ENHANCING MANUAL SCAN REGISTRATION USING AUDIO CUES

T. Ntsoko[a], G. Sithole[b]

University of Cape Town, Engineering and the Built Environment Faculty, Geomatics Division, 7700 Rondebosch Cape Town,
South Africa - [a]ntstha019@myuct.ac.za, [b]george.sithole@uct.ac.za

KEY WORDS: Point Clouds, Augmentation, Audio, Interaction, Processing, Visualisation

ABSTRACT:

Indoor mapping and modelling requires that acquired data be processed by editing, fusing, formatting the data, amongst other operations. Currently the manual interaction the user has with the point cloud (data) while processing it is visual. Visual interaction does have limitations, however. One way of dealing with these limitations is to augment audio in point cloud processing. Audio augmentation entails associating points of interest in the point cloud with audio objects. In coarse scan registration, reverberation, intensity and frequency audio cues were exploited to help the user estimate depth and occupancy of space of points of interest. Depth estimations were made reliably well when intensity and frequency were both used as depth cues. Coarse changes of depth could be estimated in this manner. The depth between surfaces can therefore be estimated with the aid of the audio objects. Sound reflections of an audio object provided reliable information of the object surroundings in some instances. For a point/area of interest in the point cloud, these reflections can be used to determine the unseen events around that point/area of interest. Other processing techniques could benefit from this while other information is estimated using other audio cues like binaural cues and Head Related Transfer Functions. These other cues could be used in position estimations of audio objects to aid in problems such as indoor navigation problems.

1. INTRODUCTION

Acquired point clouds for indoor mapping and modelling often need to be processed. The common processing tasks include registration, cleaning and simplification, amongst others for successful indoor modelling and mapping of environments. Due to the sizes of point clouds, processing is usually automatic. However, manual or semi-automatic interventions will often be required to refine results from automatic processing. Currently, the interaction the user has with the point cloud while processing is visual. However, visual interaction has limitations. Such limitations exist where the data the user needs to interact with is out of the user's field of view.

One way to deal with this limitation is to augment audio in point cloud processing. This augmentation entails associating points of interest in point clouds with audio objects. The objective of this study is to investigate the augmentation of visual point cloud processing with audio and the limitations associated with this augmentation. Specifically, audio augmentation in coarse scan registration is investigated to demonstrate the idea.

Audio augmentation could also be used to solve problems that exist in indoor mapping and modelling. These include navigation of modelled indoor scenes, enhanced perception of the modelled environments and determination of the occupancy of space of the modelled indoor scenes, amongst others (Zlatanova et al., 2013). The use of audio cues by estimating the position, depth and occupancy of space of an audio object could help in addressing these problems. This can be done by associating points of interest in indoor data with audio objects and using audio cues to estimate relevant spatial information.

The investigation begins by looking at the coarse scan registration process and observing instances where the process is hindered by limitations arising due to visual cues. Augmenting audio in those instances will then be suggested.

The implementation of augmenting audio in point clouds is then discussed. The limitations of this implementation and their implications to coarse scan registration will also be discussed. Following this, the benefits of augmenting audio in coarse scan registration and other point cloud processing techniques will be discussed.

2. AUDIO AUGMENTATION IN COARSE SCAN REGISTRATION

Audio augmentation entails associating objects of interest in the point cloud with audio objects. Using audio cues, spatial information such as the depth and occupancy of space of the audio object can be estimated. This will in turn potentially allow the user to retrieve spatial information of objects of interest in the point cloud which are associated with audio objects. Audio augmented coarse scan registration is discussed here.

2.1 Coarse Scan Registration

Zlatanova et al. (2013, pg. 64) noted that "various data are registered or fused to deliver an amalgamated scene of an indoor environment." The scans captured are fused together by rotating and translating the target (misaligned) scan with respect to the reference scan. Coarse registration entails roughly aligning multiple point clouds in preparation for an automatic fine alignment. The purpose of coarse alignment is to facilitate a fast convergence of automatic fine alignment algorithms.

The target and the reference scans are related by a rigid Euclidean transformation given by equation 1 (Brenner et al., 2007).

$$\vec{x_1} = R\vec{x_2} + \vec{t} \tag{1}$$

Using this transformation, misaligned scans can be coarsely aligned. However, coarse registration is generally a poor exercise (Xie et al., 2010). The following text will propose how this can be improved when audio is augmented in the process.

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-4-187-2014

2.2 Coarse Scan Registration with Audio Augmentation

Audio augmentation in scan registration is explained with the aid of figure 1.

