# Auditory and Visio-Temporal Distance Coding for 3-Dimensional Perception in Medical Augmented Reality

Felix Bork\*

Bernhard Fuerst<sup>†</sup>

Anja-Katharina Schneider<sup>‡</sup>

Francisco Pinto§

Christoph Graumann<sup>¶</sup>

Nassir Navab

Technische Universität München, Munich, Germany Johns Hopkins University, Baltimore, MD, United States

## ABSTRACT

Image-guided medical interventions more frequently rely on Augmented Reality (AR) visualization to enable surgical navigation. Current systems use 2-D monitors to present the view from external cameras, which does not provide an ideal perception of the 3-D position of the region of interest. Despite this problem, most research targets the direct overlay of diagnostic imaging data, and only few studies attempt to improve the perception of occluded structures in external camera views. The focus of this paper lies on improving the 3-D perception of an augmented external camera view by combining both auditory and visual stimuli in a dynamic multi-sensory AR environment for medical applications. Our approach is based on Temporal Distance Coding (TDC) and an active surgical tool to interact with occluded virtual objects of interest in the scene in order to gain an improved perception of their 3-D location. Users performed a simulated needle biopsy by targeting virtual lesions rendered inside a patient phantom. Experimental results demonstrate that our TDC-based visualization technique significantly improves the localization accuracy, while the addition of auditory feedback results in increased intuitiveness and faster completion of the task.

Index Terms: Medical Augmented Reality, Multi-Sensory Environment, Temporal Distance Coding, Auditory and Visual Stimuli.

## **1** INTRODUCTION

In Augmented Reality (AR), the perception of the real world is enhanced by incorporating virtual data that appears to coexist in the same space [3]. In most cases, these augmentations are limited to visual overlays, ranging from simple virtual annotations [33] to complex photo-realistic renderings [1]. However, as AR systems advance, the integration of data from different sensors is increasingly investigated, for example olfactory, auditory or haptic data.

The medical field has long been recognized as an application area of AR with potential for great benefit [4]. In the past decade, image-guided surgeries have experienced increased popularity for many different applications [11, 37]. Biopsies are one particular group of interventions increasingly performed under imageguidance. They are of great importance for evaluating lymph node involvement in cancer [19], and for the staging of suspicious lesions detected by pre-interventional imaging [9, 36]. Both procedures can be performed by either invasive (open) or needle biopsy. During open biopsy, the skin of the patient is cut and the region of interest is resected, increasing the risk of interventional bleeding and post-operative infection. In contrast, needle biopsies are less invasive, but require a higher precision in targeting the region of interest [10, 30], for instance of abnormal breast lesions [31]. The prostate [26], the liver [25], and the lung [20] are other typical biopsy sites. Increasing the likelihood of hitting the desired biopsy target and therefore preventing false-negative diagnoses is one of the main objectives of current research. An intra-operative view, augmented with pre-interventional information, may help improve the 3-D perception and therefore the localization accuracy of needle biopsies [27].

In this paper, we propose a new kind of multi-sensory AR environment consisting of auditory and visual augmentations of an external camera view to improve localization accuracy of needle biopsies. It is based on a technique called Temporal Distance Coding (TDC), first suggested as a general concept for improving Mixed Reality visualizations by Furmanski et al. [13]. In TDC, the point in time at which a virtual object is rendered depends on its distance to a certain reference point, e.g. the tip of the biopsy needle. A propagating virtual shape initialized at the reference point controls the start of the object's rendering period. This paradigm of augmenting an external camera view based on user actions with a dynamic surgical tool may help improve the perception of the 3-D location of virtual objects of interest. In addition to the visual augmentation, a major contribution of this work is the incorporation of auditory feedback. A repeating tone similar to the one of a metronome is played every time the propagating shape has reached a multiple of a specific distance. Another bell-like tone indicates the intersection of the propagating virtual shape with an object of interest and therefore the start of its rendering period.

# 2 BACKGROUND & RELATED WORK

Our proposed multi-sensory AR environment combines both visiotemporal and auditory stimuli. While various systems using the former have been developed, acoustic signals have not been studied in the medical context yet. In this section, we review proposed medical AR systems using visual augmentations and general systems integrating auditory feedback with a focus on perception.

