3D6-1
Invited

Paradigm shift from high-speed single-point to high-speed multipoint algorithms for multimodal acoustic displays

Ryuji Hirayama[1], Diego Martinez Plasencia[1], Roberto Montano-Murillo[1], and Sriram Subramanian[1]

r.hirayama@ucl.ac.uk
[1]Multi-Sensory Devices Group, University College London, 66-72 Gower Street London, WC1E 6EA, UK
Keywords: Acoustic levitation, Volumetric display, Multimodal display, Mixed reality.

ABSTRACT

High-speed ultrasound-field control using phased arrays of transducers (PATs) have allowed creating interactive multimodal (i.e. visual, haptics and audio) displays in mid-air. We introduce some algorithmic and technical advances in PATs that have allowed such high update rates (i.e. >10 kfps), having shifted from the high-speed single-point algorithm to the high-speed multipoint one.

1 Introduction

Phased Arrays of Transducers (PATs) provide accurate control of the phase and amplitude of dense arrays transducers, allowing dynamic control of ultrasound fields with various applications in creating parametric audio [1], providing haptics feedback [2–4] and levitating materials in mid-air [5].

One of the emergent uses for acoustic levitation is to create a mid-air display. For example, a sparse set of levitated particles can be used as voxels (volumetric pixels) to represent visual content in mid-air [6–10]. In addition, levitated particles can be used as anchors to suspend projection surfaces (i.e. light-weight fabric) that can act as mid-air screen [11]. More recently, the use of levitated particles moving at high speeds has allowed the creation of volumetric displays using the Persistence of Vision (PoV) effect [12–14]. Due to the capability of creating such mid-air visual content together with tactile and audio, systems using PATs are considered as new multimodal mixed-reality (MR) displays that user can see, touch, and hear (Fig. 1a) [12,14].

To exploit the PoV effect, particles need to scan the content within 0.1 s to reveal the 3D shape without flickering [15]. The use of a single particle maximizes the stiffness of the levitation trap, allowing higher accelerations. However, paths of content that single particles can travel within 0.1 s are limited to simple and small ones (Figs. 1b and 1c). In contrast, although the use of several high-speed particles reduces maximum accelerations that can be applied to each particle, this allows a higher level of versatility on how the power of the PAT is distributed, resulting in more flexible content delivery (Fig. 1d).

High update rate (i.e. typically more than 10 kfps) of PATs is essential to create such multimodal displays. Therefore, the computational performance needs to fulfill this requirement to realize interactive systems. In this talk, we introduce how the PAT algorithm has been advanced to achieve high-speed single-point sound-field control first [12] and then extended to multipoint [14].

2 High-speed implementation of algorithms

Our experimental setup is composed of two opposed arrays of 16 × 16 transducers, with a top-bottom arrangement separated by 24 cm (Fig. 1a). This setup is suitable for creating levitation traps with maximum vertical trapping stiffness [10]. Addition of levitation signatures to focal points is a simple but powerful way to create multipoint levitation [7] and thus was adopted in our method (i.e. adding a phase delay of π rad to the transducers in the top array).