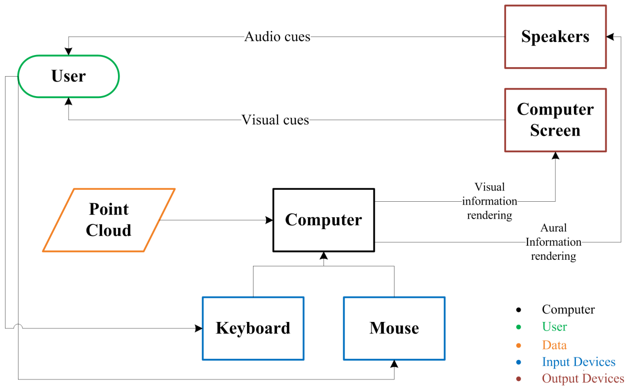

Figure 1: Audio augmentation in coarse scan registration.

The *user* controls the manual coarse scan registration task. This allows the *user* to interact with the *computer* through the *input devices*. This interaction allows the *user* to receive processing output through the *output devices*.

Using the *computer*, an **audio context** is created. This entails connecting with the *computer's* audio device and opening it to render audio. The objects (e.g., point clouds) are rendered and displayed on the screen. Simultaneously the audio cues are also rendered and presented to the user through speakers. The audio cues generated should complement the rendered imagery so that the *user* experience is enhanced and seamless.

Rendering the audio cues will require the objects be treated as emitters of sound. Audio objects emanate audio as required by the *user*. The emanated audio is received by the *user*. The next critical event in the *computer* is **audio augmentation** which entails 'attaching' an audio object to an area of interest in the *point cloud*.

Input devices (the **keyboard** and the **mouse**) allow the *user* to manually transform the target scan. The manipulation commands are sent directly to the *computer*. Interactions with *input devices* begin a feedback loop in which visual and audio renderings are updated by the *computer* and presented by the computer screen and speakers until interaction with the *input devices* ceases.

The *user* receives information of target scan transformation via *output devices* (**computer screen** and the **speakers**). The computer screen provides information to the user in the form of **visual cues**, whereas the speakers provide it in the form of **audio cues**. In general, speakers refer to any audio output device. Headphones were used in this work. Providing information through these devices is triggered by the *input devices* when they send scan transformation commands to the *computer*, which in turn communicates this manipulation through *output devices*.

Audio cues can potentially provide alignment information to the *user* that the visual cues are not able to communicate. Consider a scanned building shown in figure 2, where two separate scans (blue and red) were done from separate perspectives. To align these two scans the transformation process begins with the rotation, where the coordinate system of the misaligned scan is aligned with that of the reference scan.

2.2.1 Coarse Rotation with Audio Augmentation:

The rotation is performed using the rotation matrix R of equation 1. The right scan is rotated to align with the left scan. The resulting rotation aligns the scans.

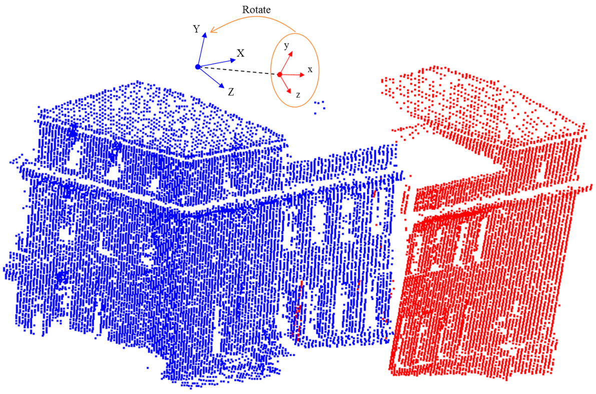

Figure 2: Scanned building with separate scans with the misaligned scan needing to be rotated.

The orientation of the misaligned scan can be tied to the orientation of an audio object attached to it. As a result, as the scan is rotated in space, the orientation of the audio object will change as well. This will provide aural information about the scan's orientation. This link is expressed in equation 2,

$$R_{object}(R_x, R_y, R_z) = \alpha, \beta, \kappa \tag{2}$$

where, R_x, R_y and R_z are the orientations of the scan in X, Y and Z-axes with respect to the reference scan. α, β and κ are the orientations of the audio object in the reference system in units of degrees, giving the orientation of the audio object R_{object}. Associating the magnitude by which the misaligned scan needs to be rotated with audio intensity, the user could use this intensity to estimate if the proper rotation has been done. With reference to figure 2, audio augmentation can potentially lead to the red scan being rotated to align with the blue scan.

