Applied Acoustics 76 (2014) 230–237

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Automatic 3D scanning surface generation for microphone array acoustic imaging

Mathew Legg *, Stuart Bradley

Physics Department, University of Auckland, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 13 December 2012
Received in revised form 12 June 2013
Accepted 12 August 2013
Available online 14 September 2013

Keywords:
Microphone array
3D
Acoustic imaging
Structured light
Beamforming
CLEAN-SC

ABSTRACT

This work presents a new technique for automatically generating the 3D scanning surface for acoustic imaging using microphone arrays. Acoustic images, or maps, of sound coming from spatially distributed sources, may be generated from microphone array data using algorithms such as beamforming. Traditional 2D acoustic maps can contain errors in the near-field if the object being imaged has a 3D shape. It has been shown that using the 3D surface geometry of an object as a scanning surface for beamforming can provide more accurate results. The methods used previously to generate this 3D scanning surface have either required existing CAD (Computer-Aided Design) models of the object being acoustically imaged or have required separate equipment which is generally bulky and expensive. The new method uses one or more cameras in the array, a data projector, and structured light code to automatically generate the 3D scanning surface. This has the advantage that it is inexpensive, can be incorporated as an add-onto existing microphone arrays, has short scan time, and is capable of being extended to imaging dynamic scenes. This technique is tested using beamforming and *CLEAN-SC* (CLEAN based on spatial Source Coherence) algorithms for a spherical array and an Underbrink multi-arm spiral array. For sound sources located about 1.2 m from the array, the mean position errors obtained are 6 mm. This is a quarter of the diameter of the mini-speakers being used as a sound sources.

1. Introduction

The microphone phased array (also known as the acoustic camera) has been developed as a tool to enable the position and magnitude of sound sources to be identified. Microphone phased arrays are widely used by industries such as aeroplane and automotive manufacturers to identify sound sources. A commonly used acoustic imaging algorithm is beamforming [1,2], which uses delaying and summing of microphone channel data to obtain acoustic images or 'maps'. However, the beamforming maps contain an interference pattern artefact referred to as sidelobes. Image sharpening or deconvolution techniques have been developed to remove these sidelobes and attempt to obtain the true sound source distribution. Examples of these algorithms are *DAMAS* (Deconvolution Approach for the Mapping of Acoustic Sources) [3] and *CLEAN-SC* (CLEAN based on Source Coherence) [4]. The other methods used to generate acoustic maps from microphone phased array data are acoustic holography [5] and inverse methods [6].

1.1. Acoustic imaging using traditional 2D scanning surfaces

Acoustic imaging techniques, such as beamforming and deconvolution, have traditionally used the assumption that the acoustic sources lie on a plane. A 2D scan surface is used which is oriented perpendicular to the array's principle forward direction (the array *Z*-axis). This can lead to errors in the resulting acoustic maps if the sound sources are offset from the 2D scanning surface. These errors appear as projection/parallax errors in the plotting of the acoustic maps [7] and incorrectly estimated sound pressure levels (SPL) and location of sound sources [8–10]. These beamforming magnitude and position errors result from incorrect focus (time delays) being used for the beamforming.

1.2. Acoustic imaging using 3D scanning surfaces

Beamforming and deconvolution have been performed using a 3D grid [11–14]. However, unless the microphone arrays surround the object being imaged, there is poor resolution in the array *Z*-axis. Another problem is that these 3D grids can contain a large number of scan points, making deconvolution of these 3D grid beamforming maps very computationally expensive.

An alternative technique that has been developed is to use a scanning surface for beamforming which corresponds to the 3D

* Corresponding author.
E-mail addresses: m.legg@auckland.ac.nz, mleg010@auckland.ac.nz (M. Legg), s.bradley@auckland.ac.nz (S. Bradley).

0003-682X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apacoust.2013.08.008

surface geometry of the object that is being acoustically imaged. This method is said to provide the correct beamforming focus (time delays) for sound propagating from the surface of the object and should provide more accurate measured sound pressure levels, positions, and plotting. Different methods have been used to generate the 3D scanning surfaces for acoustic imaging. *GFaI* uses a CAD model of an object [15] or a laser scanner for acoustic imaging of the interior of rooms [16]. Pininfarina Full Scale Wind Tunnel tests obtained the 3D surface geometry of the exterior of cars, using stereoscopic imaging and a static pattern projected onto the object. Irimia et al. [10] (*MicrodB/LMS*) uses a laser scanner, attached to a solid spherical array, to obtain the 3D geometry of the interior of cars for acoustic imaging. Another acoustic imaging technique that uses the surface geometry of an object is the acoustic holography technique referred to as conformal mapping. *Brüel & kJær* measure the surface geometry of an object using a sonic contact measurement tool. However, this process can be slow and, therefore, is generally only applied to small areas [17].

1.3. New 3D acoustic imaging technique

The automatic 3D acoustic imaging technique presented in this work was developed as part of research comparing 2D and 3D acoustic imaging methods [26,29,30]. This work was the first to investigate deconvolution (image sharpening) of beamforming maps which were generated using the 3D surface geometry of an object as a scanning surface. A key requirement was the ability to accurately determine this 3D surface geometry relative to the microphone array. This paper presents a technique which was developed to allow the 3D scanning surface to be automatically generated using structured light scanning [18,19]. Structured light scanning was decided on because it is robust, inexpensive, fast, and capable of imaging dynamic scenes. Most microphone phased arrays contain a camera and hence only the addition of a data projector and software is required to convert these microphone arrays into a structured light scanner.

2. Acoustic imaging theory

Consider sound propagating through air from sound sources located at $\vec{\boldsymbol{X}}_s$ to M microphones in an array having coordinates $\vec{\boldsymbol{X}}_m$. The time domain microphone data may be converted into the frequency domain by dividing it into J blocks of length K, and for each data black obtaining a $(M \times 1)$ element array of FFT (Fast Fourier Transform) coefficients $\boldsymbol{U}$ for each frequency bin. The resulting frequency domain microphone data may be modeled as

$$\boldsymbol{U}(j) = \sum_{s=1}^{S} \mathbf{C}(\vec{\boldsymbol{X}}_s) Q(\vec{\boldsymbol{X}}_{s,j}) + \boldsymbol{E}(j), \quad j = 1 \ldots J \tag{1}$$

where $\boldsymbol{C}$ is the $(M \times 1)$ array propagation vector, Q is the sound source strength, and $\boldsymbol{E}$ is an $(M \times 1)$ element array of uncorrelated channel noise coefficients. A generic term for the array propagation vector may be given by

$$C_m(\vec{\boldsymbol{X}}_s) = \frac{\exp\left(-i\omega\sigma_m(\vec{\boldsymbol{X}}_s)\right)}{\Lambda_m(\vec{\boldsymbol{X}}_s)}, \tag{2}$$

where ω is the angular frequency, $\sigma_m(\vec{\boldsymbol{X}}_s)$ is the acoustic propagation time from the sound source at $\boldsymbol{X}_s$ to a microphone at $\vec{\boldsymbol{X}}_m$, and $\Lambda_m(\vec{\boldsymbol{X}}_s)$ is a term to allow for spherical spreading [20].

