GS-PAT: High-Speed Multi-Point Sound-Fields for Phased Arrays of Transducers

DIEGO MARTINEZ PLASENCIA, University of Sussex, UK
RYUJI HIRAYAMA, University of Sussex, UK
ROBERTO MONTANO-MURILLO, University of Sussex, UK
SRIRAM SUBRAMANIAN, University of Sussex, UK

Phased Arrays of Transducers (PATs) allow accurate control of ultrasound fields, with applications in haptics, levitation (i.e. displays) and parametric audio. However, algorithms for multi-point levitation or tactile feedback are usually limited to computing solutions in the order of hundreds of sound-fields per second, preventing the use of multiple high-speed points, a feature that can broaden the scope of applications of PATs. We present *GS-PAT*, a GPU multi-point phase retrieval algorithm, capable of computing 17K solutions per second for up to 32 simultaneous points in a mid-end consumer grade GPU (NVidia GTX 1660). We describe the algorithm and compare it to state of the art multi-point algorithms used for ultrasound haptics and levitation, showing similar quality of the generated sound-fields, and much higher computation rates. We then illustrate how the shift in paradigm enabled by *GS-PAT* (i.e. real-time control of several high-speed points) opens new applications for PAT technologies, such as in volumetric fully coloured displays, multi-point spatio-temporal tactile feedback, parametric audio and simultaneous combinations of these modalities.

CCS Concepts: • **Computing methodologies** → **Physical simulation**.

Additional Key Words and Phrases: Multi-point phase optimization; Parametric sound; Ultrasound Levitation; Mid-air haptics; Phased Arrays of Transducers

1 INTRODUCTION

Phased Arrays of Transducers (PATs) provide accurate control of the phase and amplitude of dense arrays of transducers (e.g. 16x16 transducers), and have demonstrated capabilities to present visual, auditive and tactile content.

Visual content using PATs typically relied on sparse sets of levitated particles, initially constrained to specific locations or arrangements [Ochiai et al. 2014; Omirou et al. 2015]. Algorithmic advances allowed unconstrained 3D positioning of single particles first [Marzo

Authors' addresses: Diego Martinez Plasencia, University of Sussex, UK, dm372@sussex.ac.uk; Ryuji Hirayama, University of Sussex, UK, r.hirayama@sussex.ac.uk; Roberto Montano-Murillo, University of Sussex, UK, r.montano-murillo@sussex.ac.uk; Sriram Subramanian, University of Sussex, UK, sriram@sussex.ac.uk.
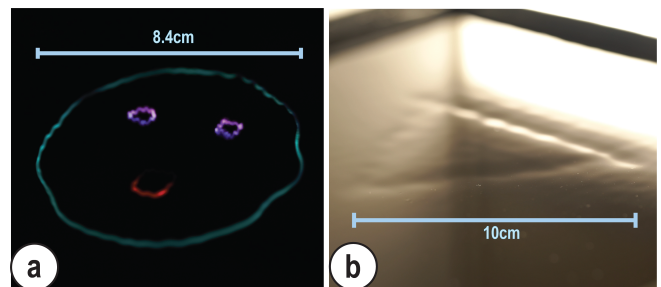
Fig. 1. Example applications enabled by GS-PAT. (A) Volumetric POV visual content using 6 high-speed scanning particles. (B) Multi-point tactile feedback using three high-speed moving tactile points.

et al. 2015] and then extended this to several particles [Marzo and Drinkwater 2019]. Recently, the use of a single levitated high-speed particle [Hirayama et al. 2019] has allowed the creation of 3D displays using the Persistence of Vision (POV) effect. For audio, PATs have allowed the steering of single [Ochiai et al. 2017; Olszewski et al. 2005] and then multiple [Shi et al. 2015] audible beams in different directions. Haptics also demonstrated mid-air feedback, first at a single tactile point in 3D [Hoshi et al. 2010], then at several tactile points [Carter et al. 2013; Inoue et al. 2015; Long et al. 2014] and finally using a single high-speed tactile point [Frier et al. 2019].

From an algorithmic point of view, all these PAT approaches have followed a similar evolution, with single-point approaches demonstrating an initial potential [Carter et al. 2013; Marzo et al. 2015; Olszewski et al. 2005]; multi-point extending it [Long et al. 2014; Marzo and Drinkwater 2019; Shi et al. 2015]; and single-point high-speed algorithms demonstrating further applications, such as better tactile feedback [Frier et al. 2019], or even simultaneous delivery of visual, tactile and auditive content [Hirayama et al. 2019].

The next natural step in this evolution lies in exploiting multi-point sound-fields computed at high rates (e.g. such as >10KHz, as used for single-particle POV displays [Hirayama et al. 2019]; or 17KHz, used for single-point haptics [Frier et al. 2019]). However, this step remains hindered by the performance of existing multi-point algorithms, typically limited to computing rates of hundreds of hertz. This paper fills this gap, and demonstrates how this shift in paradigm from a single high-speed point to multiple high-speed points enables new applications of PATs for all 3 modalities.

We present *GS-PAT*, a high performance approximation to the Gerchberg-Saxton (GS) algorithm [Gerchberg 1972] specifically tailored to phase and amplitude PATs. *GS-PAT* reduces the algorithmic complexity of GS, as to deal with more points or PAT setups with
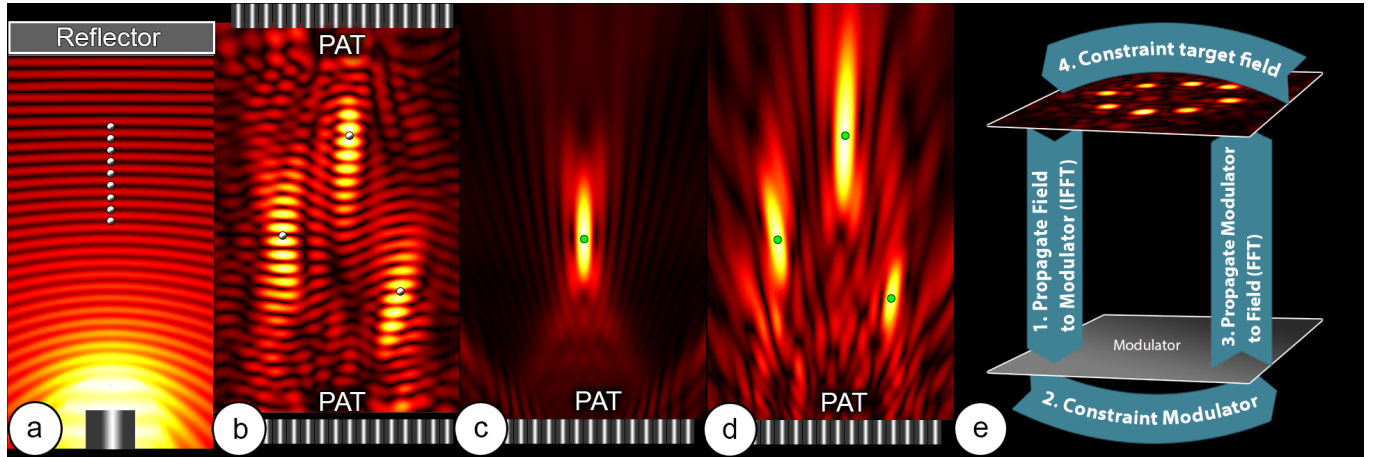
Fig. 2. Summary of approaches for sound-field computation. (a) Standing waves allow simple levitation, but no control of particle position. (b) Addition of levitation signatures allows multi-point levitation. (c) Single-point tactile feedback is computed as a focusing pattern. (d) Complex summation of focusing patterns allows for multi-point tactile feedback. (e) Summary of stages in a classic Gerchberg-Saxton (GS) iterative algorithm.

more transducers. For a PAT setup with 512 transducers and 32 (tactile or levitation) points, our algorithm can compute 17K sound-fields per second in a consumer-grade GPU (GeForce GTX 1660).

We compare our solution to state of the art algorithms used in multi-point levitation [Marzo and Drinkwater 2019] and haptics [Long et al. 2014] showing that we can achieve the high computing rates required for our applications, with minimum sacrifices in the quality of the sound-fields generated and retaining the capabilities of multi-point algorithms to avoid destructive interference.

We then illustrate how the shift in paradigm enabled by *GS-PAT* (i.e. from single to multiple high-speed points) provides great versatility in sound-field control and enables novel applications for PATs, such as in creating volumetric POV content using several particles (see Figure 1a), novel types of tactile feedback (see Figure 1b), audio and combinations of these modalities.

2 RELATED WORK AND BACKGROUND

2.1 Sound-field computation & multimodal UIs
Sound-field computation in the scope of interactive applications usually refers to audible sound, either to model the sound generated by 3D objects [Wang and James 2019; Wang et al. 2018; Zheng and James 2011] or to capture the propagation of sound in computer generated 3D scenes [Raghuvanshi and Snyder 2014, 2018].

We focus instead on interactive applications based on interference of single frequency ultrasound fields. Such fields were first observed to trap particles in the lobes of a standing wave more than 150 years ago [Stevens 1899], as shown in Figure 2a. Standing waves have been used to create displays using particles as tangible pieces of information [Ochiai et al. 2014; Omirou et al. 2015, 2016; Sahoo et al. 2016], but with very limited control of individual particle positions.

Multi-point algorithms [Marzo and Drinkwater 2019] allowed displays with free control of the position of several levitated particles (Figure 2b), or used particles as anchors to suspend projection surfaces which can act as mid-air displays [Morales et al. 2019].