2.2.2 Coarse Translation with Audio Augmentation:

Manual scan translation is done post rotation. The aim of the translation is to minimise the separation (depth) between surfaces. In figure 3 the two surfaces are shown separated by depth d. Using coordinates of a common target scanned when each of the scans were attained, this depth can be determined using equation 3. In this equation, x_1, y_1 and z_1 are the coordinates of target T in the blue scan and x_2, y_2 and z_2 are the coordinates of the same target (T$^{'}$) in the red scan.

$$d \quad = \quad \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{3}$$

Equation 4 shows one possible way of associating depth with sound intensity.

$$\begin{aligned} \phi(d) &= M - F \times V \\ &= M - F \times 20log_{10}(\frac{s_d}{s_0}) \end{aligned} \tag{4}$$

In equation 4, V is the loss of intensity in decibels (dB) determined by the initial audio object distance s_0 and the current distance s_d from the listener. V is the commonly used intensity attenuation model in 3D audio simulation applications. M and

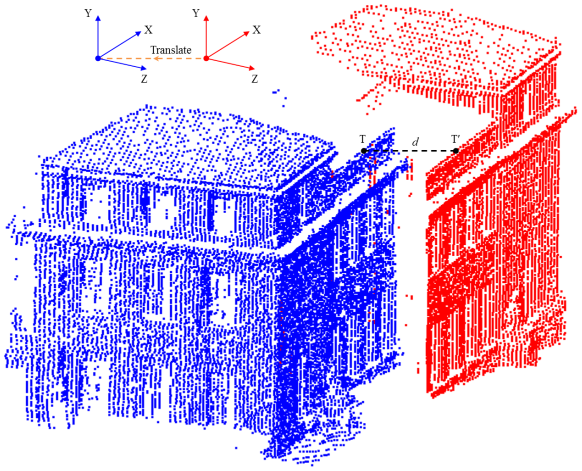

Figure 3: Scanned building with separate scans with the misaligned scan needing to be translated through depth d.

F are user-defined values, where M is the desired initial audio intensity in decibels and F is a factor which controls the rate at which the intensity drops with depth. $\phi(d)$ is the sound intensity as a function of depth d. Figure 4 demonstrates how F affects the change in intensity as the depth changes. In this case, M was chosen to be 35 dB.

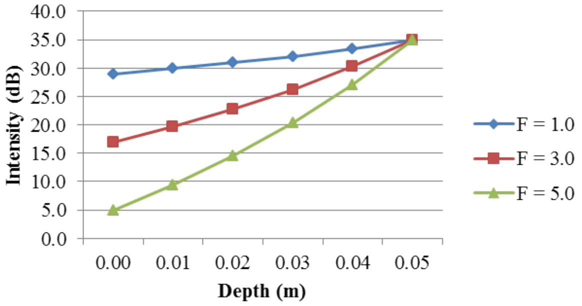

Figure 4: Intensity depth plot where the intensity of the audio object is tied to the depth between scans and the attenuation depends on factor F.

Reverberation of an audio object can be used in situations where the *user*, immersed in the point cloud attempts to move a surface in relation to another. In this situation, the *user* might need to know how the surfaces are moving in relation to each other in areas which are out of view. This is illustrated in figure 5. The *user* is immersed in the point cloud and is only able to view points inside the view frustum.

In the figure, red lines represent a misaligned scan/surface, while blue lines represent an aligned one or reference scan. While the *user* attempts to move surface B to align with surface A, there could be undesirable movement elsewhere. Or, the *user* might only be interested in the rate at which these surfaces are converging towards each other. A reverberant audio object could aid in alerting the *user* of this.

The reverberations of an audio object attached to surface A, could be tied to the average distance between the points of the two surfaces. This average distance can be determined using equation 5, where,

$$D = \frac{\sum_{k=1}^{n} d_k}{n} \tag{5}$$

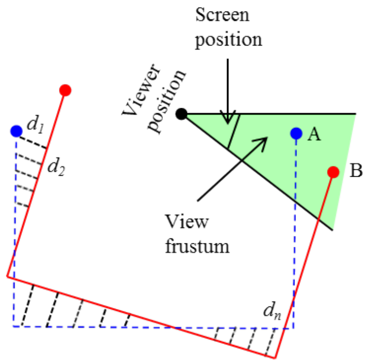

Figure 5: User moves surface B towards surface A while immersed in the point cloud.

D is the average distance, d_k is the k^{th} shortest distance between points of surface A and B and n is the number these calculated distances.

The reverberation of the audio object attached to surface A can then be a function of average distance D. This relationship could be defined using equation 6, where, τ is the total reverberation that will be experienced, measured in decibels (dB).

$$\tau(D) = R \tag{6}$$

2.3 Discussion

The stated examples of how audio could be augmented to enhance scan registration are not exhaustive. Depending on the user's preference, different audio cues could be used to enhance different aspects of scan registration. For example, instead of using audio intensity to help the user appreciate the depth between scans, audio pitch variations could be used.

Other point cloud processing techniques could benefit from audio augmentation. For example, in cleaning of scans, associating sound with the separation of points from their parent surfaces could aid in identifying outliers. Here, outlying points will have distinctly different sounds from those points that lie on a surface and can therefore be removed. In point cloud simplification tasks, the user might need to seek out places where there has been over or under simplification. This process can be enhanced by associating sound with the function of the local surface curvature and the point density.