A frequency domain beamforming $(N \times 1)$ acoustic map $\boldsymbol{b}$ may then be generated by defining a grid of scan points located at $\vec{\xi}$ and calculating

$$\boldsymbol{b}(\vec{\xi}_n) = \left\langle \left| \mathbf{w}^{\dagger}(\vec{\xi}_n) \boldsymbol{U} \right|^2 \right\rangle, \quad \vec{\xi}_n = \vec{\xi}_1 \ldots \vec{\xi}_N \tag{3}$$

where $\langle\cdot\rangle$ is the time average and $\mathbf{w}$ is an $(M \times 1)$ array steering vector. A generic term for the mth element of the array steering vector may be given by

$$w_m(\vec{\xi}_n) = \frac{\exp(i\omega\tau_m(\vec{\xi}_n))}{D_m(\vec{\boldsymbol{X}}_s)}, \tag{4}$$

where ω is the angular frequency, τ_m and is the beamforming time delay, and D_m is an amplitude correction factor. Using Eq. (1) and assuming a unit magnitude source, one may model Eq. (3) as

$$\boldsymbol{b}(\vec{\xi}_n) = \mathbf{w}^{\dagger}(\vec{\xi}_n) \boldsymbol{C}(\vec{\boldsymbol{X}}_s) \boldsymbol{C}^{\dagger}(\vec{\boldsymbol{X}}_s) \mathbf{w}(\vec{\xi}_n). \tag{5}$$

Expanding this equation and substituting in Eqs. (2) and (15) gives

$$\boldsymbol{b}(\vec{\xi}_n) = \sum_{m,m'=1}^{M,M} \frac{\exp\left(i\omega\left[d\sigma_{mm'}(\vec{\boldsymbol{X}}_s) - d\tau_{mm'}(\vec{\xi}_n)\right]\right)}{D_m(\vec{\xi}_n)\Lambda_m(\vec{\boldsymbol{X}}_s)\Lambda_{m'}(\vec{\boldsymbol{X}}_s)D_{m'}(\vec{\xi}_n)} \tag{6}$$

where $d\sigma_{mm'}(\vec{\boldsymbol{X}}_s)$ and is the difference in the propagation time from the sound source at $\vec{\boldsymbol{X}}_s$ to the two microphones respectively at coordinates $\vec{\boldsymbol{X}}_m$ and $\vec{\boldsymbol{X}}_{m'}$ and $d\tau_{mm'}(\vec{\xi}_n)$ is the corresponding difference in the beamforming time delays used for theses two microphones for scan point $\vec{\xi}_n$.

2.1. Beamforming phase error

Traditionally acoustic maps have been generated using 2D scanning surfaces. If the scanning surface is offset from the sound source location, errors can occur in the amplitude and position of sound sources in beamforming acoustic maps. These errors may result from incorrect focus (time delays) being used for the beamforming. To explain this, one may consider sound waves propagating under free-field conditions through still, homogeneous air from a sound source located at position $\vec{\boldsymbol{X}}_s$ to two microphones located respectively at positions $\vec{\boldsymbol{X}}_m$ and $\vec{\boldsymbol{X}}_{m'}$ in an array containing M microphones. The difference in acoustic propagation time between the two microphones may be described by

$$d\sigma_{mm'}(\vec{\boldsymbol{X}}_s) = \frac{\|\vec{\boldsymbol{X}}_m - \vec{\boldsymbol{X}}_s\| - \|\vec{\boldsymbol{X}}_{m'} - \vec{\boldsymbol{X}}_s\|}{c}, \tag{7}$$

where c is the speed of sound.

Now consider that the beamforming is performed for the nth scan point with coordinate $\vec{\xi}_n$. The difference between the beamforming time delays for the two microphone signals would be

$$d\tau_{mm'}(\vec{\xi}_n) = \frac{\|\vec{\boldsymbol{X}}_m - \vec{\xi}_n\| - \|\vec{\boldsymbol{X}}_{m'} - \vec{\xi}_n\|}{c}. \tag{8}$$

The maximum constructive interference of the two beamformed microphone signals should occur when $d\sigma_{mm'}(\vec{\boldsymbol{X}}_s) = d\tau_{mm'}(\vec{\xi}_n)$.

The beamforming phase error for two channels of microphone data may, therefore, be described by

$$\Delta_{mm'}(\vec{\boldsymbol{X}}_s, \vec{\xi}_p) = \frac{2\pi c}{\lambda}\left[d\sigma_{mm'}(\vec{\boldsymbol{X}}_s) - d\tau_{mm'}(\vec{\xi}_p)\right], \tag{9}$$

where λ is the wavelength and $\vec{\xi}_p$ is the theoretical coordinate in an acoustic map where a peak should occur due to a sound source at $\vec{\boldsymbol{X}}_s$. A phase error of zero will give maximum constructive interference, while a phase error of π will give the maximum destructive interference of the beamformed microphone signals. Using perspective projection from the center of the microphone coordinates $\vec{\boldsymbol{X}}_{mc}$ to a sound source at $\vec{\boldsymbol{X}}_s$, the theoretical coordinates of $\vec{\xi}_p$ may be modeled as

$$\vec{\xi}_p = (\vec{\boldsymbol{X}}_s - \vec{\boldsymbol{X}}_{mc})\frac{Z}{Z_s} + \vec{\boldsymbol{X}}_{mc}, \tag{10}$$

where Z is the z component of $\vec{\xi} - \vec{X}_{mc}$ and Z_s is the z component of $\vec{X}_s - \vec{X}_{mc}$.

As an example, consider the microphone, sound source, and 2D scanning surface configuration shown in Fig. 1. The theoretical phase error obtained during beamforming for this configuration, may be approximated as

$$\Delta_{mm'}(\vec{X}_s, \vec{\xi}_p) \approx \frac{\pi D^2}{\lambda}\left[\frac{1}{R} - \frac{1}{r}\right]. \quad (11)$$

This equation demonstrates that the beamforming phase error increases as.

- The offset $|R - r|$ of the 2D scanning surface from the sound source location increases.
- The wavelength λ reduces.
- The distances R and r reduces.
- The separation D of the microphones increases.