Interference patterns creating focusing points (Figure 2c), Bessel beams [Norasikin et al. 2019], diffraction gratings [Norasikin et al. 2018] or boundary holograms [Inoue et al. 2019] have also been used to create tactile sensations [Hoshi et al. 2010], directional sound [Olszewski et al. 2005; Yoneyama et al. 1983] or even to create olfactory and taste experiences [Hasegawa et al. 2017; Vi et al. 2017].

Even if PATs have shown the ability to stimulate every modality, interactive experiences usually focus on only one of them, using other technologies to stimulate other modalities. For instance, ultrasonic tactile feedback has been combined with multiple imaging solutions, like tabletops [Carter et al. 2013], VR headsets [Georgiou et al. 2018; Sand et al. 2015a], mid-air screens [Monnai et al. 2014; Sand et al. 2015b], real objects [Freeman et al. 2019; Norasikin et al. 2018] or replicas of real object's light-fields [Yoshida et al. 2017].

However, PAT experiences simultaneously delivering several modalities have started to be demonstrated only very recently [Hirayama et al. 2019; Shakeri et al. 2019]. Our algorithm (i.e. GS-PAT) and approach (i.e. use of high-speed multi-point sound-fields) open up new possibilities for visual, audio and tactile modalities independently, but also tighter integration across these three modalities and support for setups with more transducers, all of them key to enable exploration of new multimodal applications using PATs.

2.2 Multi-point Phase Optimization Algorithms
2.2.1 Field generated by one transducer and single-point fields: The field generated at a point z by a transducer t activated with a specific amplitude a_t and phase φ_t can be determined as:

$$\Psi_{t,z} = a_t \cdot e^{(\varphi_t)i} \cdot P_{t,z} \cdot \Phi_{t,z} \qquad (1)$$

Where $P_{t,z}$ refers the scalar directivity function of our transducers, approximated as a piston model; and $\Phi_{t,z}$ models the complex phase propagation, approximated as a point-source [O'Neil 1949].

$$P_{t,z} = \frac{2 \cdot J_1(k \cdot r \cdot \sin(\theta))}{k \cdot r \cdot \sin(\theta)} \cdot \frac{P_{ref}}{\mathbf{d}(\mathbf{p}_t, \mathbf{p}_z)}; \Phi_{t,z} = e^{(k \cdot \mathbf{d}(\mathbf{p}_t, \mathbf{p}_z))i} \qquad (2)$$

Here, P_{ref} represents the transducer's reference pressure at 1m distance at the transducer's maximum amplitude (i.e. $a_t = 1$); r represents the transducer's radius; k is the wavenumber ($k = \omega/c_0$); $\mathbf{d}(\mathbf{p}_t, \mathbf{p}_z)$ is the Euclidean distance between the transducer at position $\mathbf{p}_t$ and the point at $\mathbf{p}_z$; θ is the angle between the transducer's normal and point z; and J_1 is a Bessel function of the first kind.

Equations (1) and (2) can be used to compute the required transducer activation (i.e. $a_t \cdot e^{(\varphi_t)i} = P_{t,z} \cdot \Phi_{t,z}^* \cdot a_z$) to create a single-point focusing pattern at point z (see Figure 2c), as used for tactile stimulation [Hoshi et al. 2010]. The addition of levitation signatures allows the creation of single levitation traps from these focusing patterns [Marzo et al. 2015] and such single-point algorithms have already been successfully adapted to high computing rates [Frier et al. 2019; Hirayama et al. 2019].

2.2.2 Multi-point sound-fields and phase optimization.

Computing sound-fields creating several focusing points (i.e. for haptics, see Figure 2d) or traps (i.e. for levitation, as in Figure 2b) requires more complex algorithms, harder to adapt to high computing rates or PAT setups with a higher number of transducers.

Existing multi-point algorithms are summarized here, together with their algorithmic complexity (using Big-O notation), and our *GS-PAT* algorithm is then described in the following section.

Let $\mathbf{Z}$ represent the set of desired control points (focus or traps), and $\mathbf{T}$ represent the set of transducers in our PAT arrangement (with $|\mathbf{Z}| << |\mathbf{T}|$). For simplicity, each element x in $\mathbf{Z}$ or $\mathbf{T}$ (control point or transducer) is represented as a tuple $\{a_x, \varphi_x, \mathbf{p}_x\}$ while their equivalent complex representations are denoted by sets ζ and $\boldsymbol{\tau}$ (i.e. $\boldsymbol{\tau} = \{a_t \cdot e^{(\varphi_t)i}, \forall t \in \mathbf{T}\}$). While the position ($\mathbf{p}_z$) and amplitude ($a_z$) of the points in $\mathbf{Z}$, and the position of the transducers ($\mathbf{p}_t$) are known (i.e. input variables), the transducers' activation (a_t, φ_t) and the target point phases (φ_t) are unconstrained degrees of freedom (i.e. output variables), which are tuned to reconstruct the intended field at the target points. Also, some PAT devices might only allow phase control of the transducers (i.e. $a_t = 1, \forall t \in \mathbf{T}$).

A first order approximation of the sound-field generated by a PAT can be described as a simple linear equation system [Andrade et al. 2018], defined by the forward propagation $\mathbf{F}$ of any given transducer activation $\boldsymbol{\tau}$ to the points in $\mathbf{Z}$:

$$
\zeta = \begin{bmatrix} P_{1,1} \cdot \Phi_{1,1} & \cdots & P_{\mathbf{T},1} \cdot \Phi_{\mathbf{T},1} \\ \vdots & \ddots & \vdots \\ P_{1,\mathbf{Z}} \cdot \Phi_{1,\mathbf{Z}} & \cdots & P_{\mathbf{T},\mathbf{Z}} \cdot \Phi_{\mathbf{T},\mathbf{Z}} \end{bmatrix} \cdot \boldsymbol{\tau} = \mathbf{F} \cdot \boldsymbol{\tau} \tag{3}
$$

Solving the inverse problem (i.e. the activation $\boldsymbol{\tau}$ required to produce a given field ζ) is possible, but it leads to an underdetermined system (i.e. $|\mathbf{Z}| << |\mathbf{T}|$) and requires point phases (φ_z) to be known in advance.

Long et al. [Long et al. 2014] approach this using an Eigen solver to retrieve phases φ_z (i.e. $O(|\mathbf{Z}|^3)$) and Tikhonov regularization , to overdetermine the system and solve for a_t and φ_t (i.e. $O(|T|^{2.53})$, using fast multiplication algorithms [Le Gall 2012]). This results in a relatively more complex global optimization, with complexity heavily determined by $|\mathbf{T}|$, as it solves for each transducer in τ to approximate the target field while retaining low variability on transducers' intensity.

Other algorithms rely on heuristics to produce faster solvers (i.e. $< O(|\mathbf{T}|^{2.53})$). Instead of optimizing for the transducer's activation (i.e. a_t, φ_t), these algorithms approximate it as a summation of the individual contributions of each point in the target field $\mathbf{Z}$, reducing the problem to a single degree of freedom per point (i.e. φ_z).

For instance, the back-propagation (*BP*) algorithm [Marzo and Drinkwater 2019] in Equation 4 simply propagates each point in $\mathbf{Z}$ to the elements in $\mathbf{T}$. This implicitly sets the phase of all target points to $\varphi_z = 0$ and leads to suboptimum solutions (e.g. target points can unnecessarily interfere destructively with each other).

$$
a_t \cdot e^{(\varphi_t)i} = \frac{\sum_z^Z P_{t,z} \cdot \Phi_{t,z}^* \cdot a_z}{|\sum_z^Z P_{t,z} \cdot \Phi_{t,z}^* \cdot a_z|} \quad , \forall t \in \mathbf{T} \tag{4}
$$

The family of Gerchberg-Saxton (GS) algorithms [Gerchberg 1972] provide a heuristic approach to compute target phases (φ_z) that optimize field reconstructions (i.e. avoid destructive interference) and they have been used in a wide range of imaging applications [Dainty and Fienup 1987; Fogel et al. 2016].

However, traditional GS methods rely on assumptions that make them unsuitable for PAT applications. First, most imaging applications rely on square-shaped modulators and dense target fields defined in a 2D plane parallel to the modulator and uniformly sampled, which allows them to rely on Fourier transforms for efficient computation. Second, GS was intended for optical modulators which control phase or amplitude (but not both). In contrast, our target fields are a sparse collection of points, unevenly located anywhere in 3D; PAT modulators can vary in shape (e.g. our non-square top-bottom setup, in Figure 5) and can control both phase and amplitude.

The GS algorithm was reformulated for PATs as the *IBP* algorithm [Marzo and Drinkwater 2019], composed of the four steps in Table 1. Please note super-indices (e.g. $\zeta^{(n)}, \boldsymbol{\tau}^{(n)}$) denote variables' values across each of the N iterations. Also, the order in the steps has been modified to highlight its parallelism to a classic GS (i.e. step 1 in [Marzo and Drinkwater 2019] corresponds to step 3 in Table 1), with no effects on algorithm's behaviour or convergence.

While solving many limitations within the original algorithm (i.e. non-flat PAT arrangements, arbitrary 3D positions), *IBP* is still limited to phase-only PAT modulators. *GS-PAT* removes this unnecessary constraint, improving accuracy, and it also simplifies some of the steps, leading to large performance gains.