## 2.1 Medical Visual Augmented Reality

Medical AR applications developed for needle biopsies initially focused on the visualization of occluded instruments, for instance showing the projection of a biopsy needle onto an ultrasound image in an AR view [32]. This was found to significantly increase the biopsy accuracy [27]. The system of Wacker et al. augmented a live video stream with MRI images acquired prior to the intervention [35]. The guidance of the needle is supported by rendering a virtual disk around the target lesion. The diameter of the disk depends on the distance between needle and lesion and decreases as the needle is inserted towards the lesion. Both approaches require wearing Head-Mounted Displays (HMDs), which have significantly improved over the last years as a result of their introduction in the gaming and entertainment industry. However, HMDs

<sup>\*</sup>e-mail:felix.bork@jhu.edu

<sup>&</sup>lt;sup>†</sup>e-mail:be.fuerst@jhu.edu

<sup>&</sup>lt;sup>\*</sup>e-mail:anja.schneider@tum.de

<sup>§</sup>e-mail:pinto@jhu.edu

<sup>¶</sup>e-mail:c.m.graumann@jhu.edu

e-mail:nnavab1@jhu.edu

still face critical handling challenges and raise concerns regarding reliability during medical procedures, which have prevented their wide-spread adoption up to now [29].

In general, current systems for surgical navigation use external cameras and monitors to present data inside the operating room. This does not obscure the surgeon's view of the patient and the consequences of a system failure are minimized. Furthermore, by augmenting the external camera view with additional information, no new hardware or technology needs to be introduced. Augmented external camera views have been demonstrated by Nicolau et al. for liver punctures [22] and liver thermal ablations [23]. Both approaches also allow the rendering of a virtual view from the tip of the biopsy needle. First experiences with medical AR in neurovascular interventions by Kersten-Oertel et al. have indicated increased understanding of the topology, potential reduction of surgical duration and increase in accuracy [16].

When the AR visualization is implemented as a simple superimposition of virtual objects on the video stream, the virtual objects appear to float above the anatomy. This lack of correct depth perception has been recognized as a major challenge for AR visualization [3]. The human brain uses several monocular and binocular cues to assess the depth of an object, some of which are occlusion and motion parallax. By rendering a virtual window which changes the position based on user interaction, the perception of the virtual object may be improved [8]. More recent approaches aim at changing the transparency of the video stream in certain regions to create a see-through effect. For instance, Bichlmeier et al. calculate a transparency value depending on the view direction, distance from the focus region and the surface curvature for each pixel in a video stream [7]. In contrast to attempts aimed at making the augmentation more realistic, non-photorealistic rendering and transparency calculation based on the pq-space of the surface are used to improve depth perception by Lerotic et al. [18]. This can be useful in scenarios where the occluded objects may appear similar to the surface or need to be in the surgeon's center of attention. The major drawback of these approaches is that knowledge of the surface is required. This is associated with additional hardware in the operating room and therefore difficult to obtain in many clinical scenarios.

In this work, we propose the use of Temporal Distance Coding to combine both accurate localization and improved 3-D perception of an augmented external camera view. In addition to that, we incorporate auditory feedback and aim at reducing the procedure time while simultaneously further increasing the accuracy and intuitiveness of our technique.

# 2.2 Audio Augmented Reality

Existing research publications concerned with the topic of audio AR can roughly be divided into two main groups: those that focus on purely auditory augmentations and those that combine both auditory and visual stimuli. Early work by Mynatt et al. presents a system that uses infrared signals emitted by active badges to detect location changes and to trigger auditory cues in an office environment [21]. Bederson et al. introduce a prototype system utilized to guide visitors through a museum by playing recorded audio messages in the vicinity of interest points [5]. Similar systems were developed for outdoor environments. Rozier et al. present the concept of audio imprints, short audio notes, that can be placed inside the real world using a GPS system [28]. A linear story line of a cemetery tour is complemented with location based audio messages by Dow et al. [12]. However, the addition of a visual interface for displaying quantity and type of surrounding audio notes is recognized as a potential improvement in both of these systems. Haller et al. present such a hybrid system, consisting of both auditory and visual augmentations [15]. A simple pen as input device is used to position 3-D sound sources, represented as virtual loudspeakers, into the real 3-D world using an intuitive drag and drop scheme. No