However, it will also depend on the relative positions of the microphones. This beamforming error is eliminated if $R = r$. This is achieved if a 3D scanning surface is used which corresponds to the 3D surface geometry of the object emitting sound.

3. Equipment and methodology

Figs. 2 and 8(a) respectively show the structured light setup used for a 72 element, 0.6 m diameter spherical array and a 72 element, 0.9 m diameter Underbrink multi-arm spiral array. Although a full size data projector was used here, an inexpensive mini-data projector could instead have been used.

3.1. Calibration

3.1.1. Camera and projector calibration

To achieve accurate structured light scans, the cameras and projector need to be calibrated. The projector can be modeled as a camera where the light travels in the opposite direction to usual. The calibration was achieved using a combination of modified versions of Lanman and Taubin's structured light *MATLAB* code "mlStructuredLight" [21], Falcao et al.'s "Camera-Projector Calibration Toolbox" [22], and Bouguet's "*MATLAB* Camera Calibration Toolbox" [23].

Calibration parameters were obtained by projecting a checkerboard pattern onto the screen which has a second, printed, checkerboard pattern glued onto it, see Fig. 3. An image is captured of the projected checkerboard pattern. The projected pattern is then turned off. Without moving the screen, an image is then captured of the printed checkerboard pattern on the screen. This is repeated for a range of positions and orientations of the screen. Camera calibration is first performed. For each image of the printed checkerboard pattern, the pixels corresponding to the checkerboard pattern corners are obtained. These, combined with the dimensions of each checker square, are used to obtain the camera's intrinsic and extrinsic calibration parameters [24,25,23]. The intrinsic parameters describe the projection of light from a point in front of the camera onto a 2D pixel coordinate behind the

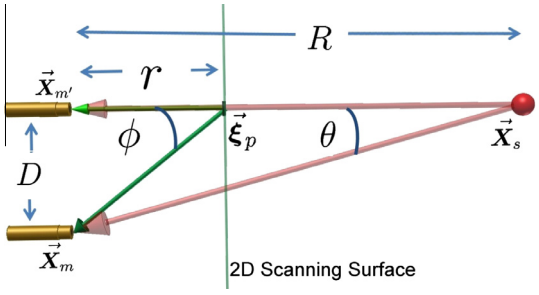

Fig. 1. Simplified geometry for two microphones in a planar array to enable a general/simplified beamforming phase error equation to be obtained.

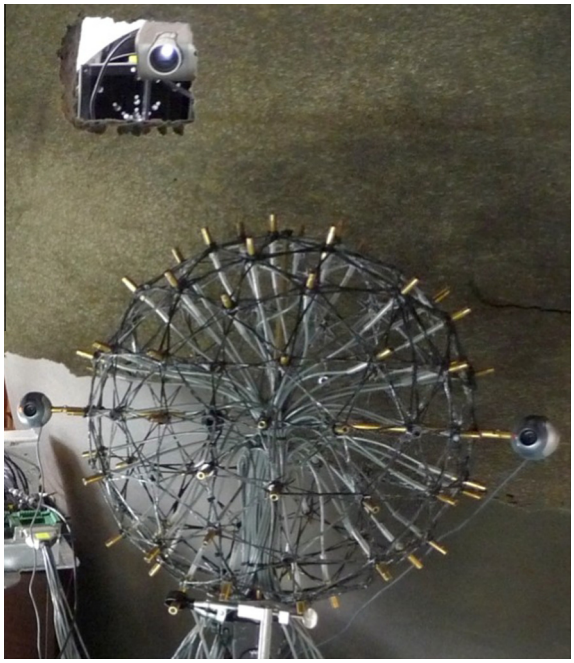

Fig. 2. Structured light experimental setup for a spherical microphone array containing three cameras and a data projector used to obtain 3D surface scans of an object for acoustic imaging.

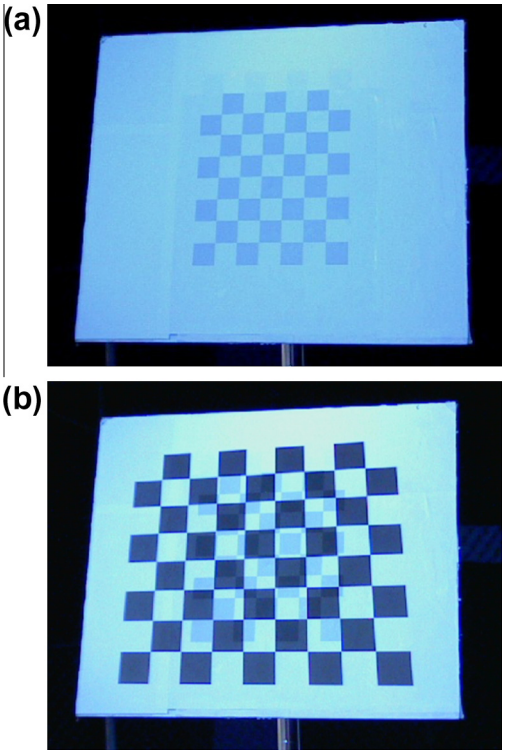

Fig. 3. Calibration rig used for structured light calibration. Image (a) shows the printed checkerboard pattern, which is glued onto its surface. A second checkerboard pattern, which is projected onto the calibration rig, is shown in image (b).

camera lens and include factors such as focal length and lens distortion. The extrinsic parameters describe the position and orientation of a camera in a checkerboard reference frame. This calibration method was also be used for stereo camera calibration. The projected checkerboard parameters were obtained in a similar manner using the images of the projected checkerboard pattern and the camera calibration data.

3.1.2. Microphone coordinate calibration

Acoustic imaging using a microphone array requires that the coordinates of the microphones be known accurately. Ref. [26] details the methodology used for the spherical array to automatically obtain the coordinates of the microphones in the reference frame of one of the cameras in the array. Since the structured light scans were obtained in this same reference frame, this allowed the scan points obtained during the structured light scan to be used as acoustic imaging scan points without the need for any further rotation or translation. The Underbink multi-arm spiral array camera alignment had been achieved using a manual process before the automatic calibration method described above was developed.

3.2. Description of structured light implementation

Binary and gray code structured light scanning software, written by Lanman and Taubin [21], has been used in this work to generate the structured light scans. A sequence of images, consisting of columns and rows of light and dark stripes are projected onto an object. One or more cameras are used to capture images of this sequence on the surface of the object. The deflection of these stripes enables a 3D point cloud to be generated using triangulation. These gray code structured light scans take about 15 s to capture the raw data files. Other structured light scanning techniques, such as color coding, could have been used which allow real time scans to be obtained.