Table 1. Summary of GS iterative algorithm adapted to multi-point sound-field computation ($O(N \cdot |\mathbf{T}| \cdot |\mathbf{Z}|)$)

1. Back-propagate field Z to modulator T:	$O(|\mathbf{Z}| \cdot |\mathbf{T}|)$
$\xi^{(n)} = \left\{ \sum_z^Z P_{t,z} \cdot \Phi_{t,z}^* \cdot a_z^{(n-1)} \cdot e^{i\varphi_z^{(n-1)}}, \forall t \in \mathbf{T} \right\}$	
2. Constrain modulator's output:	$O(|\mathbf{T}|)$
$\boldsymbol{\tau}^{(n)} = \left\{ \frac{t}{|t|} \cdot a_t^{(0)}, \quad \forall t \in \xi^{(n)} \right\}$	
3. Propagate modulator T to field Z:	$O(|\mathbf{Z}| \cdot |\mathbf{T}|)$
$\Gamma^{(n)} = \left\{ \sum_t^T P_{t,z} \Phi_{t,z} \cdot a_t^{(n)} \cdot e^{i\varphi_t(n)}, \forall z \in \mathbf{Z} \right\}$	
4. Constrain target field:	$O(|\mathbf{Z}|)$
$\zeta^{(n)} = \left\{ \frac{z}{|z|} \cdot a_z^{(0)}, \quad \forall z \in \Gamma^{(n)} \right\}$	

3 GS-PAT: MULTI-POINT SOUND-FIELDS SIMPLIFIED

Traditional approaches use a matrix $\mathbf{F}$ (see Equation 3), containing the usual forward propagation of the transducers' activation to each of our target points in $\mathbf{Z}$, and its conjugate $\mathbf{F}^*$ to back-propagate from points to transducers. This is inaccurate in terms of amplitude, but allowed as transducers' amplitude is discarded in GS step 2.

We instead use a normalised matrix $\mathbf{B}$, which describes the (back) propagation of each point z to each transducer in $\mathbf{T}$ for a focussing point with exact amplitude 1Pa. Matrix $\mathbf{B}$ is key to allow us to define the amplitude of target points in actual Pascals, and also to provide accurate reconstruction prediction in the last step of our method. This allows us to reformulate the GS approach by computing a two-step propagation matrix $\mathbf{R} = \mathbf{F} \cdot \mathbf{B} \in \mathbb{C}^{|\mathbf{Z}| \times |\mathbf{Z}|}$, as follows:

$$\mathbf{R} = \mathbf{F} \cdot \begin{bmatrix} \frac{P_{1,1}}{\left|\Sigma_t^{\mathrm{T}} P_{t,1}\right|^2} \cdot \Phi_{1,1}^* & \cdots & \frac{P_{1,z}}{\left|\Sigma_t^{\mathrm{T}} P_{t,z}\right|^2} \cdot \Phi_{1,\mathbf{Z}}^* \\ \vdots & \ddots & \vdots \\ \frac{P_{\mathrm{T},1}}{\left|\Sigma_t^{\mathrm{T}} P_{t,1}\right|^2} \cdot \Phi_{\mathrm{T},1}^* & \cdots & \frac{P_{\mathrm{T},\mathrm{Z}}}{\left|\Sigma_t^{\mathrm{T}} P_{t,z}\right|^2} \cdot \Phi_{\mathrm{T},\mathbf{Z}}^* \end{bmatrix} = \mathbf{F} \cdot \mathbf{B} \quad (5)$$

The resulting matrix $\mathbf{R}$ describes the back-propagation from each point z to each transducer t in the PAT setup (GS step 1) and their forward propagation to each point z (step 3). We approximate the GS algorithm as a combination of two steps in each iteration: Equation 6 uses $\mathbf{R}$ to combine the forward and backward propagation of our target points (GS steps 1 and 3), while Equation 7 enforces the amplitude constraints of our target points (GS step 4).

$$\Gamma^{(n)} = \mathbf{R} \cdot \zeta^{(n-1)} \quad (6)$$

$$\zeta^{(n)} = \left\{ \frac{z^{(n)}}{\left|z^{(n)}\right|} \cdot a_z^{(0)}, \quad \forall z^{(n)} \in \Gamma^{(n)} \right\} \quad (7)$$

Our approximation offers several benefits. First, it avoids GS step 2 (i.e. Constrain modulator's output). This ensures 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 $\mathbf{R}$ to remain constant across iterations. Normalizing in GS step 2 would require recomputation of, at least, GS steps 1 & 2 in each

iteration, leading to an overall complexity of $O(N \cdot |\mathbf{T}| \cdot |\mathbf{Z}|)$, dependent on $|\mathbf{T}|$, and limiting scalability to larger PAT setups. Instead, we compute $\mathbf{R}$ only once at the beginning of the process (i.e. $O(|\mathbf{T}| \cdot |\mathbf{Z}|)$), and each iteration only deals with one multiplication by matrix $\mathbf{R}$ (i.e. $O(|\mathbf{Z}|^2)$ in Equation 6) and one normalization (i.e. $O(|\mathbf{Z}|)$ in Equation 7). This leads to an overall complexity of $O\left(N \cdot |\mathbf{Z}|^2 + |\mathbf{T}| \cdot |\mathbf{Z}|\right)$, making *GS-PAT* well suited for high-speed multi-point applications and better scalability in terms of $|\mathbf{T}|$.

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

$$\zeta^\Omega = \left\{ \frac{z^{(n)}}{\left|z^{(n)}\right|^2} \cdot \left(a_z^{(0)}\right)^2, \quad \forall z^{(n)} \in \Gamma^{(n)} \right\} \quad (8)$$
$$\boldsymbol{\tau}^\Omega = \mathbf{B} \cdot \zeta^\Omega$$

The activation τ^Ω can be directly used to create tactile points (see Figure 5f&g), but also to create levitation traps (see Figure 5d&e) by adding a levitation signature (i.e. we add π radians to the transducers in the top array, to maximise vertical trapping [Marzo and Drinkwater 2019]). The activation must also be capped to the transducers' operational range ($a_t \in [0, 1]$) . This must be considered when defining target amplitudes (i.e. a_z). Too high amplitudes (i.e. beyond the capabilities of the PAT device) will lead to transducers' amplitudes being capped and performance will degrade to that of a phase only solver. Finally, capped transducers' amplitude must be discretized and mapped to duty cycle using to the nonlinear relationship in Equation S4 in [Hirayama et al. 2019].

3.1 High-performance GPU implementation

We implemented *GS-PAT* using OpenCL and clBlas [Knox 2013], as summarized in Figure 3. The solver can compute several solutions (i.e. arrangements of $|\mathbf{Z}|$ target points or geometries) in parallel, with $G \in \mathbb{N}$ representing the number of geometries/solutions computed in parallel. Grey boxes represent data, purple boxes represent the OpenCL kernels implemented and red ones represent calls to clBlas complex matrix multiplication cgemm.
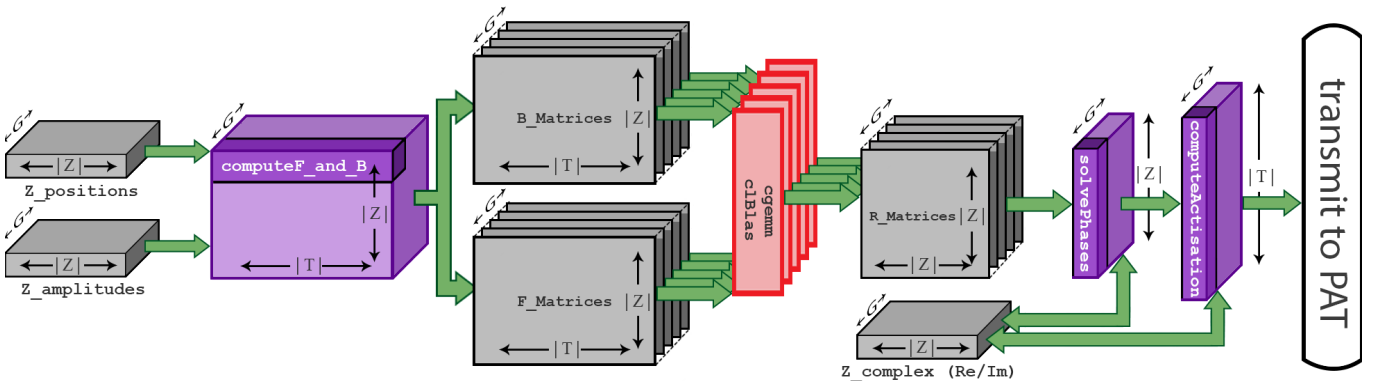

Fig. 3. Summary of our openCL based implementation, detailing data buffers (grey boxes), computing kernels (purple boxes for ours; red for clBlas) and shaded areas indicating local workgroups used. Size of buffers, kernels and workgroups denoted in terms of the number of target points ($|z|$), transducers ($|T|$) and geometries computed in parallel (G).

The dimensions of the boxes represent the amount of data to be processed, in terms of numbers of transducers ($|\mathbf{T}|$), number of target points ($|\mathbf{Z}|$) and number of geometries (G). These dimensions match the size of the global workgroup sizes used by each kernel, while the smaller shaded areas represent local workgroup sizes (in terms of $|\mathbf{T}|$, $|\mathbf{Z}|$ and G). Synchronization events and some auxiliary data structures (i.e. location of each transducer, pre-computed buffers to compute $P_{t,z}$ and $\Phi_{t,z}$) are used, but they are omitted in Figure 3 for brevity.

The buffers Z_positions and Z_amplitudes provide the input to our solver, describing the desired location and amplitudes of the target points for each geometry (i.e. our target set $\mathbf{Z}$), while Z_complex contains the complex representation of target points across our iterations (i.e. $\zeta^{(n)}$).