evaluation results are reported to support their work, though. The concept of Audio Stickies for mobile AR applications is introduced by Langlotz et al. [17]. These short spatial audio annotations are visually represented by differently colored dots and are modulated in terms of loudness and stereo channel depending on the users position and orientation. A usability-centric explorative field study was conducted to analyze the system, which proved to provide the user with valuable information. However, the overlapping of multiple sound sources, also known as sound clutter, was identified as a maior challenge in real-life situations. Vazquez-Alvarez et al. report that only up to two simultaneously playing sound sources can still be perceived as such by the user [34]. Closely attended by this is the study on accurate localization of 3-D sound sources. Early work focused on estimating the azimuth or direction of 3-D sound sources, while recently auditory distance perception (ADP) is studied more extensively [38, 2]. Works by Behringer et al. [6] and Otskui et al. [24], in which audio AR is an integral part of the user interface, are of greater relevance to our proposed solution. The former presents an AR system for maintenance and error diagnostics, which uses 3-D audio techniques to indicate objects outside of the users field of view. In the ladder, a novel mixed reality interaction paradigm for manipulating complex 3-D models is presented. Virtual elastic bands represent connections between objects which the user can break by pulling an object out of a specified area. Different sounds complement the visual augmentation and indicate successfully broken and newly established connections.

To the best of our knowledge, the application of audio AR in combination with visual feedback has not yet been explored in the medical context. We propose using sound to complement a previously purely visual AR environment for 3-D localization of virtual structures. As the human auditory system is very good at detecting rhythmic irregularities [14], we use the familiar tone of a metronome for indicating equidistant steps of the propagating virtual shape and a bell-like tone for the intersection of the propagating virtual shape with an object of interest indicating the begin of its rendering period.

#### **3** AUDITORY AND VISIO-TEMPORAL DISTANCE CODING

In this section, we give a detailed description of how auditory and visual stimuli are incorporated in our proposed multi-sensory AR environment. Both perception enhancing techniques are based on the distance between a reference point, for instance the tip of a tracked surgical tool, and the virtual object of interest.

## 3.1 Visio-Temporal Distance Coding

Visio-temporal distance coding is comprised of two main components: a propagating shape  $\Pi$  and an animation cycle for each of the multiple (*n*) objects of interest  $\Omega_i$ . Intersections of the propagating shape  $\Pi$  with fixed objects of interest  $\Omega_i$  allow the dynamic perception of relative distances.

#### Visualization of Distance-Encoding Propagating Shape

The virtual shape  $\Pi$  propagates through the environment at a constant speed. Once triggered, it is initialized at the tool tip and propagates until it reaches maximum propagation distance  $d_{max}$  at time  $t_{max}$ . Different types of propagating virtual shapes are selectable, such as a plane, hemisphere or sphere. However, if the penetration distance of the tool is of crucial importance, then non-uniformly propagating shapes, such as plane and hemisphere, could be misleading. Especially in scenarios where the tool penetration distance exceeds the distance of the object of interest from the biopsy entry site, users may be confused as the object would not get hit by the non-uniformly propagating shape and therefore not get rendered. Hence, we will use a sphere centered at the tip of the surgical tool whose propagation is represented by an increasing radius rather than distance from the tool tip for the rest of this paper. Fig. 1

illustrates such a propagating virtual sphere rendered in our virtual testing environment.



Figure 1: The propagating shape is implemented as an epicentric sphere expanding from the reference point towards its maximum expansion  $d_{max}$ . The biopsy needle is represented by a virtual red-white striped cylinder. In addition to the propagating shape, multiple intermediate shapes  $\Pi_{\tau_k}$  at steps  $\tau_k$  as well as a shape representing the maximum propagation amount are rendered in wire-frame mode.

## Animation Cycle for Regions of Interest

The rendering period of duration *T* for objects of interest  $\{\Omega_i\}_{i=1}^n$  is initiated upon intersection with the virtual propagating shape  $\Pi$  at a time  $t_i = t_c(d_i)$ , where  $d_i$  is the distance between the object of interest  $\Omega_i$  and the reference point, for instance a tracked needle tool, and  $t_c()$  is the function to determine the interaction (see Sec. 3.3). Consequently, objects closer to the reference point are animated earlier than objects farther away. If the distance between an object of interest and the reference point is greater than  $d_{max}$ , then it will not be rendered at all. The animation for the visio-temporal distance coding defines the transparency of the objects of interest  $\Omega_i$ , and can be formulated as a function depending on the a control function  $\Psi(t)$  and a set of indicator functions  $\mathbb{1}(A, x)$ :

$$\alpha_{\mathbf{\Omega}_i}(t,d_i) = \boldsymbol{\psi}(t) \cdot \mathbb{1}([t_i,t_i+T],t) \cdot \mathbb{1}([0,d_{max}],d_i), t \in [0,t_{max}],$$
(1)

with the indicator function  $\mathbb{1}(A, x)$  defined as

$$\mathbb{1}(A,x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}.$$
 (2)

The function  $\psi(t)$  controls the smoothness of the animation of the object of interest. In this work, a simple step function is used for  $\psi(t)$ . Thus, lesions are immediately visible upon collision with the propagating shape.

