2+(y_n-y_0)^2}{2r}. \quad (4)$$

Here r' is the distance from the m-th row and n-th column transducer to the focal point, and r'' is the distance from the transducer to the arbitrary point on the focal plane. In going from the first line to the second line of (2), the Fresnel approximation [18] is applied to r' and r'' as shown in (3) and (4). (x_m, y_n) is written as $(md+\xi, nd+\xi)$ where ξ is an offset of the transducer position. $\exp(-jkr')$ is the phase control factor to focus the ultrasound and $\sqrt{2}\, p_r \exp\{j(kr''-\omega t)\}$ is the spherical wave from the individual transducer. j is the imaginary unit and k [rad/m] is the wavenumber. The function sinc is defined as $\mathrm{sinc}(x, y) \equiv \sin(x)\sin(y)/xy$. Transformation of variables is done as $v_x \equiv (k/r)(x_0-x_c)$ and $v_y \equiv (k/r)(y_0-y_c)$.

$$\varphi(x_0, y_0) \equiv \frac{k}{2r}\left(x_0^2+y_0^2-x_c^2-y_c^2\right) - \frac{\xi+(M-1)d}{2}v_x - \frac{\xi+(N-1)d}{2}v_y \quad (5)$$
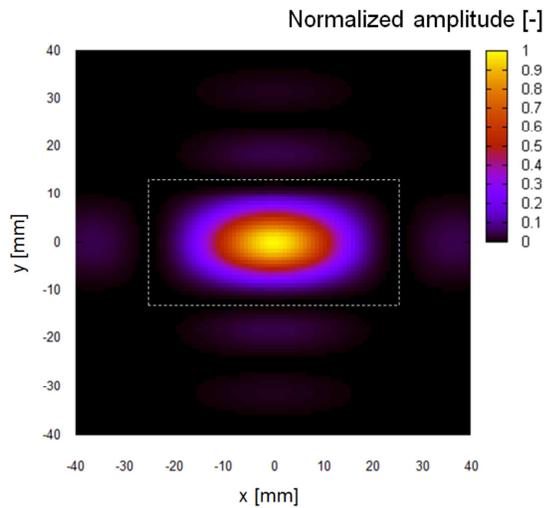
Figure 4. Simulated distribution of focused radiation pressure. Here the size of the transducer array is supposed to be 100 mm and 200 mm in X and Y directions, respectively. The dotted line indicates the edge of the main lobe (focal point).
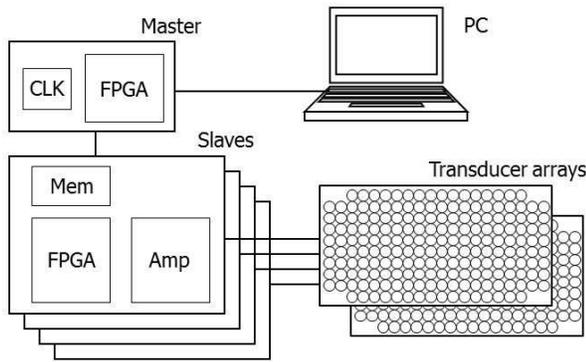

Figure 5. Block diagram of tactile display. Each slave circuit drives 96 ultrasound transducers.

is a resulting phase delay.

Equation (2) indicates that the spatial distribution of ultrasound generated from a rectangular array is nearly sinc-function-shaped. It was confirmed that the theory predicts well the distribution of the radiation pressure generated by the prototype device in [13]. The sizes of the edge of the main lobe (i.e. the diameters of the focal point) in X and Y directions are given by $4\pi r/kMd$ and $4\pi r/kNd$, respectively. For example, the diameters in X and Y directions are 51 mm and 26 mm, respectively, when $r = 300$ mm, $Md = 100$ mm, and $Nd = 200$ mm. Fig. 4 shows the theoretical distribution of the radiation pressure normalized so that the maximum value is equal to 1.0. The main lobe is accompanied by the side lobes whose normalized amplitudes are about 0.05 or less. The dotted line outlines the edge of the main lobe, which consists of the zeros surrounding the main lobe. Because the width of 51 mm is too wide to target a single finger, multi fingers or a palm is the target of the prototype.
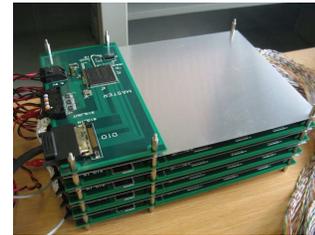
2.3 Tactile Display