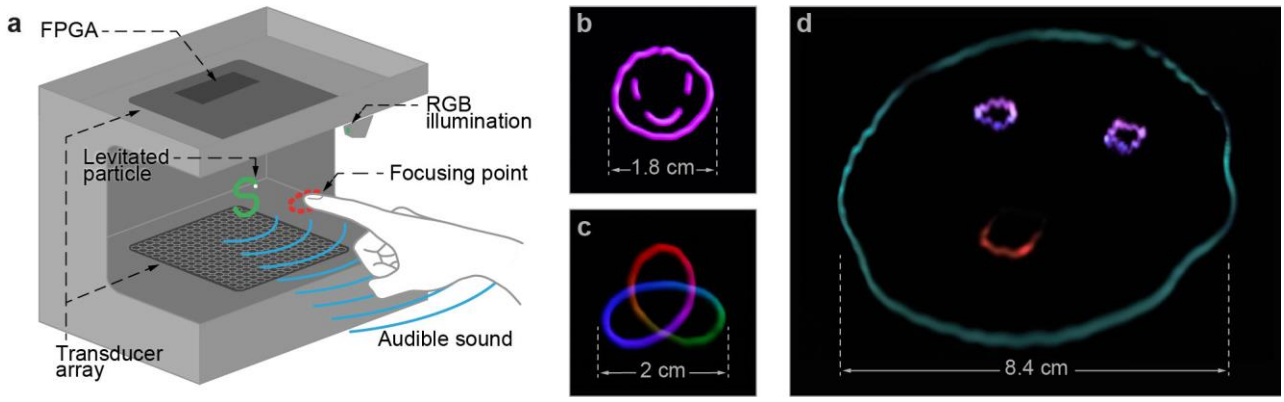

Fig. 1 Multimodal acoustic display [12,14]: (a) A geometrical description of our multimodal acoustic display's concept. Examples PoV images using a single (b, c) and multiple (6 points in total) points (d).

The square of the acoustic pressure (Pa) before applying the levitation signature to the focal points strongly correlates with trapping stiffness after applying the signature. Thus, the key to create better content across levitation and haptic domains is to accurately generate high-intensity focal points at high rates.

2.1 Sound fields generated by PATs

Before explaining methods of generating focal points, we first describe how the activations of the PATs influence the target sound fields. Let $\boldsymbol{\tau} = [\tau_1 \quad \tau_2 \quad \cdots \quad \tau_N]^T$ represent the activations of N transducers and $\boldsymbol{\zeta} = [\zeta_1 \quad \zeta_2 \quad \cdots \quad \zeta_L]^T$ represent the pressure at set of L points generated by the transducers. These sound pressures can be represented as complex numbers or equivalent phase and amplitude (i.e. $\tau_n = a_n e^{i\varphi_n}$ and $\zeta_l = a_l e^{i\varphi_l}$), given the single frequency nature of our problem (i.e. 40 kHz). Then, the relationship between $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$ can be described as simple linear equation systems [16], defined by the propagation matrix $\boldsymbol{F}$ using the scalar directivity function ($P_{n,l}$) of our sound sources approximated as a piston model, and the complex phase propagation ($\Phi_{n,l}$) approximated as a spherical sound source:

$$\boldsymbol{\zeta} = \begin{bmatrix} P_{1,1}\Phi_{1,1} & \cdots & P_{1,N}\Phi_{1,N} \\ \vdots & \ddots & \vdots \\ P_{L,1}\Phi_{L,1} & \cdots & P_{L,N}\Phi_{L,N} \end{bmatrix} \boldsymbol{\tau} = \boldsymbol{F\tau}. \qquad (1)$$

$$P_{l,n} = \frac{2J_1\big(kr\sin(\theta_{l,n})\big)}{kr\sin(\theta_{l,n})}\frac{P_{ref}}{d_{l,n}}. \qquad (2)$$

$$\Phi_{l,n} = e^{(kd_{l,n})i}. \qquad (3)$$

Here, P_{ref} represents the transducer's reference pressure at 1 m distance at the transducer's maximum amplitude; r represents the transducer's radius; k is the wavenumber; $d_{l,n}$ is the Euclidean distance between the transducer n and the point l; $\theta_{l,n}$ is the angle between the transducer´s normal and point l; and J_1 represents a Bessel function of the first kind.

2.2 High-speed single-point sound-field control

To create a single-point focal point at l, the conjugate of complex phase propagation $\Phi_{n,l}$ can be used as an activation of the transducers:

$$\tau_n = \Phi_{n,l}^* = e^{(-kd_{l,n})i}. \qquad (4)$$

Note here that amplitudes of the transducers can be adjusted to control the intensity of the focal point, allowing amplitude modulation to create haptics or audio.