While a single propagating shape can be used to assess the relative distance of virtual lesions to the reference point as well as distance ratios between virtual lesions, determining the absolute distance is significantly more difficult. To overcome this, we propose rendering equidistantly spaced intermediate shapes  $\Pi_{\tau_k}$  in addition to the propagating shape  $\Pi$ . At every time step

 $\tau_k$  such an intermediate shape is rendered, where  $\tau_k$  is calculated as:

$$\tau_k = \frac{k\Delta}{\nu}, \text{ where } k \in \mathbb{N} : 0 < k \le \frac{d_{max}}{\Delta},$$
(3)

with  $\Delta$  the distance between two intermediate shapes  $\Pi_{\tau_k}$ . Pseudocoloring is used as an assisting depth cue by interpolating the color of the shapes between red (close) and blue (far).

# 3.2 Auditory Distance Coding

Similar to the visio-temporal distance coding, the auditory distance coding is applied both to the propagating shape  $\Pi$  and the objects of interest  $\{\Omega_i\}_{i=1}^n$ . In the sonification process, we encode the propagation of the virtual shape with regular, metronome-like tones. This acoustic feedback is played at time steps  $\tau_k$  in conjunction with the rendering of an intermediate shape  $\Pi_{\tau_k}$  and is aimed at improving the understanding of scale and velocity v of the propagating shape and therefore the overall intuitiveness of our visualization. By counting the number of tones signaling the elapsed propagation distance, the user can estimate the distance to a virtual lesion using the auditory feedback.

Upon collision with the virtual propagating shape, a second tone is played and coincides with the beginning of a lesion's rendering period. Requirements for this tone are a short duration to minimize the time between visual and auditory stimuli an easy distinction to the regular auditory feedback of the shape  $\Pi$ . For our experiments, we chose a bell-like tone with a high pitch satisfying both the aforementioned characteristics.

# 3.3 Determination of Interaction Time

In order to determine the time  $t_i = t_c(d_i)$  at which the propagating shape intersects a region of interest, simple collision detection algorithms are employed. A set of *m* spheres  $\{s_c\}_{c=1}^m$  is defined to provide an approximation of the complex surface of the objects of interest. These spheres are equally spaced, and each is defined by their center  $\vec{c_c}$  and radius  $r_c$ . Collisions are detected when the propagating sphere  $\Pi$  with its radius  $r_{\pi}$  centered at  $\vec{c_{\pi}}$  intersects with one of the surface spheres:

# 4 EXPERIMENTS AND RESULTS

In a clinical scenario, a segmentation of a medical image is computed, and the resulting surfaces are rendered into an external camera view. Surgeons perform needle biopsies based on the augmented view and haptic feedback (touch), in combination with their knowledge of the anatomy. To evaluate our auditory and visiotemporal distance coding environment, we simulated needle biopsies and removed haptic feedback (in terms of varying tissue densities) in conjunction with the anatomical constraints to limit the number of interfering variables in our setup. Fig. 2 illustrates the auditory and visio-temporal distance coding over time.

#### Experimental Setup

Similar to currently deployed clinical systems, our experimental hardware setup included a video camera, an infrared tracking system to detect spherical, retro-reflective markers attached to a small frame on the back of the biopsy needle and a 2-D monitor. A Logitech HD Pro Webcam C920, Logitech International S.A., Lausanne, Switzerland, was mounted together with a Polaris Vicra tracking system, Northern Digital Incorporated, Waterloo, Canada, on a weight compensating arm.

Calibration between the RGB and infrared (IR) tracking cameras was performed using a specially designed calibration device consisting of both a checkerboard pattern and optical markers. By taking multiple RGB images of the calibration device for various locations and simultaneously capturing its IR-tracked pose, 3D-2D



Figure 2: Individual steps of an auditory and visio-temporal distance coding animation cycle: regular auditory feedback for the propagating sphere (a, c, e, g, h), and irregular acoustic signals to indicate the intersection of the propagating sphere with three objects of interest (b, d, f). A virtual model of the biopsy needle (gray) is overlayed on top of the video stream.

correspondences are obtained to calculate the transformation between the optical centers of the two cameras.

