Voice radio communication, pedestrian localization, and the tactical use of 3D audio

John-Olof Nilsson[1], Christian Schüldt[1,2], and Peter Händel[1]

[1]Signal Processing Lab, ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden

[2]Limes Audio, Umeå, Sweden

Abstract—The relation between voice radio communication and pedestrian localization is studied. 3D audio is identified as a linking technology which brings strong mutual benefits. Voice communication rendered with 3D audio provides a potential low secondary task interference user interface to the localization information. Vice versa, location information in the 3D audio provides spatial cues in the voice communication, improving speech intelligibility. An experimental setup with voice radio communication, cooperative pedestrian localization, and 3D audio is presented and we discuss high level tactical possibilities that the 3D audio brings. Finally, results of an initial experiment, demonstrating the effectiveness of the setup, are presented.

I. INTRODUCTION

Pedestrian localization and voice radio communication are key technologies for achieving situational awareness, efficient control, and coordination in dismounted units. They form the basis for higher level functionalities such as navigation, target acquisition, threat alert, and threat cueing and have both been extensively studied in the literature. However, apart from performance limitations of specific technical solutions, both technologies have more general limitations in their user interfaces, which have received less attention. At the commanding level, the localization information is naturally provided with a top-down perspective on a display, together with graphical command and communication tools. Unfortunately, due to secondary task interference, watching such a display is not feasible at all times for all levels of personnel. This will limit and delay the localization information penetration in the unit. Further, voice radio conversations can be hard to comprehend compared to face-to-face conversations; especially when multiple agents communicate on the same channel. Consequently, good radio discipline is required, which prevents natural conversations and limits the pace of the information exchange. This in turn fundamentally limits the level of situational awareness, coordination, and control, which can be achieved with voice radio communication [1,2]. Therefore, in this article we describe how these user interface limitations of pedestrian localization and voice radio communication can be mitigated by tightly integrating them to create 3D audio effects. A small tactical unit experimental setup, illustrated with two agents in Fig. 1, with 3D audio based on the integration of an infrastructure free cooperative pedestrian localization system and voice radio communication, is presented. Further, the tight integration is argued to bring a multitude of other high level tactical possibilities. Finally, initial experimental results of the effectiveness of the suggested setup are presented.
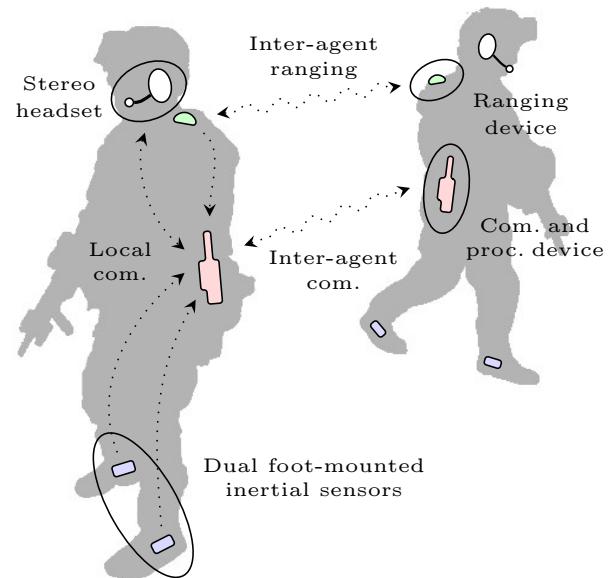

Fig. 1. Illustration of the integrated localization and voice radio communication system setup. Localization is performed by foot-mounted inertial sensors and inter-agent ranging. The processing is done on a common communication (com.) and processing (proc.) platform. The integration makes it possible to create a dynamic 3D audio effect on the voice radio. This improves speech intelligibility and gives the agents a sense of location for their colleagues.