The limitations of audio augmentation in scan registration will be tested. This should highlight the possibilities of augmenting audio to enhance manual coarse scan registration.

3. REVIEW OF PRINCIPLES OF AUDIO

Kapralos et al. (2003, pg. 1) observed that "hearing can serve to guide the visual attention and therefore eases the burden off the visual system." The user can therefore take advantage of this in doing depth and occupancy of space estimations. These are done with respect to a coordinate system whose origin is at the centre of the listener's head.

Depth estimation is illustrated in figure 6. This figure illustrates the idea of a listener estimating the depth from which the sound source (red sphere in the figure) is located. For example, in this figure, the correct estimation of the source's depth would be X m.

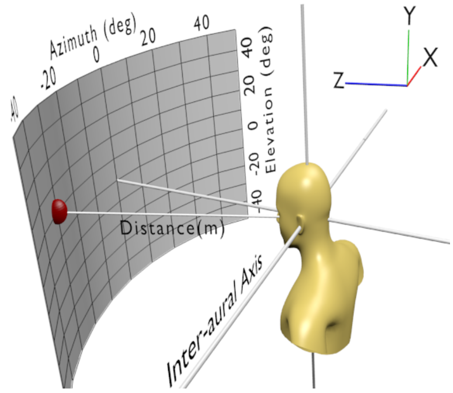

Figure 6: Depth (in meters) Estimation of an audio object. The red sphere represents the sound source whose spatial information is to be extracted.

The auditory system uses various audio cues to discern spatial information. The auditory cues that will be discussed in this study are, reverberation, intensity and frequency/pitch. Audio cues are not only limited to these ones. In another study where position estimations were done, the binaural cues and Head Related Transfer Function (HRTF) audio cues were studied and used.

3.1 Reverberation

Emitted sound gets reflected off objects around the sound source, leading to the reverberation sensation (Kapralos et al., 2003). The manner in which the reflections occur depends on the objects surrounding the sound source (Begault, 1994). This could inform the listener about the nature of the surroundings of the sound source.

Using reverberation, a sound object associated with an object of interest in the point cloud gives the user the opportunity to estimate the surroundings of that object. The occupancy of space could therefore potentially be determined in this manner.

3.2 Intensity

According to Zahorik et al. (2005) and Väljamäe (2005), intensity of an audio source can be used effectively as a depth cue. As intensity gets attenuated with, distance, this cue can also alert the listener of changes of depth. Kapralos et al. (2003) provided the following model to show the attenuation of audio intensity with depth:

$$L_{loss} = 20 \times log_{10}(\frac{s_d}{s_0}) \tag{7}$$

where, L_{loss} is the loss in intensity measured in decibels (dB), s_0 is the initial audio source depth and s_d is the current depth between the listener and the audio source. This model given by equation 7, follows the inverse square law of audio intensity attenuation. For every doubling of depth of the sound source from the listener, a 6 dB loss in source intensity is experienced (Begault, 1994; Mershon and King, 1975; Shinn-Cunningham, 2000; Zahorik et al., 2005). Using audio intensity, the depth of an object in the point cloud associated with an audio object can potentially be estimated and this will be explored.

3.3 Frequency

Frequency of audio can be a useful depth cue (Handel, 1989; Kapralos et al., 2003). Kapralos et al. (2003) furthermore noted that spectral changes can provide relative depth information, unless the listener has prior knowledge of the source, in which case absolute depth information can be provided.

Greater attenuation is experienced for higher frequency components as the depth between the source and the listener increases (Handel, 1989; Kapralos et al., 2003). The complexity of spectral changes of sound, make this notion of frequency being used as depth cues quiet contradictory, especially in auditory depth simulations (Handel, 1989). Handel (1989) and Kapralos et al. (2003) both agree that spectral changes can be used as depth cues, particularly for high frequency sounds, but they emphasise that familiarity with the sound can be very useful and lead to better auditory depth estimations.

Changes in the frequency of an audio object could provide depth information of objects in point clouds for processing. The reviewed literature suggests that audio frequency does change with depth, as already observed with audio intensity. This cue is worth exploiting in point cloud processing where depths of objects are required.

4. AUDITORY INTERFACE IMPLEMENTATION

The auditory interface was created by augmenting audio into a point cloud. An auditory interface is an environment that has sound sources emitting audio, with a listener object/virtual listener created to receive this audio. OpenAL (Open Audio Library) was used for this.

4.1 OpenAL – Open Audio Library

Wozniewski and Settel (2007, pg. 1) observed that OpenAL is “purely concerned with the spatialisation of sounds located in the scene, as a result the audio experience is focused around one user who indeed is immersed in 3D sound.” OpenAL does this through creation of a listener object and sound source objects. The listener and audio object(s) are placed according to the OpenAL coordinate system.