Two ultrasound transducer arrays are used in this research. Each array has 192 transducers that are arranged into a 100×200 mm^2 rectangle area. All the transducers on the two arrays are driven by
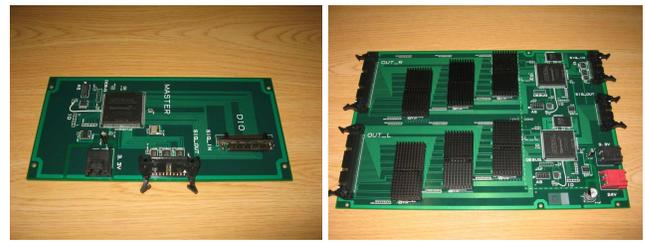
(a) Master (200×100 mm^2)   (b) Slave (290×200 mm^2)

(c) Circuits in use. Metal plates are inserted to shut off electric noise.

Figure 6. Photos of driving circuits newly developed.

the driving circuits so that a single focal point is generated. The output force of each array is 12 mN (measured).

Figs. 5 and 6 show the block diagram of the tactile display and the photos of the developed circuits, respectively. The display consists of a PC with a digital I/O PC-card, a master circuit, four slave circuits, and the transducer arrays. The master circuit has an FPGA and a 25.6-MHz oscillator which acts as the system clock. Each slave circuit has two FPGAs, two flash memories, and 96-ch amplifiers. The driving circuits are the revised version of the circuits fabricated in [13]. While the previous prototype is 22 cm^2 per transducer, the new one is 6 cm^2 per transducer.

It is described here how the driving circuits control the phase and amplitude. One cycle of 40-kHz rectangular wave is divided into 16 segments (i.e. 1.5625 μs). The phase is controlled by the position of a HIGH (= 24 V) period within the 16 segments, and the amplitude by the duration of the HIGH period (PWM). That is, the phase and amplitude are quantized in 4 and 3 bit, respectively.

The pressure is controlled not only spatially but also temporally. It can be modulated by rectangular wave whose frequency is ranging from 1 to 1000 Hz and the duty ratio is 50 percent. The frequency is quantized in 5 bit so that the selectable values are equally spaced on the logarithmic frequency scale.

The driving circuits operate as follows. First, the PC sends a command to the master circuit via the digital I/O. The command includes the target position of the focal point, the amplitude of ultrasound, and the modulation frequency. Second, the master broadcasts it to all the slaves. The flash memories on the slaves store the pre-calculated look-up tables of the phases. According to the command, each slave loads the corresponding phases from the memory, generates the wave patterns at the directed amplitude and modulation frequency, and drives 96 transducers individually through the amplifiers. The driving signal into the transducer is a 24-Vp-p, 40-kHz, and rectangular wave whose DC component is cut by an HPF.

The target area is limited to a $200\times200\times200$ mm^3 cubic area and it is divided into $5\times5\times12.5$ mm^3 sub-areas. The center positions of the sub-areas are selectable as the focal position. That is, the focal point moves among $40\times40\times16$ discrete positions. The amplitudes of all the transducers are tentatively fixed at the maximum value.
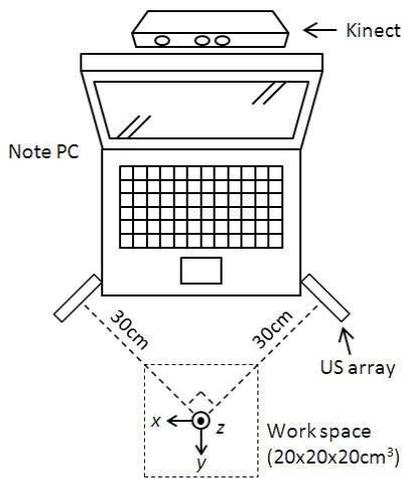
Figure 7. Arrangement of prototype system (top view). Note that the coordinate system is different from Fig. 3.