#### **Experimental Conditions**

For the purpose of evaluating the localization accuracy of our auditory and visio-temporal distance coding technique, we designed a user experiment simulating a needle biopsy procedure. Virtual lesions were positioned inside the breast area of a patient torso phantom made out of hard foam material. The task of every subject consisted in inserting an optically tracked needle into these virtual lesions. Four different conditions were tested: (A) simple overlay of virtual lesions, (B) auditory feedback only, (C) visio-temporal feedback only, and (D) combination of auditory and visio-temporal feedback. Based on these conditions, we formulated three hypotheses H1-H3, that are subject of investigation during our experiments: Conditions B, C, and D significantly outperform the simple overlay visualization A in terms of accuracy (H1), and the hybrid condition D both yields the best overall accuracy (H2) and significant improvements in task completion time over condition C, the purely visual augmentation (H3). Three sets of lesions were presented for each condition. Each set consisted of three virtual lesions, yielding a total of twelve biopsies per subject. Among the study participants, we randomized the order of conditions during the experiment. We asked all subjects to verbally confirm the correct insertion of the biopsy needle into a virtual lesion and successively computed the distance between the current needle tip and the closest point on the surface of the lesion. In case of sucessful positioning of the needle tip inside the lesion (i.e. a lesion hit), this distance was considered zero. Positions of the virtual lesions were calculated inside a cube of dimensions  $20 \times 20 \times 20$  cm below a tracked patient target (see Fig. 2). We fixed the maximum propagation distance  $d_{max}$  of the virtual shape, ensuring that all lesions are rendered when the biopsy needle was placed on the biopsy entry site of the phantom. A total of 15 subjects participated in the study with a mean age of 25.8 years (from 23 to 31 years); two female and 13 male participants. All subjects took part voluntarily and did not get any reward.

## **Evaluation Results**

Most current systems employed in clinical environments augment the regions of interest by overlaying the information without additional feedback. Therefore, a very simplified version of this AR mode was used to establish a baseline in condition (A), which led to the highest errors in localization and the lowest percentage of hit objects of interest. Auditory (B), visio-temporal (C) and the combination of auditory and vision-temporal coding (D) improved the accuracy and lead to an increased hit percentage. However, the additional information also led the users to perform the task slower, indicating that the lack of depth information in condition (A) causes an early abortion of the biopsy procedure. Overall, the results clearly show that the combination of auditory and visio-temporal coding (D) outperforms the three other AR modes in all criteria. Results are summarized in table 1.

## Statistical Evaluation

Statistical tests were performed to study the biopsy accuracy and task completion time for the four different conditions. A Friedman test was calculated to compare localization accuracy as a normal distribution of the data could not be assumed. We found a significant difference in accuracy depending on the kind of assistance that was provided to the subjects,  $\chi^2(3) = 225.53$ , p < 0.001.

Table 1: Comparison of four different conditions for AR based needle biopsies. The categories for comparison are localization error, condition completion time (9 biopsies each) and percentage of lesion hits. The simple overlay (A) does not provide any depth information, resulting in the lowest accuracy and lowest percentage of lesion hits. The combination of auditory and visio-temporal distance coding (D) enabled the users to perform best in terms of accuracy.

Conditions	Localization Error		Condition Completion Time		Lesions Hit
	Mean $\mu$	SD $\sigma$	Mean $\mu$	SD σ	
(A) Overlay (No Feedback)	25.60 mm	23.34 mm	50.64 s	21.71 <i>s</i>	14.07 %
(B) Auditory Distance Coding	0.46 mm	1.12 mm	132.89 s	29.67 s	77.04 %
(C) Visio-Temporal Distance Coding	0.82 mm	1.44 mm	156.98 s	33.29 s	62.96 %
(D) Auditory + Visio-Temporal Combined	0.24 mm	0.77 mm	127.38 s	26.16 s	81.48 %

Wilcoxon signed-rank tests with Bonferroni correction were calculated as post-hoc tests. They showed significant differences between (A) simple overlay and (B) auditory feedback (Z = -9.44, p < 0.001) as well as between (A) simple overlay and (C) visio-temporal feedback (Z = -9.45, p < 0.001) and (A) simple overlay and (D) the combination of auditory and visio-temporal feedback (Z = -9.36, p < 0.001). Furthermore the post-hoc test revealed that the accuracy was higher when using (B) sound as assistance than (C) the propagating sphere only (Z = -2.49, p = 0.013). The combination of auditory and visio-temporal feedback (D) was superior to using (B) auditory (Z = -2.06, p = 0.039) or (C) visual (Z = -4.39, p < 0.001) assistance alone.