Kernel computeF_and_B calculates matrices $\mathbf{B}$ and $\mathbf{F}$ according to Equation 5, while $\mathbf{G}$ parallel calls to clgemm are used to build matrix $\mathbf{R}$ for each geometry. Kernel solvePhases encapsulates the iterative steps in GS-PAT described by Equations 6 and 7, and the number of iterations is provided as a parameter (i.e. we used N=100 during our tests). Finally, kernel computeActivation is used to compute the phases and amplitudes to be delivered to our PAT according to Equation 8, as well as the addition of levitation signatures (if required) and discretization steps required to drive our PAT.

4 COMPARATIVE EVALUATION

We evaluated the performance of *GS-PAT*, comparing it to the two most relevant phase retrieval solvers for PATs: the heuristic solver *IBP* [Marzo and Drinkwater 2019]; and the global optimization solver *Long* [Long et al. 2014]. We also use a *Naïve* solver (see *BP*, as in Equation 4), as a baseline to assess the benefits of using phase retrieval solvers and to understand when these benefits come into effect. Our evaluation contains two sections, one testing the ability of the solvers to achieve high intensity reconstructions (i.e. avoid destructive interference) and a second one testing their ability to accurately reproduce specific intensities.

The choice of a suitable metric to assess reconstruction intensity also deserves particular consideration. We test techniques from two different domains: haptics and levitation. The first ones are intended to create high pressure focal points, measured in Pascals. The second ones create levitation traps (i.e. area of near-zero pressure surrounded by high pressures), with their quality determined by the trap stiffness (i.e. Laplacian of Gor'kov potential). As shown in Supplementary Material, the square of the acoustic pressure (Pa) before applying the levitation signature to the focus points strongly correlates ($R^2 > 0.99$) with trapping stiffness (N/m) after applying the signature. Thus, we use the acoustic pressure (i.e. amplitude of the complex pressure) as a single metric allowing for fair comparisons across levitation and haptic domains.

Also, all solvers were run according to our experimental setup (i.e. 512 transducers in a top-bottom arrangement, see Section 5.1). For fairness in comparison, both *IBP* and *GS-PAT* used N = 100 iterations, as per the original implementation by [Marzo and Drinkwater 2019]. *Long* was used with full regularization ($\gamma = 1$). The C++ implementations of the solvers tested, as well as the code used for our evaluations can be found in Supplementary Material.

4.1 Avoiding destructive interference

We compared the performance of the different solvers, when reconstructing arrangements (geometries) of $|\mathbf{Z}| \in \{2, 4, 8, 16, 32\}$ points. Particularly, we generated 5000 random geometries, 1000 for each condition in $|\mathbf{Z}|$ (i.e. 2,4,8,16 and 32 points). Each geometry was randomly generated, but the position of its points was constrained to the $12 \times 12 \times 12$cm at the centre of our experimental setup (working volume of our device) and the same set of 5000 geometries was used to test all solvers. Also, given that *IBP* is intended for phase-only modulators, we constrained the amplitude of all transducers to $a_t = 1$, so that all the solvers were delivering the same raw amount of power in all conditions.

Our evaluation measured three different parameters across the geometries tested, moving from more generic to more challenging aspects: a) average pressure across all the points in the geometry ($AVG(A)$); b) pressure of the weakest point in the geometry ($Min(A)$); and c) pressure of the weakest point for geometries showing destructive interference ($Min(A, p5)$).

Metric $Min(A, p5)$ requires us to identify geometries/cases where destructive interference occurred, which is related to the transducer arrangement used (same across all our conditions and solvers), but also to the location of the points in the geometry reconstructed [Morales et al. 2019]. To identify geometries potentially subject to destructive interference we looked at the results from our *Naïve* solver, as this solver does not optimise phases as to attenuate these effects. We selected the 5% of the geometries providing lower amplitude for their weakest point in each condition ($|\mathbf{Z}| \in \{2, 4, 8, 16, 32\}$), and attributed such low amplitudes to the presence of destructive interference. We then compared the performance of each of the four solvers for these selected cases/geometries to compute $Min(A, p5)$.

Our results are summarised in Figure 4. As expected, the amplitude of the points reconstructed tends to decrease as the number of points in the geometry increases. In general, the results also show that all 3 phase retrieval solvers (*IBP*, *Long* and *GS-PAT*) perform better than *Naïve*, but also show that the differences among the phase retrieval solvers are small, with our solver usually performing in a second position (i.e. slightly better than existing heuristic methods, but slightly worse than global ones). These results are further discussed in the following subsections.

4.1.1 Phase Retrieval Methods (IBP, Long, GS-PAT) vs Naïve. In terms of $AVG(A)$ (see Figure 4a) phase retrieval methods performed better than *Naïve*, with higher $AVG(A)$ and smaller standard deviations for all conditions, but such differences were small (i.e. average differences of ~1.2% for $|\mathbf{Z}| = 2$, going up to ~16% for $|\mathbf{Z}| = 32$ points).

These results are expected as, given the high number of transducers compared to the number of points (i.e. 512 vs 32), the *Naïve* algorithm will still deliver good average pressure in most cases, even if this comes at the expense of some points in these cases being too weak to be felt or to levitate/move a particle at high speeds.

The differences between *Naïve* and the 3 phase retrieval algorithms become more apparent when we analyse more challenging parameters (Figures 4b and c). In terms of amplitude of the weakest point ($Min(A)$), phase retrieval algorithms provide ~5% higher $Min(A)$ when compared to *Naïve* for $|\mathbf{Z}| = 2$, going up to ~ 133%
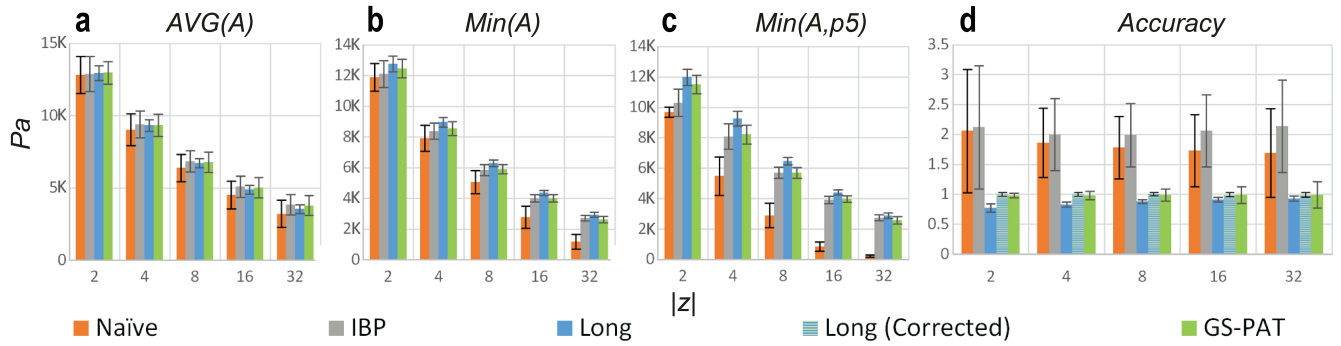
Fig. 4. Comparative performance of the multi-point solvers tested. (a) Average focusing amplitude achieved by each solver vs number of points reconstructed (1000 random geometries per case). (b) Amplitudes of the weakest point per solver vs number of points. (c) Amplitudes of the weakest points, when focusing on geometries subject to destructive interference. (d) Reconstruction accuracy of each solver vs number of points reconstructed.

higher $Min(A)$ for $|\mathbf{Z}| = 32$ points. These differences grow very significantly when we focus on cases showing destructive interference, with phase retrieval algorithms showing ~18% higher $Min(A, p5)$ than *Naive* for $|\mathbf{Z}| = 2$; double for $|\mathbf{Z}| = 8(\sim108\%$ higher $Min(A, p5)$) and ten-fold increases for $|\mathbf{Z}| = 32(\sim1125\%$ higher $Min(A, p5)$).

It is also worth noting that, even in these destructive cases, phase retrieval methods retain very similar levels of pressure to the general case in Figure 4b (95±5%, 99±3% and 96±2% for *IBP*, *Long* and *GS-PAT* respectively). Also, the large differences in $Min(A, p5)$ compared to *Naive* are not due to the selection criteria (i.e. 5% of worst geometries for the *Naive* solver). A similar analysis (5% of worst geometries for every other solver) is included in Supplementary Material, still showing better performance of all phase retrieval methods vs *Naive* and retaining good pressures, even for their worst cases.

4.1.2 Comparative performance across phase retrieval methods. The relative performance of the phase retrieval solvers (*IBP*, *Long*, *GS-PAT* varies according to the parameter ($AVG(A)$, $Min(A)$ and $Min(A, p5)$) and condition tested ($|\mathbf{Z}| \in \{2, 4, 8, 16, 32\}$), but differences are small among solvers with *GS-PAT* generally performing in a second position.

In terms of $AVG(A)$, *IBP* provides highest results but also highest standard deviations, followed by *GS-PAT* and then *Long* providing minimum $AVG(A)$ and deviations. Differences among the algorithms are in any case small, with maximum differences of ~0.3% $AVG(A)$ for $|\mathbf{Z}| = 2$, going up to 8% for $|\mathbf{Z}| = 32$.

The analysis in terms of $Min(A)$ shows the added value brought by the regularization used by Long et al's method, resulting in higher amplitudes for these weakest points and lowest standard deviations, hence providing better reconstruction even for these weaker points. *Long* is followed by *GS-PAT* in second place and with *IBP* providing lowest $Min(A)$ and highest standard deviations. Again, the differences among phase retrieval solvers are still small (maximum differences of 5.8% for $|\mathbf{Z}| = 2$; up to 11.4% for $|\mathbf{Z}| = 32$).