The time-averaged power consumption of the current system is about 150 W. According to the results in [13], 384 transducers would consume about 40 W in outputting ultrasound. That means about 2/3 of the energy is wasted as heat in the amplifier ICs. That is an issue to be solved in future.

3 AERIAL INPUT SYSTEM

3.1 Overview

An aerial-input system was developed by employing the airborne ultrasound tactile display (Figs. 1 and 7). The system consists of a laptop PC, the developed tactile display, and a hand-tracker. A circular cursor on the LCD of the laptop PC moves according to the movements of a user's hand in X and Z directions. The size of the cursor changes according to the hand movement in Y direction. The virtual touch panel is set at XZ plane (i.e. $y = 0$) and tactile feedback is provided when the user's hand is within the space where $y \leq 0$. With this system, users can click, drag, draw, select, and so on, receiving not only visual but also tactile feedback. The refresh rate of the current system is 30 Hz.

Kinect [5] is used as a depth camera in this research. It is based on an infrared pattern projected over its field of vision. The spatial resolutions are about 1 mm in all X, Y, and Z directions at a distance of 600 mm from the sensor. Currently, the nearest point and its vicinity are detected as a user's hand. An algorithm of bone estimation would be adopted to improve robustness of hand tracking.

3.2 Experiment

An experiment was conducted in order to examine how surely users could feel the position of a virtual touch screen. Ten volunteers (between 21 and 30 years old, all male, and right-handed) took part in it. The subject was instructed to sit on a chair in front of the aerial-input system, move his hand toward the PC slowly (Fig. 8), and stop his hand when he felt tactile stimulation on his hand. Then the position of his hand was recorded. After that, he put down his hand and started the next trial. The speed of hand was not controlled but 10 cm/s or less. He repeated the trial 10 times. The modulation frequency was 200 Hz. The amplitude was set to the maximum. The visual information was shut off by closing his eyes and the auditory information was blocked off by hearing a white noise with headphones.
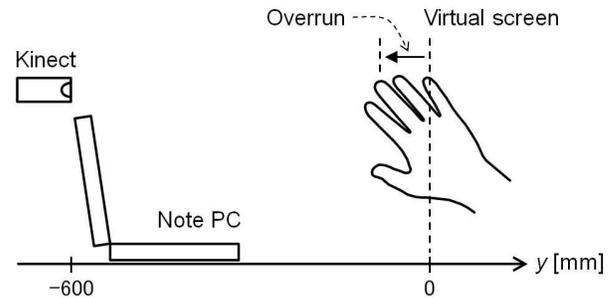

Figure 8. Settings of experiment. Note that the transducer arrays are not illustrated in this figure for simplicity.
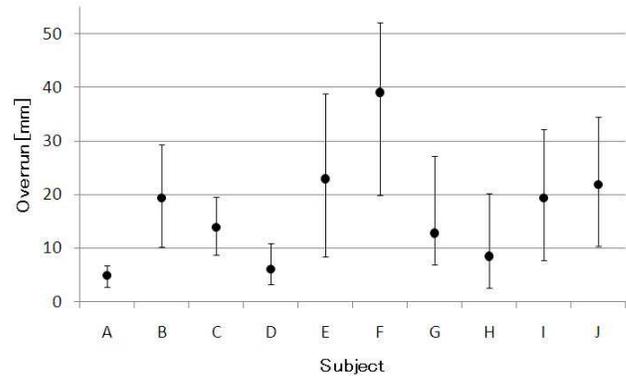

Figure 9. Experimental results. The black dot is the mean value for each subject, and the bar connects the maximum and minimum values.