3.2.1. Point cloud processing/merging

The raw point cloud obtained by a structured light scan was processed before being used for acoustic imaging. First, scan points outside the area of interest were removed. Isolated points were then identified using an algorithm that calculated the average distance of each point from its N nearest neighbors. If the average distance was greater than a cut off value, the point was assumed to be noise and was removed from the point cloud. In this example, scan points with a mean distance greater than 10 mm from their 50 nearest neighbors were considered to be noise and were removed. The resulting point clouds can still contain a large number of scan points, which could be computationally expensive if used directly for acoustic imaging. To address this, an option was added that allowed scan points to be removed which were closer than a cutoff value to their nearest neighbor. This cut off distance can be iteratively increased until the desired resolution was achieved.

The scanner can only detect the portion of the object's surface which is illuminated by the data projector and which can be seen by the camera/cameras. To generate a full model of an object, multiple scans may need to be made, with the object at different orientations, and the resulting point clouds merged. To merge two point clouds, a rigid body rotation is used to roughly align the point clouds. The parameters of this transformation may be obtained from knowledge of the extrinsic parameters of the scanner relative to the object's reference frame. Alternatively, common features in the two point clouds may be selected, manually or automatically, and the rigid body transformation parameters calculated. An iterative closest point (ICP) algorithm of overlapping sections may then be used to fine tune the merging of the point clouds.

4. Acoustic imaging using structured light scans

Acoustic imaging was performed using the processed structured light scans. An object was set-up in front of the microphone array containing M microphones. Speakers were positioned against or inserted into the surface of the object. A structured light scan was made and microphone recordings made of white noise played on the speakers. The microphone data were processed to obtain $(M \times M)$ cross spectral matrices $\mathbf{G}$. This is generated by dividing time domain data into J blocks of length K, taking a *FFT* (Fast Fourier Transform) of windowed time domain data in each block, and for each frequency bin calculating

$$\mathbf{G} = \frac{1}{W_m J}\sum_{j=1}^{J}\left[\mathbf{U}(j)\mathbf{U}^{\dagger}(j)\right], \tag{13}$$

where $\mathbf{U}(j)$ is an $(M \times 1)$ element array of *FFT* values, W_m is a windowing function value, and $\dagger$ is the Hermitian transpose. In this work, time domain data blocks of lengths $K = 4096$ samples, an overlap of data blocks of fifty percent, and a Hann window were used [27]. The cross spectral matrices were then summed into twelfth octave frequency bands. These cross spectral matrices were then calibrated using Dougherty's eigenvalue calibration method [20]. Beamforming acoustic map $\mathbf{b}$ was then generated in each twelfth octave frequency band using

$$\boldsymbol{b}(\vec{\xi}_n) = \mathbf{w}^{\dagger}(\vec{\xi}_n)\mathbf{G}\,\mathbf{w}(\vec{\xi}_n), \vec{\xi}_n = \vec{\xi}_1 \ldots \vec{\xi}_N, \tag{14}$$

where the array steering vectors were calculated using

$$w_m(\vec{\xi}_n) = \frac{\dfrac{\exp i\omega\left\|\vec{X}_{m-\vec{\xi}_n}\right\|/c}{\left\|\vec{X}_{m-\vec{\xi}_n}\right\|}}{\sqrt{\sum_{m=1}^{M}\dfrac{1}{\left(\left\|\vec{X}_{m-\vec{\xi}_n}\right\|\right)^2}}}. \tag{15}$$

Note that Eq. (14) is an expanded form of Eq. (3). Deconvolution of these beamforming maps was performed using *CLEAN-SC* [28].

The beamforming resolution of the two arrays used in this work were poor below 5–6 kHz. Also, references [29,30] had shown that, for the spherical array, the errors in 2D method for the near-field increased with frequencies above 5 kHz, while that of the 3D method remained relatively constant to over 15 kHz. Therefore, in this work, only results from one 1/12th octave frequency band, centered at 6.8 kHz, have been presented. This frequency should provide a near best case scenario for the conventional 2D method results presented in Section 5.

4.1. Spherical array

The spherical array structured light scanning system shown in Fig. 2 was used to generate structured light scans of a sheet of plexiglass with six speaker inserted into its surface, see Fig. 4. This was the calibration rig that had been used to automatically calibrate the cameras and obtain the coordinates of the microphones [26]. The plexiglass sheet was set up approximately 0.9 m from the front of the array. This structured light scan was used as to generate acoustic maps, see Fig. 5. The *CLEAN-SC* acoustic map, for the 1/12th octave band centered at 6.8 kHz, gave a mean position error of 6 mm for the speaker locations. This is 1/4 of the diameter of the mini-speakers used. This is consistent with the results presented in Refs. [29,30].

Acoustic images were then generated for a small cardboard box (approx. 220 × 260 × 290 mm). The box was set up approximately one meter from the front of the array. Speakers were positioned against its surface and a structured light scan was made, see Fig. 6. This was then used to generate acoustic maps, see Fig. 7. For this object, it was harder to determine the exact position of

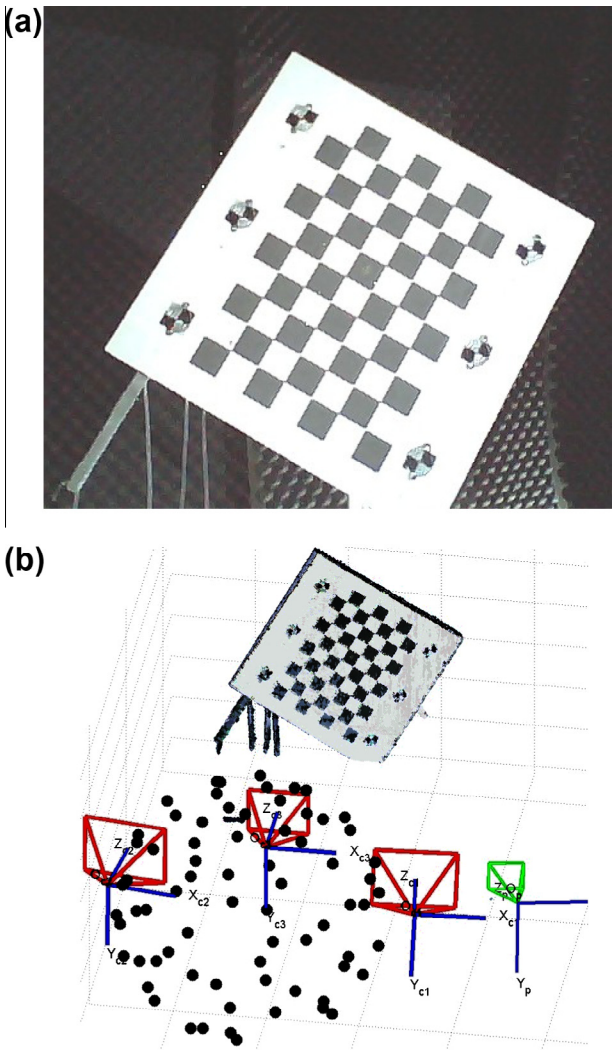

Fig. 4. The camera image in (a) shows the calibration rig which was used as a sound source object for acoustic imaging. It is composed of a piece of plexiglas with six speakers inserted into its surface, surrounding a printed checkerboard pattern. The structured light scan point cloud of this object and the relative positions and orientations of the microphone array, cameras, and projector is plotted in (b).