In our work [12], the computation of the single-point focusing was embedded into an Field-Programmable Gate Array (FPGA) to realize a high update rate (i.e. 40 kHz). As shown in Eq. (4), a phase of each transducer depends only on the distance between the positions of focal point and the transducer. Therefore, this computation of phases of the transducers can be stored in a Look-Up Table (LUT), which enables to calculate required phases at 1 clock (20ns) per transducer. Since each PAT with an FPGA has 256 transducers, the transducers' activation can be computed at 5.12 μs (i.e. ~195 kHz). Inside the FPGA, phase and amplitude are discretized and have 64 and 33 levels of resolutions, respectively.

2.3 High-speed multipoint sound-field control

The back-propagation (BP) algorithm [10] approximates the transducers' activation $\boldsymbol{\tau}$ as a summation of the individual contributions of each point $\boldsymbol{\zeta}$. The algorithm simply back propagates the points in $\boldsymbol{\zeta}$ to the transducers in $\boldsymbol{\tau}$ using $\boldsymbol{F}^*$, which is a conjugate transpose of $\boldsymbol{F}$:

$$\boldsymbol{\tau} = \boldsymbol{F}^*\boldsymbol{\zeta}. \qquad (5)$$

Here, the transducer's activation is usually normalized (i.e. all transducers' amplitude set to maximum value 1), as to maximum amplitude at the target points. This naïve approach implicitly sets the phase of all target points to $\varphi_l = 0$ and leads to suboptimum solutions (e.g. target points can unnecessarily interfere destructively with each other).

The family of Gerchberg-Saxton (GS) algorithms [17], provide a heuristic approach to compute target phases optimizing field reconstructions (i.e. avoid destructive interference). By iterating the following four steps, the Iterative BP (IBP) algorithm, which was reformulated from the GS algorithm in [10], provides the optimum set of target phases maximizing the target amplitudes. Here, (x) represents that variables' values are in x-th iteration.

1. Back-propagation: $\tau^{(x)} = F^*\boldsymbol{\zeta}^{(x-1)}$.
2. Normalize the PATs' output: $\boldsymbol{\tau}^{(x)} = \left\{\frac{\tau_n^{(x)}}{\left|\tau_n^{(x)}\right|}, n = 1 \dots N\right\}$.
3. Forward-propagation: $\boldsymbol{\zeta}^{(x)} = \boldsymbol{F\tau}^{(x)}$.
4. Constraint target points: $\boldsymbol{\zeta}^{(x)} = \left\{\frac{\zeta_l^{(x)}}{\left|\zeta_l^{(x)}\right|}, l = 1 \dots L\right\}$.

Our GS-PAT algorithm presented in [14] removes unnecessary constraint of IBP, improving amplitude accuracy, and simplifies some of the steps, leading to large performance gains. GS-PAT uses a normalized matrix $\boldsymbol{B}$, which describes the back-propagation from each point in $\boldsymbol{\zeta}$ to each transducer in $\boldsymbol{\tau}$ for a focal point with exact amplitude 1 Pa. This allows us to reformulate the GS approach by computing a two-step propagation matrix $\boldsymbol{R} = \boldsymbol{FB}$, as follows:

$$\boldsymbol{R} = \boldsymbol{F}\begin{bmatrix} \frac{P_{1,1}}{\left|\sum_{n=1}^{N} P_{1,n}\right|}\Phi_{1,1}^* & \cdots & \frac{P_{\mathrm{L},1}}{\left|\sum_{n=1}^{N} P_{L,n}\right|}\Phi_{\mathrm{L},1}^* \\ \vdots & \ddots & \vdots \\ \frac{P_{1,\mathrm{N}}}{\left|\sum_{n=1}^{N} P_{1,n}\right|}\Phi_{1,N}^* & \cdots & \frac{P_{\mathrm{L,N}}}{\left|\sum_{n=1}^{N} P_{L,n}\right|}\Phi_{L,N}^* \end{bmatrix} = \boldsymbol{FB}. \quad (6)$$

We approximate the GS algorithm as a combination of two steps in each iteration: Eq. (6) uses $\boldsymbol{R}$ to combine the forward and backward propagation of our target points (GS steps 1 and 3), while Eq. (7) enforces the amplitude constraints of our target points (GS step 4).