These tendencies are retained for destructive cases $Min(A, p5)$. *Long* provides best results, followed by *GS-PAT* and *IBP* in the third place, with small differences among them (i.e. maximum differences of 17.7% for $|\mathbf{Z}| = 2$; decreasing to 11% for $|\mathbf{Z}| = 32$).

This evaluation shows how the added value of phase-retrieval solvers lies not so much in increasing the intensity of their reconstructions (i.e. low gains in $AVG(A)$, even if compared to *Naive*).

Their benefits lie on the ability to provide robust reconstructions even for challenging geometries (i.e. avoid destructive interference), and benefits grow with more points (i.e. more degrees of freedom to exploit and avoid interference). However, differences in performance between phase retrieval solvers are small, and the potential gains obtained from using more complex solvers (e.g. regularization methods used by *Long*) should be weighed against the much higher computing rates enabled by *GS-PAT*, which are key to enable the range of novel PAT applications exemplified in Section 5.

4.1.3 Accurate amplitude reconstruction. The previous evaluation compared the performance of the four algorithms tested when all transducers were used at full power (i.e. $a_t = 1, \forall t \in \mathbf{T}$). This is useful when algorithms are used for phase-only PATs, but also for applications aiming to maximize the total power delivered, such as multi-particle levitation or stronger tactile experiences.

However, other applications, such as the creation of parametric sound or accurate tactile experiences rely on the recreation of specific pressure levels, and this evaluation focuses on these aspects.

Like before, we generated 5000 random geometries, 1000 for each condition ($|\mathbf{Z}| \in \{2, 4, 8, 16, 32\}$ points), with points constrained to the $12 \times 12 \times 12$cm at the centre of our experimental setup, and the same geometries tested across all solvers. As a critical difference, the target amplitude of each point varied between $0.1 \cdot A\_z$ and $0.9 \cdot A\_z$, with A_z being the average amplitude measured for each $|\mathbf{Z}|$ condition during the previous test. This conservative range for target amplitudes was used to ensure that the solvers should be able to achieve the target amplitudes for the points tested (i.e. reconstruction errors due to limitations in each solver, not due to solvers recreating amplitudes beyond the capabilities of the PAT).

Our results are summarized in Figure 4d 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 (*Naive* and *IBP*) could not achieve accurate amplitude reconstructions, producing points that exceeded intended amplitudes ($AVG(Accuracy) \gg 1$) and with very high standard deviations.

Long produced results with very low standard deviation (i.e. similar accuracy for all points reconstructed), but the amplitudes achieved were always smaller than the target (i.e. $AVG(Accuracy)$ <1). $AVG(Accuracy)$ was worse for low number of points (i.e. ~0.77 and ~0.83 for $|\mathbf{Z}| = 2$ and $|\mathbf{Z}| = 4$ points, respectively), but improved for geometries with more points (i.e. $0.91 < AVG(Accuracy) < 1$, for $|\mathbf{Z}| \geq 16$ points). This could be the result of the regularization policy used, which intends to homogenise transducers' output rather than reconstruction accuracy. A simple correction for Long is provided in Supplementary Material (and shown in Figure 4d), achieving accurate reconstructions ($AVG(Accuracy) \simeq 0.99$, for all $|\mathbf{Z}|$) and retaining similarly low variability.

Finally, *GS-PAT* produced $AVG(Accuracy) \simeq 0.98$ across all conditions, but the variability increased with the number of points ($SD(Accuracy) > 0.14$, for $|\mathbf{Z}| \geq 16$), indicating that the algorithm can only provide accurate reconstructions for soundfields using a reduced number of points, such as those used in our examples.

4.2 Algorithm computing performance

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 $|\mathbf{T}| = 512$ transducers, for a varying number of points and on different graphics cards.

Particularly, we tested geometries with $|\mathbf{Z}| = \{2, 4, 8, 16, 32\}$ points on three different GPUs: a low-end GPU (NVIDIA GTX 1050), a mid-end GPU (NVIDIA GTX 1660 Ti) and a high-end GPU (NVIDIA RTX 2080), with results summarized in Table 2.

Our algorithm provides a vast increase in performance (i.e. 100x) when compared to previous reference implementations of multi-point algorithms. For instance, Marzo et al. report computing rates of 90 *gps* ($|\mathbf{T}| = 512$; $|\mathbf{Z}|$ not specified), while Long et al. report rates of ~100*gps* with full regularization and ~200*gps* with no regularization ($|\mathbf{Z}| = 32$; $|\mathbf{T}|$=256).

More importantly, our implementation can, for any number of points and even for a low-end GPU, compute more than 10K*gps* in real time. This is the lower limit of for optimum particle control identified by [Hirayama et al. 2019], allowing maximum acceleration of the particles levitated. Also, being computed in real time allows *GS-PAT* experiences to be interactive. Mid-end GPUs will be required to match the update rates used for tactile spatio-temporal modulation (i.e., [Frier et al. 2019] reported using 17K*gps*, although

Table 2. Geometries per second (gps) computed by our implementation for 3 different graphics cards, according to the number of points in the geometry.

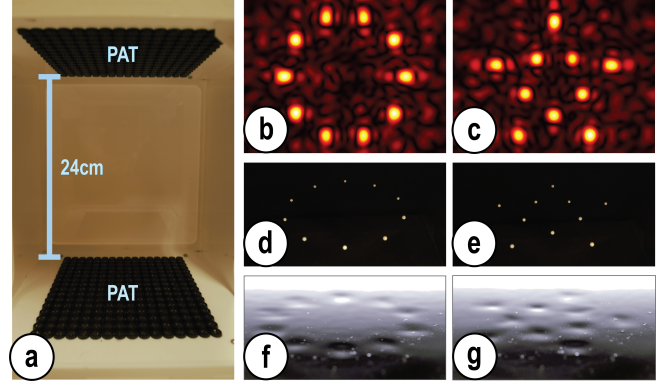| $|\mathbf{Z}|$	GTX 1050	GTX 1660TI	RTX 2080
2	15905 *gps*	23711 *gps*	23248 *gps*
4	14141 *gps*	21717 *gps*	23198 *gps*
8	13797 *gps*	19427 *gps*	23045 *gps*
16	13681 *gps*	18026 *gps*	18709 *gps*
32	11773 *gps*	17168 *gps*	17212 *gps*

Fig. 5. (a) Overview of our experimental top-bottom setup. (b, c) Example sound-fields, and resulting levitation (d, e) and tactile points on liquid (f, g).

no justification is provided as per why this particular rate is used). Also, while our update rates cannot deliver the whole auditory spectrum (i.e. 44.1KHz), these are in all cases enough to encode human voice (8KHz, as used in G.711), to address the human primary auditory range (i.e. 2KHz-5KHz) [Tubbs 2010], or even the notes in a piano (i.e. A0 (27.5Hz) - C8 (4186 Hz)), as shown in our video figure.

5 EXAMPLE APPLICATIONS

This section describes examples, created to showcase the potential of *GS-PAT* and the benefits from transitioning from high-speed single-point fields to high-speed multi-point fields with adjustable control of each points' intensity. Our exploration demonstrates advances in three related application areas: multi-particle POV display, mid-air haptics, sound, as well as combinations of these modalities.

5.1 Hardware and Experimental setups

Our applications and experiments were created using PATs of 16×16 transducers, designed as an extension of the Ultraino platform [Marzo and Drinkwater 2019], modified for faster communication rates. The arrays used Murata MA40S4S transducers (40KHz, 10.5mm diameter (~1.2λ), delivering ~ 8.1Pa at 1m distance when driven at 20 Vpp). A Waveshare CoreEP4CE10 Field Programmable Gate Array (FPGA) board was used to receive phase and amplitude updates from the CPU, using a USB FT245 Synchronous FIFO Interface at 40Mbyte/sec and allowing for 20K phase and amplitude updates per second. For levitation examples, we used two opposed arrays, aligned on top of each other and with a separation of 24cm (see Figure 5a). For our sound-field measurements (e.g. Figure 7), we used a modified 3D printer (OpenBuilds Sphinx 55), with extruder replaced by a calibrated Brüel & Kjer 4138-A-015 microphone, reporting pressure in Pa (rms). For direct visualizations (e.g. Figure 5f&g), we used an edge lit tray 20cm below the top array, filled with a mixture of honey and water to adjust its viscosity.

5.2 Multi-particle POV particle displays

Particle based displays allow the creation of volumetric content in mid-air, for users freely placed around the display and not requiring any instrumentation (e.g. headsets). Such displays have been
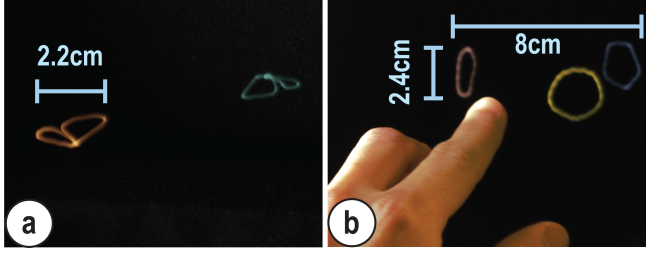
Fig. 6. Examples of POV content. (a) Two independent butterflies, showing dynamic animation. (b) Three circles with real time control using gestures.

Fig. 7. Tactile approaches and main parameters: (a) Spatio-temporal modulation technique; (b) multi-point extension enabled by GS-PAT.

implemented first by detecting and illuminating free moving dust particles in mid-air [Perlin et al. 2006], and later by actively trapping, moving and illuminating particles using lasers [Smalley et al. 2018] or PATs [Hirayama et al. 2019]. The Persistence of Vision (POV) effect is exploited, with particles scanning the content within 0.1s to reveal the 3D shape [Bowen et al. 1974] and only one trapped particle at a time had been used so far.