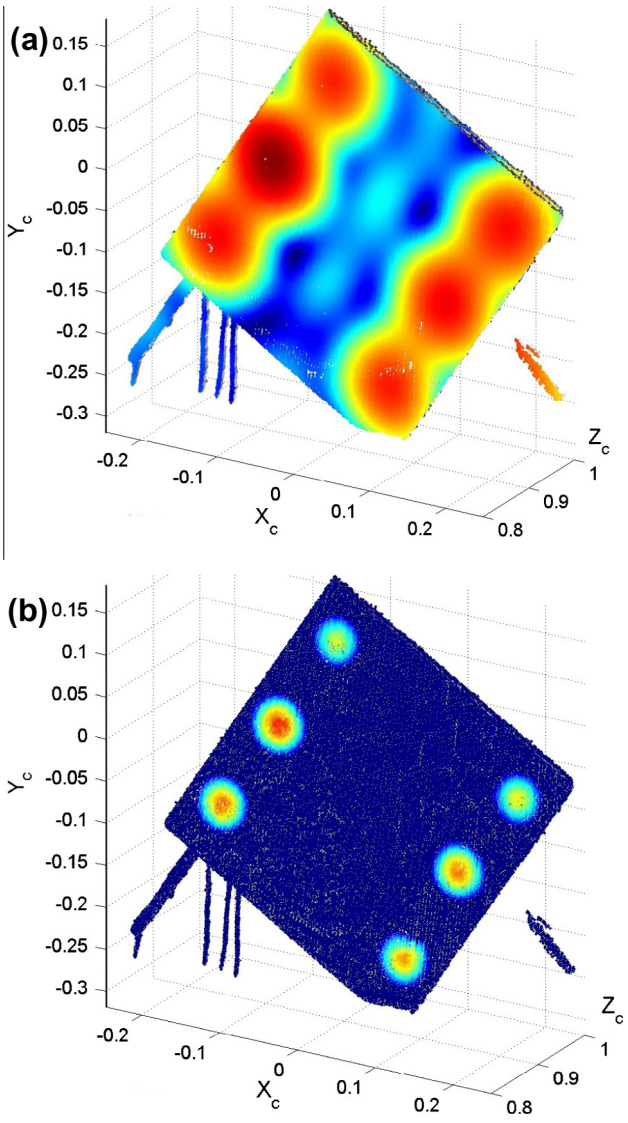

Fig. 5. Graphs (a) and (b) respectively show acoustic beamforming and *CLEAN-SC* maps generated using the structured light scan shown in Fig. 4. A mean position error of 6 mm was obtained.

the center of the speakers. Excluding the speaker which was placed behind the box, a mean speaker position error of 12 mm was obtained for the 1/12th octave band centered at 6.8 kHz. Some additional errors were expected to be introduced by the box causing scattering of the sound propagating to the microphones. To test this, an additional set of "free-air" microphone recordings was made, where the cube was removed but the speakers were left undisturbed. Using the same 3D scanning surface as before, a mean position error of 6 mm was now obtained with these "free-air" recordings.

4.2. Underbrink multi-arm spiral array

Acoustic imaging using structured light scans was also tested using the Underbrink multi-arm spiral array. The setup is shown in Fig. 8. The object being acoustically imaged was a board with four speakers located behind holes drilled into the board. The board was positioned at an angle to the microphone array. The coordinates of the speakers, relative to the microphone array, are given in Table 1.

A structured light scan was then made of the object, see Fig. 8(b). The acoustic maps were then generated using these structured light scans, see Fig. 9. A mean speaker position error of 6 mm was obtained for *CLEAN-SC* for the 1/12th octave band centered at 6.8 kHz. This mean error is half the diameter of the holes in the board through which the sound propagated.

5. A comparison of the 3D method with the conventional 2D scanning surface method

A comparison of the 3D method with the conventional (2D scanning surface) method was then performed. This comparison was achieved by using the same microphone data that was used in Section 4.2. However, rather then using a 3D scanning surface, four 2D scanning surfaces were used which were parallel to the array. Each planar scanning surface was positioned so that it passed through one of the speakers, see Fig. 10 for one example. This meant that, for each plane, three of the speakers where offset from the scanning surface (conventional 2D method) while one of the speakers was not offset (equivalent to the 3D method used in this

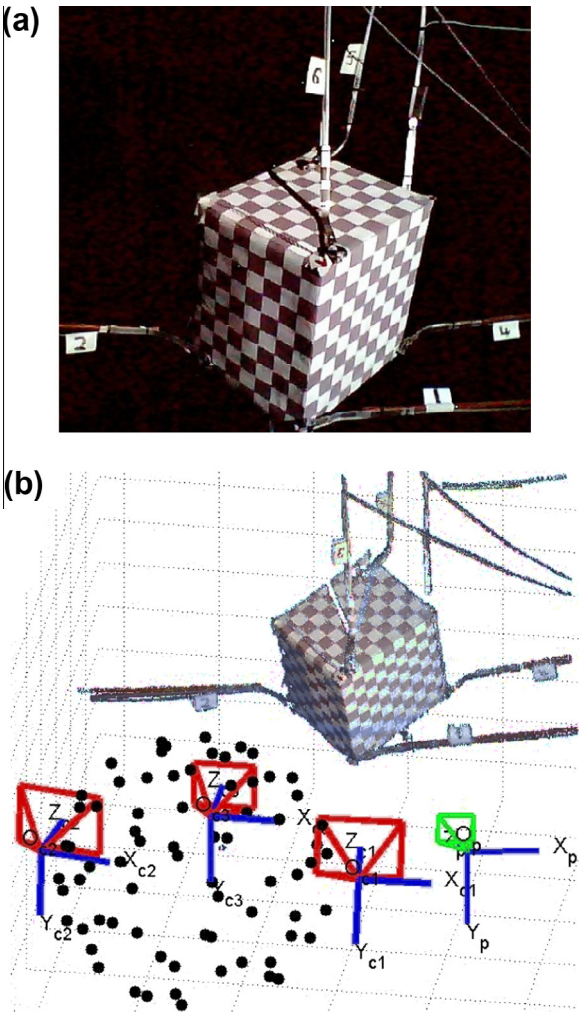

Fig. 6. Photo (a) of a cardboard box with six speakers positioned against its surface. The structured light scan of this object and the relative positions and orientations of the microphone array, cameras, and projector are plotted in (b).