The results are shown in Fig. 9. The overrun (vertical axis) is the distance travelled along Y direction after the subject's hand passed the XZ plane. In the ideal case, the overrun would be zero because the virtual touch screen is set at the XZ plane. Fig. 9 shows the mean value as a black dot and the maximum and minimum values as both ends of a vertical bar for each subject. The mean value is the offset of position recognition and it can be compensated for each person. The difference between the maximum and minimum values indicates the degree of ambiguity. The offset and ambiguity averaged among the trials of all the subjects are 17 mm and 19 mm, respectively. That ambiguity implies that an interval wider than 19 mm is required when multi layered screen is reproduced, for example.

Note that the amplitude was set at the maximum value and the information which the subjects received was 1 bit (ON/OFF) in this experiment. There are other possible ways to control the amplitude. For example, the amplitude is controlled so that it is in proportion to the overrun distance for an application of drawing with a brush or pressing a soft material. Providing more information by amplitude control is investigated in future. At that time, increasing the maximum output force of the tactile display might be also required.

4 DISCUSSIONS

There are two points of concern in the current arrangement of the aerial-input system. One is the trade-off relationship between the sizes of the whole system and the focal point, and the other is the relative configuration of the hand surface and the transducer array.

The diameter of the focal point is in inverse proportion to the size of the array, as formulated in Section 2.2, just like a wider
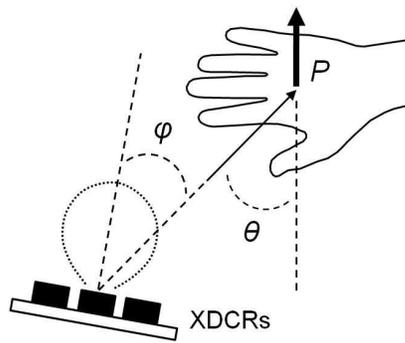
Figure 10. Relative configuration of hand and transducers. The incident angle and the directivity should be considered.

aperture lens achieves a higher resolution. In this research, the size of the transducer array is determined as 100×200 mm^2 so that the whole system is not so space-consuming. That results in the diameters of the focal point which are 5.1 mm and 2.6 mm in horizontal and vertical directions, respectively (Fig. 4). The horizontal diameter is too wide to stimulate a single finger. Before this research, the hexagonal and quadrate arrays were used, and so such an issue of unevenness of an array was not obvious. That issue is to be considered when one intends to mount the transducer arrays onto, for example, the frame of the laptop PC.

The other point is the angle between the hand and the array affects the intensity of the resulting force (Fig. 10). When the ultrasound wave incidents on the hand surface at the incident angle θ [rad], the resulting force is perpendicular to the hand surface and it is in proportion to $\cos\theta$ [14]. Besides, the directivity of the ultrasound transducer [17] affects the force. Based on the far-field theory, the directivity of ultrasound from a piston is represented as $2J_1(ka\sin\varphi)/ka\sin\varphi$, where J_1 is the Bessel function of the first kind of order 1, k is the wavenumber, a is the radius of the diaphragm of the transducer, and φ is the angle from the axial direction. In this research, without any optimization, θ and φ are set to be $\pi/4$ rad and 0 rad, respectively. Then the resultant force normal to the hand produced by two transducer arrays is 12 mN × $\cos(\pi/4)$ × 2 = 17 mN. Before this research, the hand was supposed to be parallel to (i.e. $\theta = 0$ rad) and just above (i.e. $\varphi = 0$ rad) the transducer array and there was no need to take those angles into account. In in designing the tactile display using multi arrays, it is better to consider these angles.

5   CONCLUSION

In this paper, an aerial-input system with aerial tactile feedback was presented. It enables users to operate computers feeling not only visual and audio feedback but also tactile feedback directly on their hands. An ultrasound-based tactile display is used for tactile feedback and a depth camera is used for hand-tracking. Owing to the two technologies, people can use the aerial-input system with their bare hands.