1. Back- and forward-propagations: $\boldsymbol{\zeta}^{(x)} = \boldsymbol{R\zeta}^{(x-1)}$.
2. Constraint target points: $\boldsymbol{\zeta}^{(x)} = \left\{\frac{\zeta_l^{(x)}\left|\zeta_l^{(0)}\right|}{\left|\zeta_l^{(x)}\right|}, l = 1 \dots L\right\}$.

Our approximation offers several benefits. First, it avoids GS step 2, ensuring that transducers' amplitude remains as an additional degree of freedom, making our solver compatible with PATs operating over both phase and amplitude. Second and most important, avoiding GS step 2 allows matrix $\boldsymbol{R}$ to remain constant across iterations. Thus, we compute $\boldsymbol{R}$ only once at the beginning of the process, and each iteration only deals with one multiplication by matrix $\boldsymbol{R}$ and one normalization. This makes GS-PAT well suited for high-speed multi-point applications and better scalability in terms of the number of transducers.

The set $\boldsymbol{\zeta}^{(x)}$ provides an estimate of the final amplitudes of the generated field at each point. The final step in our method uses $\boldsymbol{\zeta}^{(x)}$ and $\boldsymbol{B}$ to correct the amplitudes of our control points and to compute the final transducer activation, as follows:

$$\boldsymbol{\zeta} = \left\{ \frac{\zeta_l^{(x)} \left|\zeta_l^{(0)}\right|^2}{\left|\zeta_l^{(x)}\right|^2}, l = 1 \dots L \right\}. \quad (2)$$

$$\boldsymbol{\tau} = \boldsymbol{B}\boldsymbol{\zeta}. \quad (3)$$

3 Results

3.1 Single-point multimodal acoustic display

As described in 2.1, the FPGA implementation of the single-point algorithm allows computation of the transducers' activation at 5.12 μs (i.e. faster than 40 kHz). This means its computing performance is only limited by the transducers' frequency (i.e. 40 kHz), allowing very high particle speeds of up to 8.75 m/s and 3.75 m/s in the vertical and horizontal directions. The high update rate also enables the system to create tactile and audio content together with visual content (like Figs. 1b and 1c) by time-multiplexing them, however with significant decrease in particle accelerations and some audible artefacts [12].

3.2 Multipoint multimodal acoustic display

We evaluated the performance of our GS-PAT implementation by testing the number of geometries per second (gps) that the algorithm could compute for a setup with $N = 512$ transducers, for a varying number of points and on different graphics cards. Particularly, we tested geometries with $L = \{2, 4, 8, 16, 32\}$ points on three different GPUs: a low-end, a mid- end, and a high-end GPUs, with results summarized in Table 1. Our algorithm provides a vast increase in performance (i.e. 100x) when compared to previous reference implementations of multipoint algorithms. For instance, Marzo et al. report computing rates of 90 gps ($N = 512$; L not specified), while Long et al. report rates of ~100 gps with full regularization and ~200 gps with no regularization ($L = 32$; $N = 256$).

Table 1. Computational performance of GS-PAT (geometries per second) with different GPUs.

L	GTX 1050	GTX 1660TI	RTX 2080
2	15,905	23,711	23,248
4	14,141	21,717	23,198
8	13,797	19,427	23,045
16	13,681	18,026	18,709
32	11,773	17,168	17,212

Fig. 2 summarizes the comparative performance of the different algorithms: Naïve, IBP [10], Long [4] and GS-PAT [14]. Although there is no significant difference in average amplitudes (Fig. 2a) between Naïve and the other phase retrieval algorithms, Naïve cannot deliver strong focal points especially when we analyze more challenging parameters when the number of points increase (Figs. 2b and 2c). This is because the Naïve cannot avoid destructive interference between the target points. Differences in performance between phase retrieval solvers are small, and the potential gains obtained from using more complex solvers should be weighed against the much higher computing rates enabled by GS-PAT, which are key to enable the range of novel PAT applications like multimodal acoustic displays.