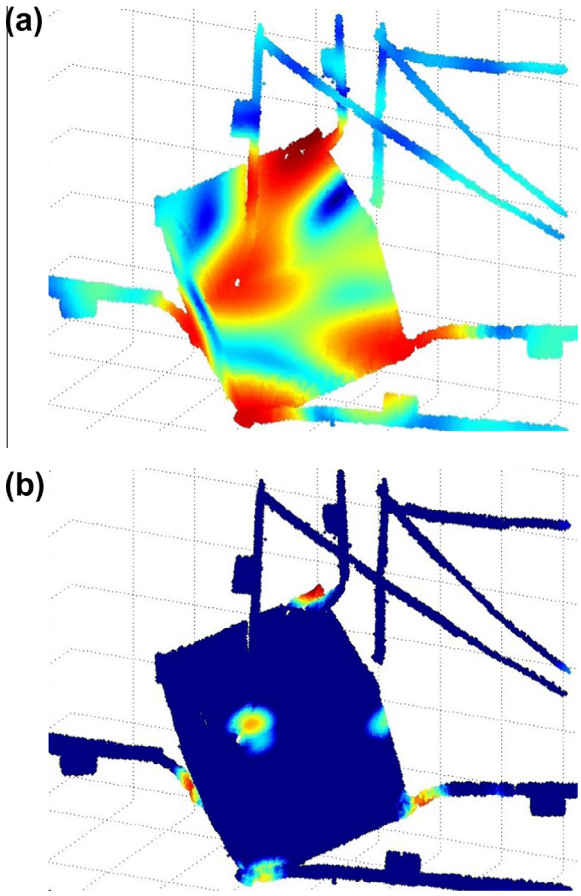

Fig. 7. Graphs (a) and (b) respectively show acoustic beamforming and *CLEAN-SC* maps generated using the structured light scan of the cardboard box with speakers on its surface shown in Fig. 6.

work). This methodology enabled a comparison to be made of the accuracy of the conventional 2D method with that of the 3D method.

Fig. 10 shows the experimental 2D acoustic map obtained for the scanning surface closest to the array. Table 2(a) and (b) respectively give the position and sound pressure levels errors for each speaker and each of the four planes. The diagonal elements correspond the 3D method, while the off-diagonal elements correspond to the 2D method with different amounts of offsets. The mean position error for the diagonal element (3D method) was 6.6 mm. However, for the off-diagonal elements (2D method), the mean position error was 57 mm and the mean reduction in sound pressure level was −4 dB. As the offset of the 2D scanning surface from the speaker increased, the error in the position and magnitude of the sound sources increased.

This can be explained in terms of increased beamforming phase error with increased offset of the scanning surface from the sound source location. Using perspective projection, one may use Eq. (10) to calculate a theoretical scan point where a peak in the 2D acoustic map should occur due to a sound source located at $\vec{X}_s$. Table 3(a) gives the offsets of the four speakers from the corresponding points $\vec{\xi}_p$ in the four sets of 2D scanning surfaces. Table 3(b) shows the resulting mean theoretical beamforming phase error that results from these offsets calculated using Eq. 9. The near zeros phase errors for the diagonal (3D) elements indicates that constructive interference was occurring during beamforming for the 3D case. However, the phase errors of nearly π for the off-diagonal (2D) elements which had the most offsets indicates that a high degree of destructive interference was occurring during beamforming for these (2D) elements.

6. Possible extensions

To obtain realtime scans of moving objects, techniques such as color coding structured light scanning could be used. Alternatively, new 3D scanning technology which has become available since this work was started could be used. This includes the *Xbox Kinect*[1], the *Leap Motion*[2], and a low cost time of flight camera from Intel[3]. The ability to obtain real-time 3D surface scans of the surface of an object is becoming cheaper and more easily available. Future work would incorporate sensors such as these into a microphone array for real-time 3D surface acquisition for acoustic imaging.

A draw-back of using a standard structured light system is that the scan has a limited field of view. For imaging enclosures, such as the interior of a car, one might wish to image the full interior. Multiple camera/projector pairs which cover the entire field of interest could be used or an omnidirectional system could be investigated.

[1] http://www.microsoft.com/en-us/kinectforwindows/.
[2] https://www.leapmotion.com/.
[3] http://click.intel.com/intelsdk/Default.aspx.

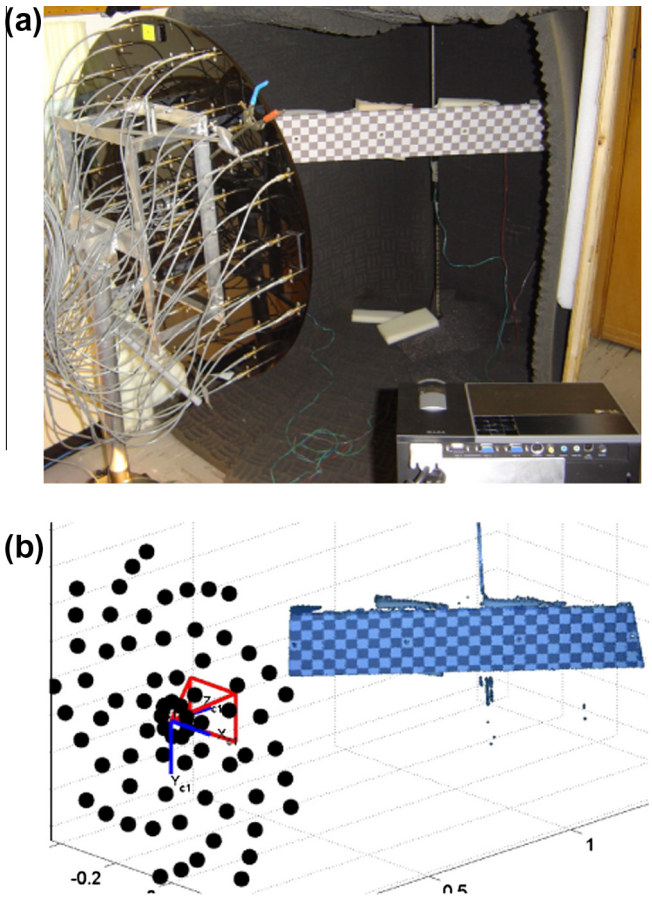

Fig. 8. Photo (a) showing the experimental setup combining an Underbrink multi-arm spiral array with a structured light scanning system. The object being scanned is a board with four speakers positioned behind holes. The resulting structured light scan, shown in (b), is used for acoustic imaging.