OpenAL Coordinate System: OpenAL uses a right-handed coordinate system to define the spatial attributes of the listener and sound source objects. In a default frontal view, the X-axis points to the right, the Y-axis points up and the Z-axis points towards the viewer (Creative-Labs, 2010). In figure 7, a listener object is shown with one sound source object in the scene.

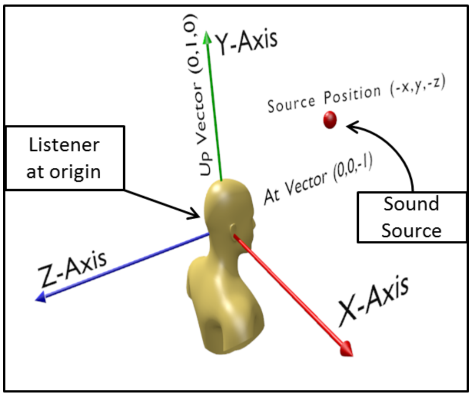

Figure 7: A virtual listener and a sound source are shown here. The listener’s orientation is such that its *up vector* is (0,1,0) and its *at vector* is (0,0,-1).

The Listener Object: The OpenAL listener has a *3D position* which can be set. This *position*, combined with the *position* of the sound source(s), influences the aural experience the listener receives when audio is emitted. The listener has the *master gain* attribute which controls the loudness with which the listener experiences emitted audio. Lastly, the listener has an *orientation vector* to define its orientation in 3D space. The *orientation vector* is split into two vectors, each with three elements: the *up vector* and the *at vector*. The *up vector* defines which way *up* the listener is directed. The *at vector* defines the direction the listener is looking *at*. The virtual listener is synced with the user through an audio output device. In this work, headphones were chosen as the most appropriate manner for the user to receive audio. (See figure 8.)

The Sound Source Object: The sound source object is the source of the emitted audio and received by the virtual listener. The *position* of a sound source can be set in 3D space. This influences the aural experience of a virtual listener when sound is emitted. The *pitch multiplier* attribute allows the user to change the *pitch*. The *source gain* helps in adjusting the *gain/volume*. A *source gain* of zero means that a source will not be heard at all when audio is emitted. The *source gain* has a bearing on the intensity with which a virtual listener experiences the emitted sound. The *direction vector* of a sound source sets the *direction* which a sound source is facing. This has a bearing on how well the listener will hear the audio. The audio data to be emitted by a sound source is stored in a buffer object.

4.2 Creating an Auditory Interface

An auditory interface was created for the point cloud. The objective here was to turn objects (clusters of points) in point clouds into sound sources and to have a listener object that will listen to the emitted sound. The point cloud was partitioned using an octree. The purpose of this partitioning was to allow for unconnected objects represented in the point cloud to be treated as separate objects. These separate objects are taken as nodes whose purpose will be explained later. The point cloud was partitioned using the *by node width* method.

With the point cloud partitioned, the objects/nodes in the octree can now be treated as sound sources. Sound sources can be assigned to the centroids of objects. The virtual listener was placed at the origin of the OpenAL coordinate system.

4.3 Experimental Set-up

The experimental set-up adopted here will be explained using figure 8. In this figure, the following are shown:

1. The tester wearing headphones (output device).
2. The computer screen (output device) displaying the visual interface.
3. The input devices – mouse and keyboard.

Interaction with the audio augmented point cloud to carry out the tests was done using response locations. Figure 9 shows an octree partitioned point cloud (the octree is omitted). The shown squares are screen projections of five randomly selected octree nodes. These offer the tester the means of interacting with the audio augmented point cloud and will be referred to as *response locations*. (Figure 9 can be assumed to be what the computer screen is displaying to the tester in figure 8.) A sound emitting source can be associated with a response location. The use of response locations offers the tester the ability to determine spatial information of the associated sound source.

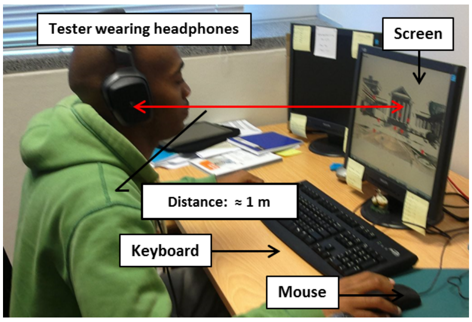

Figure 8: The test subject wearing Logitech G35 headphones while interacting with the auditory interface using input devices.