Our results are summarized in Fig. 2d in terms of accuracy, with accuracy of each point computed as the ratio between the amplitude achieved for each point and its target amplitude (i.e. accuracy = 1, for a perfect reconstruction). As expected, phase-only solvers (Naïve and IBP) could not achieve accurate amplitude reconstructions, producing points that exceeded intended amplitudes and with very high standard

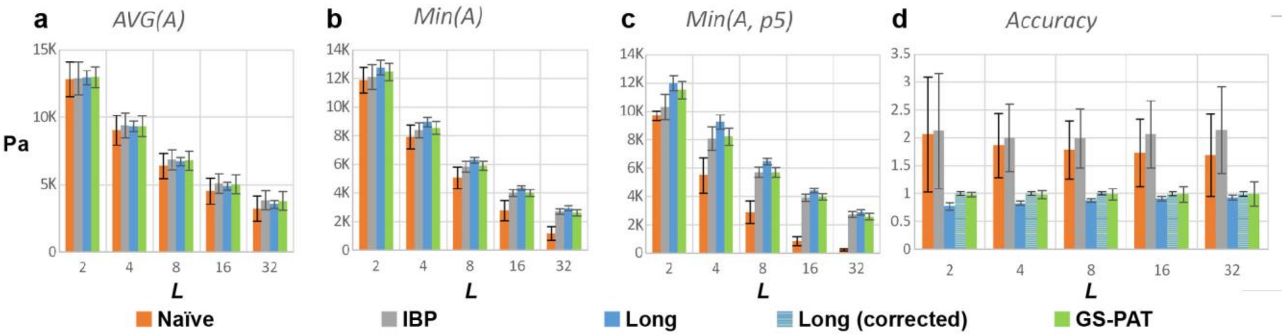

Fig. 2 Comparative performance of the multi-point solvers tested [14]: (a) Average focusing amplitude, (b) amplitudes of the weakest point, (c) Amplitudes of the weakest points, when focusing on geometries subject to destructive interface, and (d) reconstruction accuracy of each solver vs number of points reconstructed.

deviations. In contrast, GS-PAT produced an averaged accuracy ≃ 0.98 across all conditions, but the variability increased with the number of points, indicating that the algorithm can only provide accurate reconstructions for sound fields using a reduced number of points.

Such high-speed multipoint sound-field control allows new applications that single-point algorithm cannot realize, for examples, creating interactive three individual PoV content in mid-air (Fig. 3a) and delivering three modalities (visual, tactile and audio) at the same time without any audio artefacts (Fig. 3b). In addition, accurate control of amplitudes allows a higher level of versatility on how the power of the PAT is distributed, resulting in more flexible and large content delivery (Fig. 1d).

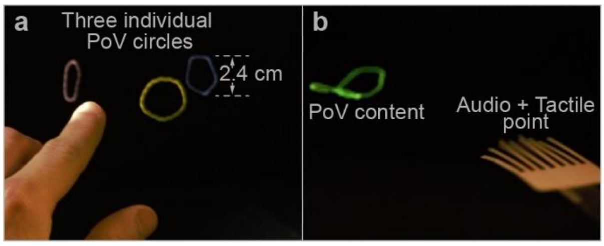

Fig. 3 Examples of applications [14]: (a) Three PoV circles with real-time control using hand gestures. (b) Creation of three modalities at the same time.

4 Conclusions

We demonstrated how the PAT algorithms have been advanced. The single-point algorithm is relatively simple and thus can be implemented using the FPGA, allowing the very high update rate (i.e. 40 kHz). Other advantages of using FPGAs for computation is it does not require any high-end processors like GPUs and the system can be standalone, making this technology more accessible. However, content that the single-point algorithm can deliver is limited to simple and small ones (Figs. 1b and 1c). In contrast, the multipoint algorithm allows more flexible and large content delivery (Figs. 1d, 3a and 3b). Next step would be to implement the multipoint algorithm using the FPGA and to realize a standalone multipoint acoustic display that can work at 40 kHz.