The use of a single particle maximizes the stiffness of the trap , allowing higher accelerations and being beneficial to render sharper features (e.g. corners). In contrast, the use of several high-speed particles reduces maximum accelerations that can be applied to each particle and can limit content to more curved paths. This reduction in acceleration is the result of the decrease in amplitude observed as the number of points increases (i.e. see Figure 4a) which, as discussed in section 4 and supplementary material, leads to lower stiffness (i.e. trapping forces) and, in turn, to lower accelerations.

However, single particles also limit versatility for the presentation of content made of independent parts (e.g. eyes and mouth in an emoji face), with the particle needing to travel between features while not being illuminated (as to not be visible). Even if impressive maximum values for speed and acceleration have been demonstrated (i.e. $8.75 m/s$ and $1600 m/s^2$), in practise examples created by [Hirayama et al. 2019] were limited to rendering 1.3 meters of *content per second* and contents of up to 2 cm in diameter.

In comparison, the use of several particles allows a higher level of versatility on how the power of the PAT is distributed, as a single particle system might not need to make use of all this power all the time, or waste it as the particle travels between features without being illuminated (no content delivered).

For instance, we managed to reproduce two butterflies of similar sizes to [Hirayama et al. 2019] (i.e. 1 particle per butterfly, wingspan $2.2 cm$, scanning speed ~$0.7 m/s$ each, see Figure 6a), rendering up to 1.4m of *content per second* in total, while showing dynamic animation (i.e. butterflies flapping their wings while flying in circles).

The use of several particles also allows content made of independent parts, which can span across longer areas (i.e. avoiding a single particle travelling from part to part). Figure 6b shows content made of 3 independent parts (1 particle per circle at 0.75m/s each, $2.4 cm\varnothing$, 2.25m of *content per second*), with each circle placed at 4 cm from the centre, covering a larger area. The user can also control the rotation of the circles using hand gestures (detected using Leap motion), a feature enabled by the real-time nature of *GS-PAT*. The ability to
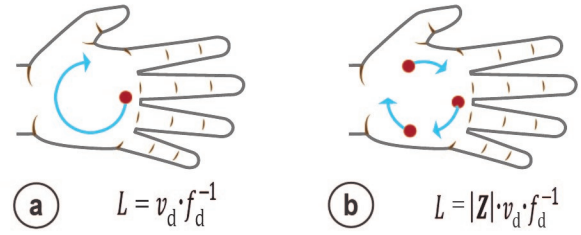
control the trapping stiffness of each particle independently also allows more flexibility on how content is presented. Figure 1 a shows an example of a larger and more complex POV content, with 3 particles being used to render the larger outline of the face ($8.4 cm\varnothing$, $0.92 m/s$ each) and, 1 particle used per extra feature (eyes $0.9 cm\varnothing$, $0.28 m/s$; mouth $1.2 cm\varnothing$, $0.38 m/s$), and power adjusted according to the complexity of each feature (i.e. relative stiffness 1, 0.82 and 0.76 for face, eyes and mouth respectively). This example illustrates how *GS-PAT* can be used to dynamically adjust intensities to the complexity and lengths of each path/feature being revealed. Such control could also be used to maximize intensity when rendering corners or high curvature features (i.e. high accelerations required), or reducing it along linear paths (i.e. low acceleration), using that power to reinforce traps delivering other parts of the content.

5.3 Multi-point mid-air haptics

Our approach can reproduce existing techniques in haptics, such as multi-point spatial multiplexing (Figure 5f & g), lateral modulation [Takahashi et al. 2018] or spatio-temporal modulation [Frier et al. 2019], but we focus our exploration on the later. The spatio-temporal modulation approach, uses a single fast moving focus point to render tactile shapes/paths on the users' skin, as illustrated in Figure 7a. The speed at which the focus point moves on the user's skin is defined as drawing speed (v_d) while the drawing frequency (f_d) is defined as the number of times per second each point on the user's skin is stimulated.

Both parameters influence the users' tactile perception [Ablart et al. 2019; Frier et al. 2018, 2019; Takahashi et al. 2019] but, for a given path length (L), both parameters also become related (i.e. $L = v_d \cdot f_d^{-1}$). To date, it has not been possible to explore their individual effects on tactile response by using (single-point) spatio-temporal modulation alone. Also, the approach is limited to single frequency tactile sensations (i.e. defined by f_d).

GS-PAT enables the use of multi-point spatio-temporal modulation (see Figure 7b), solving these issues. For any given length (L) the relationship between parameters can be mediated by the numbers of points used (i.e. $L = |\mathbf{Z}| \cdot v_d \cdot f_d^{-1}$). That is, for a fixed L and v_d, using more points increases the effective f_d delivered (i.e. f_d proportional to $|\mathbf{Z}|$). Also, for a fixed L and f_d, higher $|\mathbf{Z}|$ effectively reduces the required v_d (i.e. v_d inversely proportional to $|\mathbf{Z}|$).

We explored the technical feasibility of multi-point spatio-temporal modulation by creating three different tactile shapes (i.e. triangle, circle and square) at $11 cm$ above a single PAT array. Each shape

was created for a varying number of points ($|\mathbf{Z}| = \{1, 2, 3, 4\}$) and drawn at two different speeds ($v_d = L/10$ and $v_d = L/20 m/s$). As shown in Table 3 for any given shape and speed v_d, four different frequencies f_d can be delivered (i.e. or alternatively, the same f_d can be delivered at different v_d).

We measured the time-averaged pressure delivered for each of these tactile patterns using our B&Q measuring setup (see Figure 8a-d, remaining plots available in Supplementary Material), finding how it is possible achieve such versatile control of v_d and f_d, while producing very similar pressure distributions on the users' skin.
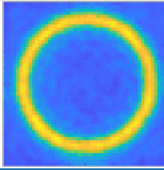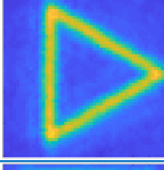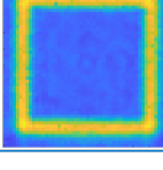
We then analysed these pressure distributions, always comparing them to those produced by single-point stimulation. More specifically, from the data for $|\mathbf{Z}| = 1$ of each shape, we created a binary mask identifying its *stimulation area* (i.e. white area in Figure 8e, where high pressure was delivered) and its *surrounding area* (in black, where minimum pressure should be found). We then used these masks to compare how pressure was distributed across areas when that shape was created with higher number of points (i.e. $|\mathbf{Z}| = \{2, 3, 4\}$).

The threshold between areas was set at 30% of the observed peak amplitude based on visual inspection of the masks. Higher threshold values retained bright halos (high-pressures) at the edges of the shapes within the *surrounding area* (masks used for visual inspection for different thresholds can be found in Supplementary Material). Finally, we computed the overall pressure delivered by each condition inside and outside the *stimulation area* (i.e. *Inside(Pa)* and *Outside(Pa)* in Table 3), as well as the difference in average pressures between the two areas (i.e. *Contrast(Pa)*, in Table 3).

Our results show that multi-point conditions enabled by *GS-PAT* tend to deliver slightly higher values of pressure inside the *stimulation area* (i.e. *Inside(Pa)*), with smother distributions of pressure (i.e. lower *SD*). However, they also tend to increase pressure in the *surrounding area* due to secondary lobes (i.e. these increase with the number of points used). The overall differences between areas (i.e. the *Contrast*), remains very similar for conditions using 1 to 3 points, but decreases when 4 points are used (i.e. lower than for the single-point condition).

These results seems to indicate that, while the use of multi-point spatio-temporal modulation provides a valuable tool to allow the independent exploration of the effects of v_d and f_d on human's

Table 3. Parameters of the tactile shapes created using multi-point spatio-temporal modulation.

Shape	|Z|	f_d (Hz)	v_d (m/s)	Inside (Pa)	Outside (Pa)	Contrast (Pa)
	1	10	2.51	904 ± 303	286 ± 63	618
		20	5.02	914 ± 316	286 ± 81	629
	2	20	2.51	991 ± 267	324 ± 96	688
		40	5.02	978 ± 267	310 ± 94	667
	3	30	2.51	981 ± 273	344 ± 83	638
		60	5.02	961 ± 275	328 ± 80	633
	4	40	2.51	957 ± 273	362 ± 88	590
		80	5.02	937 ± 269	347 ± 83	595
	1	10	2.40	906 ± 313	281 ± 85	760
		20	4.80	909 ± 303	279 ± 86	758
	2	20	2.40	946 ± 267	317 ± 94	764
		40	4.80	947 ± 288	310 ± 98	766
	3	30	2.40	913 ± 290	335 ± 84	712
		60	4.80	923 ± 308	325 ± 85	726
	4	40	2.40	900 ± 291	333 ± 80	701
		80	4.80	904 ± 313	322 ± 80	710
	1	10	3.60	875 ± 245	278 ± 53	596
		20	7.20	832 ± 254	270 ± 47	563
	2	20	3.60	939 ± 205	321 ± 76	618
		40	7.20	896 ± 232	303 ± 74	593
	3	30	3.60	940 ± 201	350 ± 78	590
		60	7.20	902 ± 220	330 ± 75	572
	4	40	3.60	914 ± 209	356 ± 80	557
		80	7.20	879 ± 223	339 ± 81	541

tactile response, the use of higher numbers of points can decrease the overall perceived intensity (i.e. *Contrast*) of the tactile sensations and power adjustments will be needed.