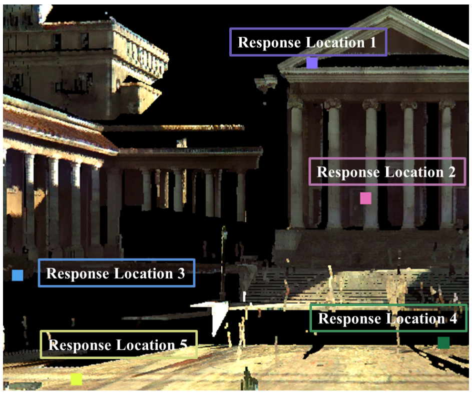

Figure 9: A point cloud with randomly chosen nodes made response locations. The octree is omitted in this figure.

4.3.1 Implementation of Depth Estimation in Point Clouds:

Euclidean depth is given by equation 8,

$$d(x, y, z) = \sqrt{(x_l - x_s)^2 + (y_l - y_s)^2 + (z_l - z_s)^2} \qquad (8)$$

where, x_l, y_l, z_l are the 3D coordinates of the virtual listener and x_s, y_s, z_s are the 3D coordinates of the sound source. The virtual listener is placed at the origin of the OpenAL coordinate system for the tests, therefore x_l, y_l, z_l are all 0.0. Because the x and y coordinates of the target sound source will keep changing, as will be explained later, the easiest depth to help explain aspects of this set of tests is the one along the Z-axis and will simply be referred to as *depth*. Sound source intensities are attenuated by the euclidean depth.

4.3.2 Implementation of Occupancy of Space in Point Clouds:

OpenAL has means of creating reverberant environments for auditory interfaces created through its effects extensions. Aural experiences from real reverberant environments are simulated in this manner in auditory interfaces. The aural experience from OpenAL depends on the type of reverberant environment simulated. OpenAL has a number of reverberant environments that can be used and these can be found from Creative-Labs (2010). The virtual listener was immersed in the following three environments where the user needed to distinguish between the reflections experienced in these environments:

1. *Cave environment* – reverberation typical of caves.
2. *Hallway environment* – reverberation experienced in hallways.
3. *Room environment* – reverberation of a normal room.

A *waterdrop* wave sound was used for this investigation. In these tests, the environment the virtual listener was immersed in was randomly picked from the listed environments. The test subject therefore did not know which environment the virtual listener was immersed in and had to make that decision based on the aural experience when sound was emitted. The aim here was to test if reverberations were experienced by the test subject.

The keyboard was used in depth and environment ambience tests. These interactions will be explained further when the respective tests are discussed. Through all the tests, one target sound source whose spatial information need to be determined was be used.

5. LIMITATIONS OF AUDIO AUGMENTED PROCESSING

5.1 Sound Source Depth Estimation

The aim here is to determine the depth of the target sound object at various response locations on the screen. The sine tone was used in these tests. One response location at a time would appear on the screen, rather than multiple. The displaying of a response location on the screen was done systematically – this is illustrated using figure 10.

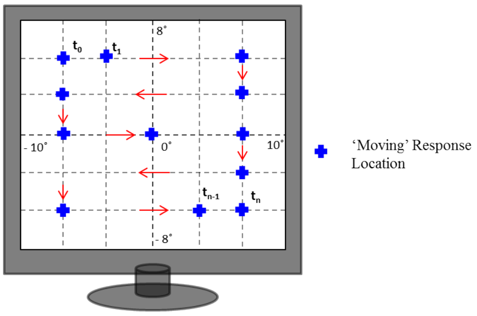

Figure 10: Displaying of one response location on the screen changing with time.

Shown in figure 10 is a response location 'moving' with time. At time t_0, the response location appears at the first location on the screen. After a set number of seconds, the response location appears at the second location, this is at time t_1. This process continues until the response location appears at the last location at time t_n, where n is the total number of locations where the response location needs to appear, minus one. This was done so that depth tests can be done at as many locations as possible. At each response location, the target audio source would have a randomly chosen depth away from the virtual listener. Depth was chosen from two sets containing three depth values each, Set A: {-0.75, -1.5, -2.25}m and Set B: {-1.0, -1.75, -2.5}m. The reason for making this random was so that the tester would have to estimate each time what the depth was. Determination of depths of sound sources is a crude exercise, as previously noted. This is why the target sound source was not made to vary between multiple depths with small variations between them. The idea here was to estimate which depth the target sound source had.

It emerged that for this work, intensity alone is not enough to act as a depth cue. As a result, pitch variation was also used to act as a depth cue. This means that at different depths, the sine tone had different frequencies, leading to these pitch variations. The sine tone was generated with a default frequency of 0.44 kHz. As an indicator of closeness of the target sound source to the virtual listener, the frequency was increased as the depth decreased. The frequency-depth pairing was therefore as follows for both sets of depths: {35.64:-0.75, 1.32:-1.5, 0.44:-2.25} and {35.64:-1.0, 1.32:-1.75, 0.44:-2.5} (kHz:m). These frequency values were chosen because between them there were good noticeable aural differences, i.e., the pitch changes were not discreet.