Table 1
Speaker coordinates relative to the microphone array for the setup shown in Fig. 8.

Speaker no.	x (m)	y (m)	z (m)
1	−0.353	0.081	0.780
2	−0.159	0.077	0.987
3	0.032	0.073	1.199
4	0.219	0.070	1.414

A single projector/camera scanner system could be attached to a robotic device allowing it to be rotated using software. A more ideal scenario, however, might be to use a computer vision code that allows the array, and hence the structured light scanner, to be moved about an object or enclosure and a 3D surface could then be built up as the array is moved [31].

Microphone arrays are commonly used for wind tunnel aeroacoustic testing. The technique presented in this paper may provide added value in these measurements, as wind tunnel models can deform in the wind (for example, the wings of aircraft scale models). However, the aeroacoustic noise may not propagate from the surface of the object but instead propagate from some point in the wake of the object. Just acoustic imaging over the surface of the object might, therefore, miss these aeroacoustic sources. A solution to this could be to include a grid of scan points surrounding the surface geometry obtained using the structured light scan. This should allow these aeroacoustic sources to be imaged.

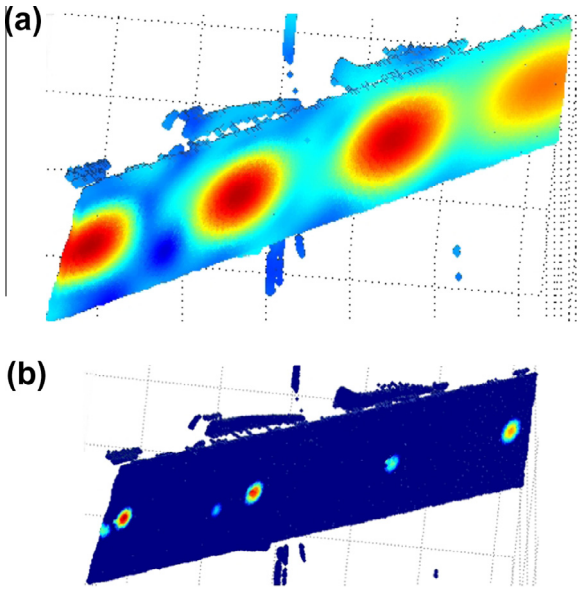

Fig. 9. Beamforming (a) and *CLEAN-SC* (b) 3D acoustic maps for the Underbrink multi-arm spiral array generated using structured light scan points. The mean position error is 6 mm.

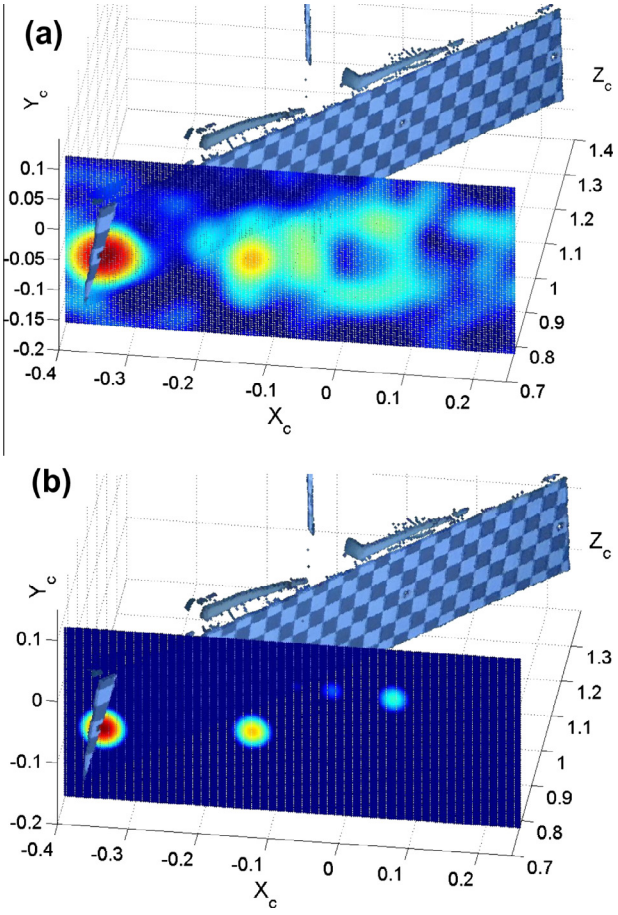

Fig. 10. Beamforming (a) and *CLEAN-SC* (b) 2D acoustic maps generated using traditional 2D scanning surfaces for the Underbrink multi-arm spiral array. To demonstrate the position and orientation of the 2D scanning surface relative to the object being scanned, the structured light point cloud has also been included.

Table 2
Table (a) shows the experimentally measured *CLEAN-SC* position errors in mm for four speakers and four sets of 2D planes. The corresponding reduction in the *CLEAN-SC* sound pressure levels in dB is given in Table (b) relative to that obtained using the 3D structured red light scanning surface technique. The diagonal (bold) components correspond to the 3D case with minimal offset, while the off-diagonal elements correspond to conventional 2D planes with different amounts of offset from the speakers.

Speaker no.	1	2	3	4
(a)				
Plane 1	**9.5**	10	110	120
Plane 2	38	**2**	8.5	7
Plane 3	94	6	**6.5**	2.5
Plane 4	210	75	5	**8.5**
(b)				
Plane 1	**0**	−4.4	−10.8	−9.0
Plane 2	−2.8	**0**	−5.0	−9.1
Plane 3	−4.4	−0.4	**0**	−2.5
Plane 4	−2.1	−2.0	0.5	**0**

Table 3
Table (a) gives the offset in mm of four sound sources relative to four 2D planes. Table (b) gives the corresponding mean theoretical beamforming phase errors. The diagonal (bold) components correspond to the 3D case with minimal offset, while the off-diagonal elements correspond to conventional 2D planes with different amounts of offset from the speakers.