For comparing the time necessary to complete the biopsies, we employed a univariate ANOVA with repeated measures which yielded significant differences at the p < 0.01 level (F(3, 42) = 111.85, p < 0.001). Post-hoc tests revealed that task completion time was significantly higher for the conditions using (B) auditory assistance (p < 0.001), (C) visual assistance (p < 0.001) or (D) the combination of auditory and visual assistance (p < 0.001) compared to the simple overlay (A). Furthermore, task completion time was higher when (C) visio-temporal feedback was presented compared to (B) the auditory condition (p = 0.004). The hybrid approach (D) improved task completion time significantly compared to (C) the use of visual feedback only (p = 0.004).

## 5 DISCUSSION

The experiments show that the combination of auditory and visiotemporal coding significantly improves the accuracy and the percentage of successfully performed needle biopsies. Although not explicitly evaluated, all participants mentioned that the hybrid condition is the most intuitive way to perform needle biopsies. This is consistent with the results observed in the experiments, which showed the highest accuracy for condition D. As conditions B and C also significantly outperformed the simple overlay visualization, it is possible to confirm both hypotheses H1 and H2. The average duration of a biopsy was significantly higher for conditions B, C, and D compared to the simple overlay condition A. This may be explained by two facts: Firstly, the lack of depth information for the simple augmentation, which caused users to guess the position rather than trying to place the needle correctly, and secondly the necessary learning phase when users are confronted with novel visualizations and user interfaces. However, the addition of auditory feedback in the hybrid approach improved task completion time significantly, therefore confirming hypothesis H3 as well. Future evaluations could compare our approach to existing visualization techniques by Bichlmaier et al. [8] or Lerotic et al. [18], that aim at improving depth perception. Currently, those solutions are applied only to the surgeon's direct view, e.g. using HMD, surgical microscopes or to the view of a medical imaging system. Therefore, we think that the comparison with a simple overlay establishes the correct baseline to properly evaluate the accuracy improvements our novel multi-sensory AR environment provides. In our experiments, we used a fixed maximum propagation distance up to which feedback about collision with virtual objects of interest is provided. In future experiments, this distance could be computed automatically and coincide with the distance of the farthermost object of interest.

Another topic for future research is that of a dynamic user interface. In this work, the virtual shapes remain static for the entire animation cycle and do not move along when motion of the needle occurs. However, needle placement is a very slow and precise task. Fast motions are not expected to occur during the procedure and our proposed solution would serve as a status update to the physician during certain parts of the biopsy in realistic application scenarios.

Auditory feedback could turn out to be challenging to incorporate in a realistic surgical scenario, since multiple sound sources, e.g. the sound from a patient vital sign monitor, are well established in the operating room. However, since the auditory feedback will only be used during a limited time of the procedure, this does not pose a major problem. Intra-interventional ultrasound guidance is used for many different needle biopsy procedures. Incorporating this imaging data into our multi-sensory AR setup could increase the acceptance and fast adoption, while providing additional feedback to validate the needle insertion, similar to the work by State et al. [32]. Indicating when the tip of the biopsy needle is successfully inserted by changing the color or shading of a virtual object could serve as additional guidance and support. However, this alone does not provide the user with feedback about the location of objects of interest and their distance to the reference point.

## 6 CONCLUSION

In this paper, we presented a novel multi-sensory augmented reality system for 3-D localization by use of auditory and visio-temporal distance coding. Acoustic and visual feedback is provided by propagating a virtual shape from a reference point, and the interaction of the shape with objects of interest. The combination of auditory and visio-temporal distance coding for medical augmented reality has the potential to improve the clinical care as higher accuracy during needle biopsies results in a lower false-negative rate. The application of this technique in the clinical routine is simple, since the risks and costs of implementation are minimal. In an simulated needle biopsy procedure, we evaluated the impact of auditory and visiotemporal stimuli on 3-D localization accuracy. Our experimental results demonstrate that our temporal distance coding-based visualization technique significantly increases the localization accuracy compared to a simple overlay of virtual objects. The addition of auditory feedback further increased accuracy and was found to be more intuitive while simultaneously yielding a significantly faster time to perform the overall procedure. The outcome of the evaluation strongly motivates the use of this system and further research to initialize pre-clinical trials as soon as possible.