References

[1] W. Gan, J. Yang, and T. Kamakura, “A review of parametric acoustic array in air,” Appl. Acoust. **73**(12), 1211–1219 (2012).

[2] T. Hoshi, M. Takahashi, T. Iwamoto, and H. Shinoda, "Noncontact tactile display based on radiation pressure of airborne ultrasound," IEEE Trans. Haptics **3**(3), 155–165 (2010).

[3] T. Carter, S. A. Seah, B. Long, B. Drinkwater, and S. Subramanian, “UltraHaptics: Multi-point mid-air haptic feedback for touch surfaces,” Proc. 26th Annu. ACM Symp. User Interface Softw. Technol. (UIST’13) 505–514 (2013).

[4] B. Long, S. A. Seah, T. Carter, and S. Subramanian, “Rendering volumetric haptic shapes in mid-air using ultrasound,” ACM Trans. Graph. **33**(6), 1–10 (2014).

[5] M. Andrade, A. Marzo, and J. C. Adamowski, “Acoustic levitation in mid-air: Recent advances, challenges, and future perspectives,” Appl. Phys. Lett. **116**(25), 250501 (2020).

[6] Y. Ochiai, T. Hoshi, and J. Rekimoto, “Pixie Dust: Graphics Generated by Levitated and Animated Objects in Computational Acoustic-Potential Field,” ACM Trans. Graph. **33**(4), 1–13 (2014).

[7] A. Marzo, S. A. Seah, B. W. Drinkwater, D. R. Sahoo, B. Long, and S. Subramanian, “Holographic acoustic elements for manipulation of levitated objects,” Nat. Commun. **6**(1), 8661 (2015).

[8] T. Omirou, A. Marzo, S. A. Seah, and S. Subramanian, "LeviPath: Modular Acoustic Levitation for 3D Path Visualisations," Proc. ACM Conf. Hum. Factors Comput. Syst. (CHI’15) 1, 309–312 (2015).

[9] D. R. Sahoo, N. Takuto, A. Marzo, T. Omirou, M. Asakawa, and S. Subramanian, “JOLED: A Mid-Air Display Based on Electrostatic Rotation of Levitated Janus Objects,” Proc. 29th ACM User Interface Softw. Technol. Symp. (UIST ’16) 437–448 (2016).

[10] A. Marzo and B. W. Drinkwater, “Holographic acoustic tweezers,” Proc. Natl. Acad. Sci. 201813047 (2018).

[11] R. Morales, A. Marzo, S. Subramanian, and D. Martínez, “LeviProps: Animating Levitated Optimized Fabric Structures using Holographic Acoustic Tweezers,” Proc. 32nd ACM User Interface Softw. Technol. Symp. (UIST ’19) 651–661 (2019).

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

[13] T. Fushimi, A. Marzo, B. W. Drinkwater, and T. L. Hill, “Acoustophoretic volumetric displays using a fast-moving levitated particle,” Appl. Phys. Lett. **115**(6), 064101 (2019).

[14] D. M. Plasencia, R. Hirayama, R. Montano and S. Subramanian, “GS-PAT: High-Speed Multi-point Sound-Fields for Phased Arrays of Transducers,” ACM Transactions on Graphics **39**, 138, (2020).

[15] R. W. Bowen, J. Pola, and L. Matin, “Visual persistence: Effects of flash luminance, duration and energy,” Vis. Res. **14**(4), 295–303 (1974).

[16] M. Andrade, T. Camargo, and A. Marzo, “Automatic contactless injection, transportation, merging, and ejection of droplets with a multifocal point acoustic levitator,” Rev. Sci. Instrum. **89**(12) 125105 (2018).

[17] R. W. Gerchberg, “A practical algorithm for the determination of phase from image and diffraction plane pictures,” Optik **35**, 237–246 (1972).