Speaker no.	1	2	3	4
(a)				
Plane 1	**2.76**	209	418	641
Plane 2	213	**1.50**	213	433
Plane 3	424	215	**1.08**	218
Plane 4	639	434	215	**1.45**
(b)				
Plane 1	**0.01**π	0.35π	0.60π	0.77π
Plane 2	0.38π	**0.01**π	0.25π	0.42π
Plane 3	0.90π	0.24π	**0.01**π	0.18π
Plane 4	1.32π	0.42π	0.18π	**0.01**π

7. Conclusion

A new technique has been presented which automatically generates the 3D surface geometry of an object for acoustic imaging using structured light scanning. The technique uses one or more cameras in the microphone array and a data projector to form a structured light scanning system. The structured light scans are then used as a scanning surface for 3D acoustic imaging using the microphone data. The advantages of this method are that it provides a relatively simple add onto most existing microphone arrays, is robust, inexpensive, and fast. Extensions are suggested which would allow the method to be extended to 3D acoustic imaging of dynamic scenes. This technique was tested with a spherical array and an Underbrink multi-arm spiral array. A mean position accuracy was obtained for *CLEAN-SC* of 6 mm, which is a quarter of the diameter of the speakers used. A comparison between the accuracy of *CLEAN-SC* for these 3D scanning surfaces and traditional 2D scanning surfaces was made for the spiral array. The measured position and magnitude of sound sources obtained by *CLEAN-SC* using the 2D scanning surfaces was shown to provide an increase in the error as the offset of the sound source location from the 2D scanning plane increased. The position error was as large as an order of magnitude bigger than that obtained using the 3D method.

References

[1] Humphreys Jr WM, Brooks TF, Hunter WW, Meadows KR. Design and use of microphone directional arrays for aeroacoustics measurements, AIAA-1998-0471; 1998.

[2] Dougherty RP. Advanced time-domain beamforming techniques, AIAA-2004-2955, In: 10th AIAA/CEAS aeroacoustics conference, Manchester, Great Britain, May 10–12, 2004; 2004.

[3] Brooks TF, Humphreys WM. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. J Sound Vib 2006;294(4–5):856–79.

[4] Sijtsma P. CLEAN based on spatial source coherence. Int J Aeroacoust 2007;6(4):357–74.

[5] Maynard JD, Williams EG. Nearfield acoustic holography: theory of the generalized holography and the development of NAH. J Acoust Soc Am 1985;78(4):1395–413.

[6] Schmitt A, Lamotte L. Source identification inside cabin using inverse methods, BeBeC-2010-05. In: Proceedings on CD of the 3rd Berlin beamforming conference, 24–25 February, 2010; 2010. <http://bebec.eu/Downloads/BeBeC2010/Papers/BeBeC-2010-05.pdf>.

[7] Dougherty RP. Extensions of DAMAS and benefits and limitations of deconvolution in beamforming. In: 11th AIAA/CEAS aeroacoustics conference (26th AIAA aeroacoustics conference), Monterey, California, USA; 2005.

[8] Döbler D, Heilmann G, Schröder R. Investigation of the depth of field in acoustic maps and its relation between focal distance and array design. In: Inter noise, Shanghai, China; 2008.

[9] Maffei M, Bianco A. Improvements of the beamforming technique in Pininfarina full scale wind tunnel by using a 3D scanning system. SAE Int J Mater Manuf 2008;1(1):154–68.

[10] Irimia C, Deblauwe F, Jannens K, Juhos Z, Ignat S. Improving the localization of noise sources inside a vehicle, vol. B. Scientific Bulletin: Automotive Series, Year XV; 2009.

[11] Brooks TF, Humphreys WMJ. Three-dimensional application of DAMAS methodology for aeroacoustic noise source definition. In: 11th AIAA/CEAS aeroacoustics conference, Monterey, California, USA; 2005.

[12] Ravetta PA. LORE approach for phased array measurements and noise control of landing gears. Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA; 2005.

[13] Xenaki A, Jacobsen F, Tiana-Roig E, Grande EF. Improving the resolution of beamforming measurements on wind turbines. In: Proceedings of 20th international congress on acoustics (ICA), Sydney, Australia; 2010.

[14] Dougherty RP. Jet noise beamforming with several techniques. In: 3rd Berlin beamforming conference, Berlin, Germany; 2010.

[15] Meyer A, Döbler D. Noise source localization within a car interior using 3D-microphone arrays. In: 1st Berlin beamforming conference, Berlin, Germany; 2006.

[16] Döbler D, Heilmann G, Meyer A, Navvab M. Fields of application for three-dimensional microphone arrays for room acoustic purposes, 4th ed. Acoustic Camera – Newsletter; 2011.

[17] Upton R, Haddad K, Sorensen J. Conformal mapping techniques for consumer products. Sound Vib 2008;42(7):8–11.

[18] Geng J. Structured-light 3D surface imaging: a tutorial. Adv Opt Photon 2011;3(2):128–60.

[19] Li Z, Curless B, Seitz SM. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In: Proceedings of the first international symposium on 3D data processing visualization and transmission, Padova, Italy; 2002. p. 24–36.

[20] Dougherty RP. Beamforming in acoustic testing. In: Mueller TJ, editor. Aeroacoustic testing. Berlin: Springer-Verlag; 2002. p. 63–97.

[21] Lanman D, Taubin G. Build your own 3D scanner: optical triangulation for beginners; 2009. <http://mesh.brown.edu/byo3d/> [accessed March 2012].

[22] Falcao G, Hurtos N, Massich J, Fofi D. Projector-camera calibration toolbox; 2009. <http://code.google.com/p/procamcalib> [accessed March 2012].

[23] Bouguet J-Y. Camera calibration toolbox for MATLAB; 2008. <http://www.vision.caltech.edu/bouguetj/calib_doc/index.html> [accessed March 2012].

[24] Zhang Z. A flexible new technique for camera calibration. Tech. Rep., Microsoft Research, Microsoft Corporation; 1998.

[25] Bradski G, Kaehler A. Learning OpenCV: computer vision with the OpenCV Library, O'Reilly Media; 2008.

[26] Legg M, Bradley S. A combined microphone and camera calibration technique with application to acoustic imaging. In: IEE transactions on image processing; 2013. http://dx.doi.org/10.1109/TIP.2013.2268974. ISSN:1057-7149. .

[27] Brooks TF, Humphreys WMJ. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. In: AIAA-2004-2954, 10th AIAA/CEAS aeroacoustics conference, Manchester, UK; 2004.

[28] Sijtsma P. CLEAN based on spatial source coherence. In: 13th AIAA/CEAS aeroacoustics conference, Rome, Italy; 2007.

[29] Legg M, Bradley S. Comparison of CLEAN-SC for 2D and 3D scanning surfaces. In: 4th Berlin beamforming conference, Berlin, Germany; 2012.

[30] Legg M. Microphone phased array 3D beamforming and deconvolution. Ph.D thesis, University of Auckland, Auckland, New Zealand; 2012.

[31] Robbel P. The kinect sensor in mobile robotics: initial experiments; 2012. <http://www.youtube.com/watch?v=dRPEns8MS2o&feature=player_embedded#at=71Personal> [accessed March 2012].