Given that the target sound source had different intensity and pitch at each depth, the tester had to make depth estimations based on these depth signatures. The tester had to press a number on the keyboard to make the depth estimations. The choices were as follows for Set A: key 0 for depth -0.75 m, key 1 for depth -1.5 m and key 2 for depth -2.25 m. Similarly, for Set B they were: key 0 for depth -1.0 m, key 1 for depth -1.75 m and key 2 for depth -2.5 m.

Depth estimations were made at a total of 72 response locations for each test. Figure 11 shows response locations and the frequency-depth pairs that the target sound object had at each response location for first test of Set A depths. (Figures for second and third tests for Set A and first, second and third tests for Set B show different frequency-depth pairs at different response locations. These are omitted here as they illustrate the same idea.) The sizes of the 'bubbles', i.e., response locations, indicate the depth value – the smallest 'bubble' for the smallest depth, etc.

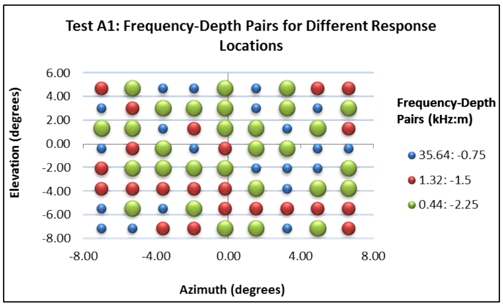

Figure 11: The Frequency-Depth pairs for different response locations for first test of set A depths.

5.2 Occupancy of Space

Here the investigation is about whether the tester can determine the occupancy of space judging from its ambience when a sound is emitted. This could serve to provide information about existing objects which the user can not visually witness. This will investigate the existence of unseen events and their surroundings when the user processes a point cloud. Three sets of tests with 50 trials each were done. In each trial, the test subject was required to make a judgement of which environment the virtual listener was immersed in by pressing a key on the keyboard.

Accurately judging the environment the listener is immersed in, indicates that the test subject does notice the reverberation. Moreover, it shows the tester's ability to identify the nature of those reflections in terms of whether they are from a closed or an open space.

6. RESULTS AND ANALYSES

6.1 Depth Estimation Results

The tester made correct depth estimations of the target sound source for all response locations in all tests. The random change of depth of the target sound source with respect to the virtual listener did not affect the estimations. As the target sound source 'moved' to the currently displayed response location, the euclidean distance between the source and the listener also changed, in turn

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-4-187-2014

changing the sound source's intensity. However, the use of different frequencies as indicators of depth aided with the depth estimations.

It emerged from the depth estimation tests that the tester needed to concentrate on the sine tone being emitted by the target audio object without distractions. With changes in the euclidean depth between the virtual listener and the target sound source, intensity changes became harder to detect. To some extent, this was true with regards to frequency changes too.

Depth estimations for Set B depths were the hardest to make. The reason for this is that Set B depths were slightly greater than those of Set A. As a result, the intensities at Set B depths were lower, leading to discreet changes when the target sound source changed depths. The frequency variations became discreet too, particularly when the target sound source was at the peripheral response locations, where the euclidean depths were greatest.

The tester required some time to make the depth estimations. Response locations appeared on the screen at a rate of 4 s giving the tester 4 s to make an estimation. In some cases, particularly for Set A tests, the tester could make estimations in about 2 s. It took the tester longer to make the estimations in some instances, especially for Set B tests. This is alluded to the fact that the variations were more discreet for these tests.

Making correct depth estimations in all instances does not imply that depth estimations can be made error free. Reviewed literature suggests, making depth estimations is a crude exercise. As a result, the focus here was on the depths that would offer less discreet variations for the sake of detecting if the depth has changed or not.

The stated results suggest that depths of audio augmented objects in point clouds can be estimated accurately using intensity and pitch as depth cues. The changes in these depths can be detected using these cues. Discreet depth changes could be harder to detect, however. Therefore, while performing point cloud processing which requires depth estimation, the user must be cognisant of this.

Using more depth variations for each set of depths could have led to different results. Having more depth variations was problematic because there were more keyboard keys to choose from, leading to errors resulting from pressing the wrong key even if the tester knew what the correct depth was.

Calibration was needed before the tests could be carried out. This was so that the tester could become familiar with the sound source since familiarity aids in depth estimations. This entailed the tester having to programmatically change the depth of the target sound source to get an aural impression at different depths. The calibration was done so that the tester knew what intensities and frequencies to expect at what depths. This calibration process proved vital in the tests as the tester had a point of reference.