Our multi-point spatio-temporal technique can also use different drawing frequencies for each point, to provide multi-frequency stimulation (i.e. each point as a carrier for a specific frequency), or with extra points reinforcing salient features [Martinez et al. 2019], such as corners. While the exploration of all these possibilities fall beyond the scope of the current paper, they serve to illustrate the versatility enabled by *GS-PAT* to create tactile experiences.

5.4 Audio and multimodal combinations

The high update rate allowed by *GS-PAT* offers interesting potential for parametric audio, either to combine audio with other modalities
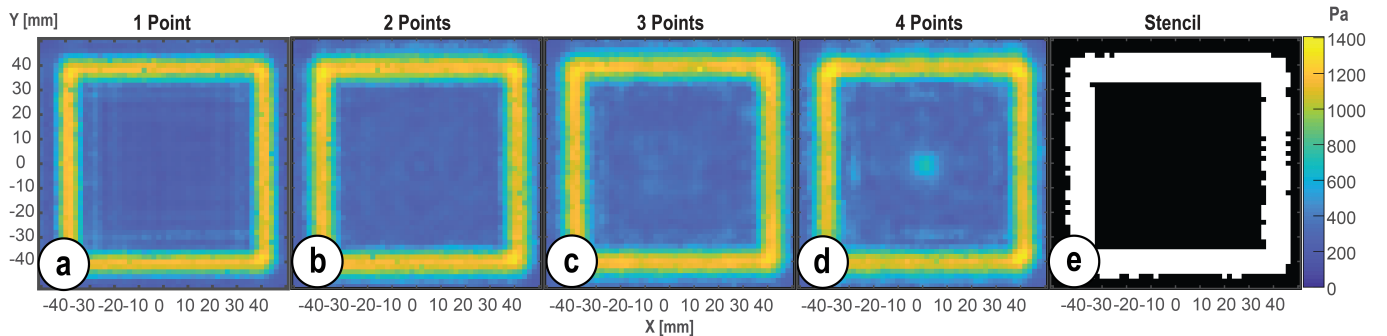

Fig. 8. Measurements of the time-averaged pressure delivered for the square for (a) 1, (b) 2, (c) 3 and (d) 4 tactile points. (e) Binary mask (derived from (a)) used to identify the stimulation and surrounding areas.
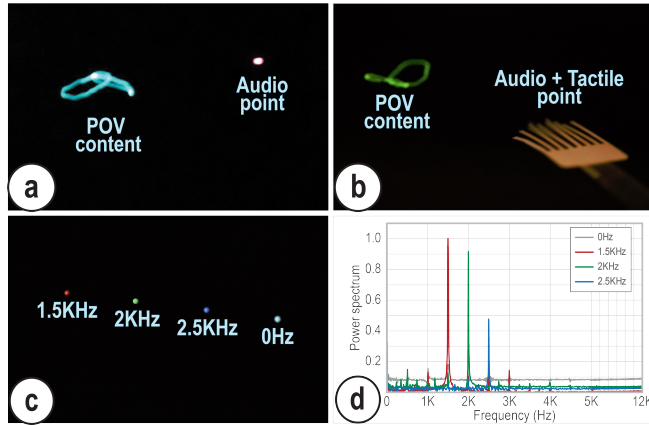
Fig. 9. Examples of multimodal applications for (a) two and (b) three modalities. (c) Simultaneous deliver of several audio streams with levitation. (d) Audio frequencies at each of the four point sources generated.

(POV displays) or to generate multiple spatialized audio streams. This is achieved by encoding each audio source independently using single-sided band-modulation [Sakai and Kamakura 2008] and using the resulting time-domain signal as the amplitude for each of our target points over time.

To showcase the flexibility enabled by *GS-PAT*, we reproduced one of the examples in [Hirayama et al. 2019], as shown in Figure 9a. We set our update rate at 10 KHz, matching their spatial multiplexing frequency. One point was used to create a POV butterfly (constant amplitude, 2.2*cm* wingspan) and the second point created a sound source at the centre of the setup (chirp signal 100Hz-5KHz ; modulation index $a = 0.4$) resulting in audible sound of 71.0dB, 69.5dB and 66.0dB for frequencies 1.5KHz, 2KHz and 2.5KHz respectively.

Besides, *GS-PAT*'s ability to control the intensity of points independently allows more flexibility to create content. By adjusting the relative intensities for levitation and audio to 1 and 0.5, we could create a larger butterfly (3.5cm), at the expense of decreasing audio levels 65.7dB, 63.1dB and 60.1dB).

The second (audible) point could also be used to deliver tactile feedback using spatio-temporal modulation (Figure 9b), allowing the delivery of all 3 modalities (visual, audio and touch) with only 2 points. It is also interesting to note that the visual content introduces a small audible component at 10Hz audible by the naked ear (also present in single-point POV).

Figure 9c illustrates the possibility to simultaneously deliver several audio streams with simultaneous levitation. We used four levitation points aligned along the centre of our system (4*cm* separation), with 3 of them each encoding a different sound (i.e. 1.5KHz, 2KHz and 2.5KHz) and the last one maintaining constant pressure (i.e. no sound). We then measured the spectrogram at each of four target points using the Brüel & Kizer 4138-A-015 microphone.

As shown in Figure 9d, good localized audible frequencies are produced at each point, with the expected presence of components at harmonic frequencies (e.g. the spectrogram at 1.5KHz, also shows small components at 3KHz, 4.5KHz , etc.). Also, minimum parasitic contribution from the other points can be found. At the same time,

all three combined sounds are audible by a user in front of the device, as shown in our video figure. It is also worth noting the absence of aliasing effects at 10 KHz (our update rate), which would be present with the time multiplexing technique used in single-point approaches [Hirayama et al. 2019].

Although *GS-PAT* still does not allow rates of 40KHz, required to cover the whole audible spectrum, it offers some advantages compared to [Hirayama et al. 2019], such as in reducing audible artefacts or allowing the recreation of several audible streams.

6 CONCLUSION

This paper presented *GS-PAT*, a phase retrieval algorithm that allows computation of multi-point sound-fields at very high interactive rates (e.g. >17KHz in an NVIDIA GTX1660), with applications in haptics, levitation, parametric audio or even to combinations of these modalities. Our comparison against related algorithms (i.e. [Long et al. 2014; Marzo and Drinkwater 2019]) allowed us to assess the benefits introduced by such phase retrieval techniques (i.e. great robustness against destructive interference cases) and showed that *GS-PAT* can provide performance in line with these slower solvers, while enabling much faster rates. Accurate amplitude reconstructions however are limited to reduced numbers of points (e.g. $|\mathbf{Z}| \leq 8$), and practitioners focusing on the use of many points and very accurate reconstructions, but slower rates should refer to global regularization based approaches (i.e. possibly with the corrections described in our Supplementary Material).

But most importantly, this paper showed how the use of *GS-PAT* enables a shift in paradigm for PAT devices. By moving from the use of single-point high-speed sound-fields, to the use of several high-speed (tactile or levitation) points, *GS-PAT* offers great versatility on the management of the power delivered by the PAT.

This versatility is illustrated in our exploration of applications which, even if preliminary, demonstrates multiparticle POV displays, with more flexible content delivery enabled by the use of several particles to render different features, or by the more flexible distribution of power according to the properties of the rendering paths. When applied to haptics, our approach enables the use of multi-point spatio-temporal modulation, offering interesting possibilities to modulate drawing speed and frequency independently, or enabling multi-frequency stimulation. We also demonstrated the possibility of delivering audible sound, while removing artefacts present in single-point approaches and extending this concept to the delivery of several audio streams or in combination with other modalities. We hope that our choice of multiplatform, open-source software and open hardware will also allow others to help us explore these new and exciting opportunities for PAT technologies.

REFERENCES

Damien Ablart, William Frier, Hannah Limerick, Orestis Georgiou, and Marianna Obrist. 2019. Using Ultrasonic Mid-air Haptic Patterns in Multi-Modal User Experiences. In *2019 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*. IEEE, IEEE, New York, NY, USA, 1–6.

Marco AB Andrade, Thales SA Camargo, and Asier Marzo. 2018. Automatic contactless injection, transportation, merging, and ejection of droplets with a multifocal point acoustic levitator. *Review of Scientific Instruments* 89, 12 (2018), 125105.

Richard W Bowen, Jordan Pola, and Leonard Matin. 1974. Visual persistence: Effects of flash luminance, duration and energy. *Vision Research* 14, 4 (1974), 295–303.

Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. Association for Computing Machinery, New York, NY, USA, 505–514.

J. Christopher Dainty and James R. Fienup. 1987. Phase retrieval and image reconstruction for astronomy.

Fajwel Fogel, Irène Waldspurger, and Alexandre d'Aspremont. 2016. Phase retrieval for imaging problems. *Mathematical programming computation* 8, 3 (2016), 311–335.

Euan Freeman, Asier Marzo, Praxitelis B Kourtelos, Julie R Williamson, and Stephen Brewster. 2019. Enhancing physical objects with actuated levitating particles. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays*. Association for Computing Machinery, New York, NY, USA, 1–7.

William Frier, Damien Ablart, Jamie Chilles, Benjamin Long, Marcello Giordano, Marianna Obrist, and Sriram Subramanian. 2018. Using spatiotemporal modulation to draw tactile patterns in mid-air. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, Springer, 270–281.

William Frier, Dario Pittera, Damien Ablart, Marianna Obrist, and Sriram Subramanian. 2019. Sampling strategy for ultrasonic mid-air haptics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–11.

Orestis Georgiou, Craig Jeffrey, Ziyuan Chen, Bao Xiao Tong, Shing Hei Chan, Boyin Yang, Adam Harwood, and Tom Carter. 2018. Touchless haptic feedback for VR rhythm games. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, IEEE, New York, NY, USA, 553–554.