#### REFERENCES

- [1] K. Agusanto, L. Li, Z. Chuangui, and N. W. Sing. Photorealistic rendering for augmented reality using environment illumination. In *Mixed and Augmented Reality, 2003. Proc.. The Second IEEE and ACM International Symposium on*, pages 208–216. IEEE, 2003. 1
- [2] P. W. Anderson and P. Zahorik. Auditory/visual distance estimation: accuracy and variability. *Frontiers in Psychology*, 5(October):1–11, 2014. 2.2
- [3] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *Computer Graphics and Applications, IEEE*, 21(6):34–47, 2001. 1, 2.1
- [4] R. T. Azuma. A survey of augmented reality. Presence: Teleoperators and Virtual Environments, 6(4):355–385, 1997. 1
- [5] B. Bederson. Audio Augmented Reality : A Prototype Automated Tour Guide. Chi '95 - Mosaic of Creativity, pages 210–211, 1995. 2.2
- [6] R. Behringer, S. Chen, V. Sundareswaran, K. Wang, and M. Vassiliou. A novel interface for device diagnostics using speech recognition, augmented reality visualization, and 3D audio auralization. *Proc. IEEE Int. Conf. on Multimedia Computing and Systems*, 1, 1999. 2.2
- [7] C. Bichlmeier, S. Heining, M. Feuerstein, and N. Navab. The virtual mirror: A new interaction paradigm for augmented reality environments. *IEEE Trans. Medical Imaging*, 28(9):1498–1510, 2009. 2.1
- [8] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab. Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In *Proc. IEEE Int. Symp. Mixed and Augmented Reality.* IEEE, 2007. 2.1, 5
- [9] C. Burke, R. Thomas, C. Inglis, et al. Ultrasound-guided core biopsy in the diagnosis of lymphoma of the head and neck. a 9 year experience. *BJR*, 84(1004):727–732, 2011. 1
- [10] M. E. Burt, M. W. Flye, B. L. Webber, and R. A. Wesley. Prospective evaluation of aspiration needle, cutting needle, transbronchial, and open lung biopsy in patients with pulmonary infiltrates. *The Annals of Thoracic Surgery*, 32(2):146–153, 1981. 1
- [11] K. Cleary and T. M. Peters. Image-guided interventions: technology review and clinical applications. *Annual review of biomedical engineering*, 12:119–142, 2010. 1
- [12] S. Dow, J. Lee, C. Oezbek, B. MacIntyre, J. D. Bolter, and M. Gandy. Exploring spatial narratives and mixed reality experiences in Oakland Cemetery. ACE '05: Proc. of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, pages 51–60, 2005. 2.2
- [13] C. Furmanski, R. Azuma, and M. Daily. Augmented-reality visualizations guided by cognition: perceptual heuristics for combining visible and obscured information. In *Proc. International Symp. Mixed and Augmented Reality*. IEEE, 2002. 1
- [14] T. D. Griffiths, C. Büchel, R. S. Frackowiak, and R. D. Patterson. Analysis of temporal structure in sound by the human brain. *Nature neuroscience*, 1(5):422–427, 1998. 2.2
- [15] M. Haller, D. Dobler, and P. Stampfl. Augmenting the reality with 3D sound sources. ACM SIGGRAPH 2002 conference abstracts and applications on - SIGGRAPH '02, page 65, 2002. 2.2
- [16] M. Kersten-Oertel, I. Gerard, S. Drouin, K. Mok, D. Sirhan, D. Sinclair, and D. L. Collins. Augmented reality in neurovascular surgery: First experiences. In Augmented Environments for Computer-Assisted Interventions, pages 80–89. Springer, 2014. 2.1
- [17] T. Langlotz, H. Regenbrecht, S. Zollmann, and D. Schmalstieg. Audio Stickies : Visually-guided Spatial Audio Annotations on a Mobile Augmented Reality Platform. pages 545–554, 2013. 2.2
- [18] M. Lerotic, A. J. Chung, G. Mylonas, and G.-Z. Yang. Pq-space based non-photorealistic rendering for augmented reality. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pages 102–109. Springer, 2007. 2.1, 5
- [19] G. H. Lyman, A. E. Giuliano, M. R. Somerfield, et al. American society of clinical oncology guideline recommendations for sentinel

lymph node biopsy in early-stage breast cancer. *Journal of Clinical Oncology*, 23(30):7703–7720, 2005. 1