6.2 Occupancy of Space

Here the investigation is on the existence of unseen events in point cloud processing using reflections of audio augmented in a point cloud. The test subject was required to identify the type of reverberant environment a sound source was in, with three choices to choose from.

For cave, hallway and room environments, the accuracies were 33.3%, 35.3% and 35.3%, respectively. The reflections experienced in these closed indoor environments are very similar, hence the difficulty in separating one from the others. For example, if the sound source was being emitted from a room environment, the tester confused it with a hallway or cave environment.

The ability of the tester to identify the ambience of an environment can provide information about the occupancy of space of a point cloud. This implies that judging by the nature of the reflections, the user might be able to infer the unseen events in audio augmented point cloud processing.

To improve the accuracy of identifying the ambience of an environment, the user needs to listen to sound being reflected in different environments and get an understanding of how it is reflected. In doing the tests, the test subject had to go through a training process in order to understand these different reflections. The ability to determine how sound is reflected in a particular situation, can potentially inform the user the nature of the unseen event and its surroundings while processing a point cloud.

7. CONCLUSIONS

Depth estimation of sound sources using intensity and frequency variations as cues can prove useful in roughly detecting the depth of points/objects of interest in point clouds. This can be useful in point cloud processing tasks. For example in a case where the depth between two surfaces to be aligned is required.

Depths to/of unseen areas of interest in a point cloud can be detected, even though this will be crude. The aspect of familiarisation with a particular tone and its intensity at different levels could prove to be vital in auditory interfaces. The value of being familiar with a sound stimuli was demonstrated by Begault (1994).

The inability to detect fine depth variations using audio cues could be limiting in audio augmented point clouds. In this respect, point cloud processing that depends on audio depth could require some time for the user to master. In coarse scan registration instances where fine depth variations are not particularly required, the depth cues could be beneficial.

With thorough knowledge of sound reflections in different environments, with practice, the user can be able to determine the nature of reflections. One can also get a sense of how clustered an environment is as that affects sound reflections depending on where the sound source is placed.

Using the nature of these reflections, the user can possibly infer the occupancy of space where visual cues are limited. As seen, some events might lead to similar reflections and therefore confuse the user. Familiarity with sound reflections could lead to better results and therefore enhance coarse scan registration. Determination of occupancy of space in indoor environments is crucial and using audio reverberation could help improve this process.

The objective of augmenting audio in coarse scan alignment for indoor mapping and modelling has been realised. The cues used in extracting specific information have been identified. The accuracies of these cues and their limitations have also been stated. Audio augmentation is not only limited to scan alignments. Other processing problems such as cleaning, simplification, hole filling could benefit from this, where information such as position, shapes and sizes of objects could be estimated using audio cues.

Position estimations of audio objects could also aid in indoor navigation problems where indoor models are used to simulate crisis response. For example, by placing audio object at a destination, the user could use audio cues to estimate the depth and position of the destination. The occupancy of spaces in indoor environments can be estimated in a similar manner using relevant audio cues.

REFERENCES

Begault, D., 1994. 3D Sound for Virtual Reality and Multimedia. Academic Press Inc.

Brenner, C., Dold, C. and Ripperda, N., 2007. Coarse orientation of terrestrial laser scans in urban environments. ISPRS Journal of Photogrammetry and Remote Sensing 63(1), pp. 4–18.

Creative-Labs, 2010. Openal.

Handel, S., 1989. Listening. Massachusetts Institute of Technology.

Kapralos, B., Jenkin, M. R. and Milios, E., 2003. Auditory perception and spatial (3d) auditory systems. Department of Computer Science, York University, Tech. Rep. CS-2003-07.

Mershon, D. H. and King, L. E., 1975. Intensity and reverberation as factors in the auditory perception of egocentric distance. Perception & Psychophysics 18(6), pp. 409–415.

Shinn-Cunningham, B. G., 2000. Distance cues for virtual auditory space. In: Proceedings of the First IEEE Pacific-Rim Conference on Multimedia, pp. 227–230.

Väljamäe, A., 2005. Self-motion and presence in the perceptual optimization of a multisensory virtual reality environment. Technical report, Chalmers University of Technology.

Wozniewski, M. and Settel, Z., 2007. User specific audio rendering and steerable sound for distributed virtual environments. In: Proceedings of the 13th International Conference on Auditory Display.

Xie, Z., Xu, S. and Li, X., 2010. A high-accuracy method for fine registration of overlapping point clouds. Image and Vision Computing 28(4), pp. 563–570.

Zahorik, P., Brungart, D. S. and Bronkhorst, A. W., 2005. Auditory distance perception in humans: A summary of past and present research. Acta Acustica united with Acustica 91(3), pp. 409–420.

Zlatanova, S., Sithole, G., Nakagawa, M. and Zhu, Q., 2013. Problems in indoor mapping and modelling. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XL(4/W4), pp. 63–68.