The key observation of the article is that the localization and the voice radio communication are enablers for and have strong mutual benefits in 3D audio. We suggest using 3D audio as a means to give a coarse perception of the relative locations of agents within a group. This is done by adding the spatial information from the pedestrian localization to the group voice radio communication to create a dynamic 3D audio effect. This will enable the agents to hear where their colleagues are. Additionally, the 3D audio will also help the group members distinguish between each other in a conversation. Such spatial audio cues have been shown to increase speech intelligibility. Together, this will lessen some of the problems of group voice radio communication. 3D audio as a means to improve speech intelligibility and to gain tactical capabilities, has been studied before, e.g. [3–6]. *However, the main contributions of this article is the recognition of 3D audio as a means to give coarse localization information; a presentation of a complete integrated pedestrian localization, voice radio, and 3D audio system implementation; and a discussion of related tactical implications.* Despite frequent tactical usage of 3D voice radio
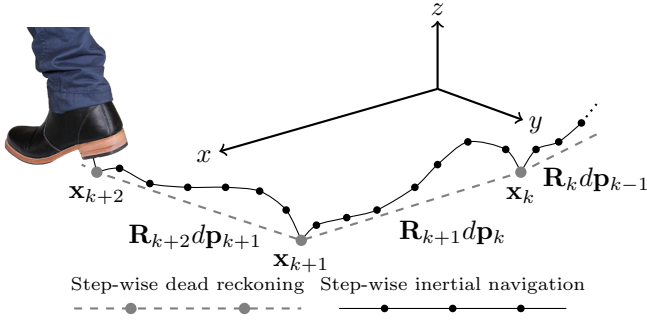
Fig. 2. Illustration of the system perspective of the foot-mounted inertial navigation. It provides displacement $d\mathbf{p}_k$ and heading changes $d\psi_k$ over individual steps. These displacement and heading changes are used to propagate the pedestrian dead reckoning, giving the pose $\mathbf{x}_k$ of an agent.

communication in virtual computer games, to our knowledge, this is the first presented system of its kind, which uses real world positioning information together with 3D audio voice radio communication.

II. PEDESTRIAN LOCALIZATION

Localization of dismounted units can be achieved from a multitude of sensor and filter technologies [7,8]. However, only a few solutions provide desirable tactical characteristics. The solution should preferably be infrastructure independent and only rely on a minimum of dynamics and environmental assumptions. At the same time, it should make do with a minimum of communication resources, while providing good relative agent pose estimates. With these requirements, in practice we are left with pedestrian dead reckoning and inter-agent ranging [9]. Recently, a few such localization solutions have appeared in the literature [9–11]. Our experimental setup relies on the one presented in [9]. Consequently, here we give a brief review of the solution and describe our implementation.

A. Localization principles

Pedestrian dead reckoning can be implemented in a number of ways [12]. However, foot-mounted inertial navigation provides a unique, high accuracy, and robust dead reckoning [13,14]. From a system perspective, the inertial navigation can be viewed as a subcomponent, which gives measurements of the displacement $d\mathbf{p}_k$ and heading change $d\psi_k$ of a step k, and related uncertainties. The state $\mathbf{x}_k$ of an agent is its position in three dimensions $\mathbf{p}_k = [x_k, y_k, z_k]$ and heading in the horizontal plane θ_k, i.e. $\mathbf{x}_k = [\mathbf{p}_k, \theta_k]$, where $[\cdot, \dots]$ is used to denote a column vector. The displacements and heading changes of individual steps can then be used to implement the dead reckoning

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{R}(\theta_{k-1})(\mathbf{u}_k + \mathbf{w}_k) \qquad (1)$$

where k is a time index, $\mathbf{u}_k = [d\mathbf{p}_k, d\theta_k]$,

$$\mathbf{R}(\theta_k) = \begin{bmatrix} \cos(\theta_k) & -\sin(\theta_k) & 0 & 0 \\ \sin(\theta_k) & \cos(\theta_k) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
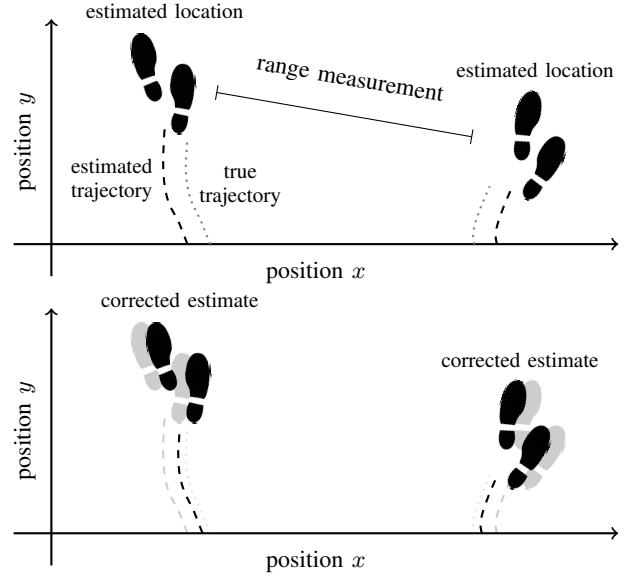