Ralph W Gerchberg. 1972. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* 35 (1972), 237–246.

Keisuke Hasegawa, Liwei Qiu, Akihito Noda, Seki Inoue, and Hiroyuki Shinoda. 2017. Electronically steerable ultrasound-driven long narrow air stream. *Applied Physics Letters* 111, 6 (2017), 064104.

Ryuji Hirayama, Diego Martinez Plasencia, Nobuyuki Masuda, and Sriram Subramanian. 2019. A volumetric display for visual, tactile and audio presentation using acoustic trapping. *Nature* 575, 7782 (2019), 320–323.

Takayuki Hoshi, Masafumi Takahashi, Takayuki Iwamoto, and Hiroyuki Shinoda. 2010. Noncontact tactile display based on radiation pressure of airborne ultrasound. *IEEE Transactions on Haptics* 3, 3 (2010), 155–165.

Seki Inoue, Yasutoshi Makino, and Hiroyuki Shinoda. 2015. Active touch perception produced by airborne ultrasonic haptic hologram. In *2015 IEEE World Haptics Conference (WHC)*. 362–367.

Seki Inoue, Shinichi Mogami, Tomohiro Ichiyama, Akihito Noda, Yasutoshi Makino, and Hiroyuki Shinoda. 2019. Acoustical boundary hologram for macroscopic rigid-body levitation. *The Journal of the Acoustical Society of America* 145, 1 (2019), 328–337. https://doi.org/10.1121/1.5087130

Kent Knox. 2013. CLBLAS. "https://github.com/clMathLibraries/clBLAS".

François Le Gall. 2012. Faster algorithms for rectangular matrix multiplication. In *2012 IEEE 53rd annual symposium on foundations of computer science*. IEEE, IEEE, New York, NY, USA, 514–523.

Benjamin Long, Sue Ann Seah, Tom Carter, and Sriram Subramanian. 2014. Rendering volumetric haptic shapes in mid-air using ultrasound. *ACM Transactions on Graphics (TOG)* 33, 6 (2014), 1–10.

Jonatan Martinez, Adam Harwood, Hannah Limerick, Rory Clark, and Orestis Georgiou. 2019. Mid-Air Haptic Algorithms for Rendering 3D Shapes. In *2019 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*. IEEE, IEEE, New York, NY, USA, 1–6.

Asier Marzo and Bruce W Drinkwater. 2019. Holographic acoustic tweezers. *Proceedings of the National Academy of Sciences* 116, 1 (2019), 84–89.

Asier Marzo, Sue Ann Seah, Bruce W Drinkwater, Deepak Ranjan Sahoo, Benjamin Long, and Sriram Subramanian. 2015. Holographic acoustic elements for manipulation of levitated objects. *Nature communications* 6 (2015), 8661.

Yasuaki Monnai, Keisuke Hasegawa, Masahiro Fujiwara, Kazuma Yoshino, Seki Inoue, and Hiroyuki Shinoda. 2014. HaptoMime: mid-air haptic interaction with a floating virtual screen. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. Association for Computing Machinery, New York, NY, USA, 663–667.

Rafael Morales, Asier Marzo, Sriram Subramanian, and Diego Martinez. 2019. LeviProps: animating levitated optimized fabric structures using holographic acoustic tweezers. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 651–661.

Mohd Adili Norasikin, Diego Martinez, Gianluca Memoli, and Sriram Subramanian. 2019. SonicSpray: A Technique to Reconfigure Permeable Mid-Air Displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces (ISS '19)*. Association for Computing Machinery, New York, NY, USA, 113–122. https://doi.org/10.1145/3343055.3359704

Mohd Adili Norasikin, Diego Martinez, Spyros Polychronopoulos, Gianluca Memoli, Yutaka Tokuda, and Sriram Subramanian. 2018. SoundBender: Dynamic Acoustic Control Behind Obstacles. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. Association for Computing Machinery, New York, NY, USA, 247–259. https://doi.org/10.1145/3242587.3242590

Yoichi Ochiai, Takayuki Hoshi, and Jun Rekimoto. 2014. Pixie dust: graphics generated by levitated and animated objects in computational acoustic-potential field. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 1–13.

Yoichi Ochiai, Takayuki Hoshi, and Ippei Suzuki. 2017. Holographic whisper: Rendering audible sound spots in three-dimensional space by focusing ultrasonic waves. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 4314–4325.

Dirk Olszewski, Fransiskus Prasetyo, and Klaus Linhard. 2005. Steerable highly directional audio beam loudspeaker. In *Ninth European Conference on Speech Communication and Technology*.

Themis Omirou, Asier Marzo, Sue Ann Seah, and Sriram Subramanian. 2015. LeviPath: Modular acoustic levitation for 3D path visualisations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 309–312.

Themis Omirou, Asier Marzo Perez, Sriram Subramanian, and Anne Roudaut. 2016. Floating charts: Data plotting using free-floating acoustically levitated representations. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, IEEE, New York, NY, USA, 187–190.

HoT O'Neil. 1949. Theory of focusing radiators. *The Journal of the Acoustical Society of America* 21, 5 (1949), 516–526.

Kenneth Perlin et al. 2006. Volumetric display with dust as the participating medium. US Patent 6,997,558.

Nikunj Raghuvanshi and John Snyder. 2014. Parametric wave field coding for precomputed sound propagation. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 1–11.

Nikunj Raghuvanshi and John Snyder. 2018. Parametric directional coding for precomputed sound propagation. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–14.

Deepak Ranjan Sahoo, Takuto Nakamura, Asier Marzo, Themis Omirou, Michihiro Asakawa, and Sriram Subramanian. 2016. Joled: A mid-air display based on electrostatic rotation of levitated janus objects. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 437–448.

Shinichi Sakai and Tomoo Kamakura. 2008. Dynamic single sideband modulation for realizing parametric loudspeaker. In *AIP Conference Proceedings*, Vol. 1022(1). American Institute of Physics, AIP Publishing, Melville, NY, USA, 613–616.

Antti Sand, Ismo Rakkolainen, Poika Isokoski, Jari Kangas, Roope Raisamo, and Karri Palovuori. 2015a. Head-mounted display with mid-air tactile feedback. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*. Association for Computing Machinery, New York, NY, USA, 51–58.

Antti Sand, Ismo Rakkolainen, Poika Isokoski, Roope Raisamo, and Karri Palovuori. 2015b. Light-weight immaterial particle displays with mid-air tactile feedback. In *2015 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*. IEEE, IEEE, New York, NY, USA, 1–5.

Gözel Shakeri, Euan Freeman, William Frier, Michele Iodice, Benjamin Long, Orestis Georgiou, and Carl Andersson. 2019. Three-in-one: Levitation, Parametric Audio, and Mid-Air Haptic Feedback. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–4.

Chuang Shi, Yoshinobu Kajikawa, and Woon-Seng Gan. 2015. Generating dual beams from a single steerable parametric loudspeaker. *Applied Acoustics* 99 (2015), 43–50.

DE Smalley, E Nygaard, K Squire, J Van Wagoner, J Rasmussen, S Gneiting, K Qaderi, J Goodsell, W Rogers, M Lindsey, et al. 2018. A photophoretic-trap volumetric display. *Nature* 553, 7689 (2018), 486.

W Le Conte Stevens. 1899. A Text-Book of Physics–Sound.

Ryoko Takahashi, Keisuke Hasegawa, and Hiroyuki Shinoda. 2018. Lateral Modulation of Midair Ultrasound Focus for Intensified Vibrotactile Stimuli. In *Haptics: Science, Technology, and Applications*, Domenico Prattichizzo, Hiroyuki Shinoda, Hong Z. Tan, Emanuele Ruffaldi, and Antonio Frisoli (Eds.). Springer International Publishing, Cham, 276–288.

Ryoko Takahashi, Keisuke Hasegawa, and Hiroyuki Shinoda. 2019. Tactile Stimulation by Repetitive Lateral Movement of Midair Ultrasound Focus. *IEEE Transactions on Haptics* (2019), 1–1.

Rhonda S Tubbs. 2010. Essentials of Audiology. *Otology & Neurotology* 31, 3 (2010), 379–380.

Chi Thanh Vi, Asier Marzo, Damien Ablart, Gianluca Memoli, Sriram Subramanian, Bruce Drinkwater, and Marianna Obrist. 2017. Tastyfloats: A contactless food delivery system. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. Association for Computing Machinery, New York, NY, USA, 161–170.

Jui-Hsien Wang and Doug L James. 2019. KleinPAT: optimal mode conflation for time-domain precomputation of acoustic transfer. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–12.

Jui-Hsien Wang, Ante Qu, Timothy R Langlois, and Doug L James. 2018. Toward wave-based sound synthesis for computer animation. *ACM Trans. Graph.* 37, 4 (2018), 109–1.

Masahide Yoneyama, Jun-ichiroh Fujimoto, Yu Kawamo, and Shoichi Sasabe. 1983. The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design. *The Journal of the Acoustical Society of America* 73, 5 (1983), 1532–1536.

Kentaro Yoshida, Yuuki Horiuchi, Seki Inoue, Yasutoshi Makino, and Hiroyuki Shinoda. 2017. HaptoCloneAR: mutual haptic-optic interactive system with 2D image superimpose. In *ACM SIGGRAPH 2017 Emerging Technologies*. Association for Computing Machinery, New York, NY, USA, 1–2.

Changxi Zheng and Doug L James. 2011. Toward high-quality modal contact sound. In *ACM SIGGRAPH 2011 papers*. Association for Computing Machinery, New York, NY, USA, 1–12.