The tactile feedback is based a nonlinear effect of ultrasound: Acoustic radiation pressure. Controlling phase delays to generate a focal point was formulated based on the Fresnel (or near-field) diffraction. After that, the current version of the tactile display was introduced. It can produce vibrations up to 1 kHz. The diameters of the focal point in the horizontal and vertical directions are 51 mm and 26 mm, respectively.

An experiment was conducted to examine how surely users could feel the produced tactile stimulation. As a result, it was confirmed that people could feel the tactile stimulation rightly around the intended position and they could localize it within 19 mm in the depth direction.

Finally, the points of concern in the proposed arrangement were pointed out: one is the trade-off relationship between the sizes of the ultrasound transducer array and the focal point, and the other is the relative configuration of the hand and the array.

REFERENCES

[1]   WiimoteProject, http://johnnylee.net/projects/wii/.

[2]   aeroTAP, http://www.aerotap.net/?en.

[3]   T. Niikura, Y. Hirobe, A. Cassinelli, Y. Watanabe, T. Komuro, M. Ishikawa: In-air Typing Interface for Mobile Devices with Vibration Feedback, Proc. SIGGRAPH 2010, Emerging Technologies, article no. 15, 2010.

[4]   Mgestyk, http://www.mgestyk.com/.

[5]   Kinect, http://www.xbox.com/kinect.

[6]   M.B.Khoudja, M.Hafez, J.M.Alexandre, and A. Kheddar: Tactile Interfaces: A State-of-the-Art Survey, Proc. 35th International Symposium on Robotics (ISR 2004), 2004.

[7]   CyberTouch, http://www.est-kl.com/products/data-gloves/ cyberglove-systems/cybertouch.html.

[8]   K. Minamizawa, S. Kamuro, S. Fukamachi, N. Kawakami, and S. Tachi: GhostGlove: Haptic Existence of the Virtual World, Proc. 35th International Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH 2008), New Tech Demos, article no. 18, 2008.

[9]   S.-C. Kim, C.-H. Kim, T.-H. Yang, G.-H. Yang, S.-C. Kang, and D.-S. Kwon: SaLT: Small and Lightweight Tactile Display Using Ultrasonic Actuators, Proc. 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2008), pp. 430-435, 2008.

[10] K. Sato, K. Minamizawa, N. Kawakami, and S. Tachi: Haptic Telexistence, Proc. 34th International Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH 2007), Emerging Technologies, article no. 10, 2007.

[11] Y. Suzuki and M. Kobayashi: Air Jet Driven Force Feedback in Virtual Reality, IEEE Computer Graphics and Applications, vol. 25, pp. 44-47, 2005.

[12] T. Iwamoto, M. Tatezono, and H. Shinoda: Non-contact Method for Producing Tactile Sensation Using Airborne Ultrasound, Proc. EuroHaptics 2008, pp. 504-513, 2008.

[13] T. Hoshi, M. Takahashi, T. Iwamoto, and H. Shinoda: Noncontact Tactile Display Based on Radiation Pressure of Airborne Ultrasound, IEEE Transactions on Haptics, vol. 3, no. 3, pp. 155-165, 2010.

[14] J. Awatani: Studies on Acoustic Radiation Pressure. I. (General Considerations), Journal of the Acoustical Society of America, vol. 27, pp. 278-281, 1955.

[15] T. Hasegawa, T. Kido, T. Iizuka, and C. Matsuoka: A General Theory of Rayleigh and Langevin Radiation Pressures, Acoustical Science and Technology, vol. 21, no. 3, pp. 145-152, 2000.

[16] K. Shinagawa, Y. Amemiya, H. Takemura, S. Kagami, and H. Mizoguchi: Three Dimensional Simulation and Measurement of Sound Pressure Distribution Generated by 120 ch Plane Loudspeaker Array, Proc. 2007 IEEE International Conference on Systems, Man and Cybernetics (SMC '07), pp. 278-283, 2007.

[17] Ultrasound Transducer, http://www.nicera.co.jp/pro/ut/pdf/T4010A1 (ENG).pdf.

[18] F. Träger: Springer Handbook of Lasers and Optics, Springer, 2007.

(All URLs were last accessed on 30 March 2011.)