- [20] A. Manhire, M. Charig, C. Clelland, F. Gleeson, R. Miller, H. Moss, K. Pointon, C. Richardson, and E. Sawicka. Guidelines for radiologically guided lung biopsy. *Thorax*, 58(11):920–36, 2003. 1
- [21] E. D. Mynatt, M. Back, R. Want, R. Frederick, and S. a. C. M. S. I. G. o. C.-H. Interaction. Audio Aura: Light-weight Audio Augmented Reality. 10th Annual ACM Symposium on User Interface Software and Technology, pages 211–212, 1997. 2.2
- [22] S. Nicolau, X. Pennec, L. Soler, and N. Ayache. A complete augmented reality guidance system for liver punctures: First clinical evaluation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pages 539–547. Springer, 2005. 2.1
- [23] S. Nicolau, X. Pennec, L. Soler, X. Buy, a. Gangi, N. Ayache, and J. Marescaux. An augmented reality system for liver thermal ablation: Design and evaluation on clinical cases. *Medical Image Analysis*, 13(3):494–506, 2009. 2.1
- [24] M. Otsuki, T. Oshita, A. Kimura, F. Shibata, and H. Tamura. Touch & Detach: Ungrouping and observation methods for complex virtual objects using an elastic metaphor. *IEEE Symposium on 3D User Interface 2013, 3DUI 2013 - Proc.*, pages 99–106, 2013. 2.2
- [25] F. Piccinino, E. Sagnelli, and G. Pasquale. Complications following percutaneous liver biopsy. *Journal of Hepatology*, 2(2):165–173, 1986. 1
- [26] L. V. Rodriguez, V. Larissa, M. K. Terris, and K. Martha. Risks and Complications of Transrectal Ultrasound Guided Prostate Needle Biopsy: A Prospective Study and Review of the Literature. *The Journal of Urology*, 160(6):2115–2120, 1998. 1
- [27] M. Rosenthal, J. Lee, Hirota, et al. Augmented reality guidance for needle biopsies: A randomized, controlled trial in phantoms. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI* 2001, pages 240–248. Springer, 2001. 1, 2.1
- [28] J. Rozier, K. Karahalios, and J. Donath. Hear & there: An augmented reality system of linked audio. *Proc. of the International Conference* on Auditory Display, pages 63–67, 2000. 2.2
- [29] T. Sielhorst, M. Feuerstein, and N. Navab. Advanced Medical Displays: A Literature Review of Augmented Reality. *Journal of Display Technology*, 4(4):451–467, 2008. 2.1
- [30] M. C. Skrzynski, J. S. Biermann, A. Montag, and M. A. Simon. Diagnostic accuracy and charge-savings of outpatient core needle biopsy compared with open biopsy of musculoskeletal tumors\*. *The Journal* of Bone & Joint Surgery, 78(5):644–9, 1996. 1
- [31] N. Sneige. Image-guided biopsies of the breast: Technical considerations, diagnostic challenges, and postbiopsy clinical management. In *Breast Cancer 2nd Edition*, pages 163–196. Springer Science + Business Media, 2008. 1
- [32] A. State, M. A. Livingston, W. F. Garrett, et al. Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies. In Proc. of the 23rd annual conf. on Comp. graphics and interactive techniques, pages 439–446. ACM, 1996. 2.1, 5
- [33] K. Uratani, T. Machida, K. Kiyokawa, and H. Takemura. A study of depth visualization techniques for virtual annotations in augmented reality. In *IEEE Proc.*. VR 2005. Virtual Reality, 2005. Institute of Electrical & Electronics Engineers (IEEE), 2005. 1
- [34] Y. Vazquez-Alvarez, I. Oakley, and S. a. Brewster. Auditory display design for exploration in mobile audio-augmented reality. *Personal* and Ubiquitous Computing, 16(8):987–999, 2012. 2.2
- [35] F. K. Wacker, S. Vogt, A. Khamene, J. A. Jesberger, S. G. Nour, D. R. Elgort, F. Sauer, J. L. Duerk, and J. S. Lewin. An augmented reality system for MR image–guided needle biopsy: Initial results in a swine model. *Radiology*, 238(2):497–504, 2006. 2.1
- [36] T. M. Whitten, T. W. Wallace, R. E. Bird, and P. S. Turk. Imageguided core biopsy has advantages over needle localization biopsy for the diagnosis of nonpalpable breast cancer. *The American surgeon*, 63(12):1072–8, 1997. 1
- [37] Z. Yaniv and K. Cleary. Image-guided procedures: A review. Technical report, 2006. 1
- [38] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91(3):409–420, 2005. 2.2