Fig. 3. Illustration of the correction of agents' positions based on an inter-agent range measurement. In the upper plot, two agents have drifted apart. This is perceived in the range measurement, which is used to correct the agents' pose estimates in the lower plot. Continuous corrections by inter-agent range measurements bound the relative errors and reduce the absolute errors.

is the rotation matrix from the agent frame to the navigation coordinate frame, and $\mathbf{w}_k$ is a (by assumption) white error of $\mathbf{u}_k$ with known covariance. The relation between the steps, the displacement and heading changes, and the dead reckoning is illustrated in Fig. 2. See [9] for further details.

The dead reckoning (1) trivially provides pose estimates $\mathbf{x}_k$ and a related uncertainty estimate $\mathbf{C}_k$ (error covariance) of all agents, relative some initial estimates. Unfortunately, measurement errors will inevitably accumulate in (1), causing unbounded error growth. Consequently, after some time, independent of the quality of the initial estimate, the pose estimates cannot be used to produce useable relative pose estimates, needed for group coordination and 3D audio. This, however, can be circumvented by introducing inter-agent ranging. The range measurements are used to continually adjust the relative positions of the agents. This will bound the relative pose estimate errors and reduce overall error growth. This correction is illustrated in Fig. 3. From a system perspective, this can be thought of as an adjustment of the global state estimate. Let

$$\mathbf{x} = [\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c, \dots]$$

denote the global state vector, where the time indices k have been dropped, and the alphabetic indices instead indicate different agents. Then, the ranging correction can be viewed as the update

$$\mathbf{x}, \mathbf{C}_x \overset{r}{\mapsto} \mathbf{x}^+, \mathbf{C}_x^+ \qquad (2)$$

where $(\cdot)^+$ denotes updated quantities and r is the range measurement. See [9] for further details. Note also, that the inter-agent ranging can be used to make the localization system completely independent of any extroceptic information or initial agent pose estimates [15].

Fig. 4. Hardware for localization and 3D audio voice radio communication. The OpenShoe units (integrated in the sole of the shoes) and the stereo headset are connected to the Android platform (tablet for the commander and smartphone for other agents), which rely on a WiFi network for communicating with the Android platforms of other agents. Cables are not displayed. The Ubisense radio tag provides synthetic inter-agent range measurements.
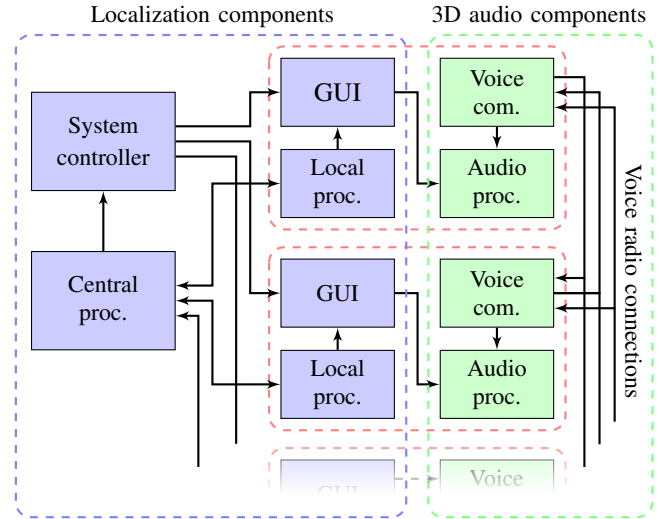

Fig. 5. Illustration of software components and their communication in a running system. Components within dashed red boxes belong to a single agent and run on his local processing platform. The common components (system controller and central processing) run on the processing platform (Android device) of one agent or potentially on a separate platform. The local processing interfaces with the sensory equipment, while the audio processing playback the voice radio to the agent.

B. Implementation

In implementing the pedestrian localization, the fundamental measurements must be acquired, (1) and (2) and the related processing performed, and the results need to be distributed. Consequently, sensory hardware, processing and communication platforms, software that implements the processing, and supportive software that controls and manage the connections and information transfer are needed. The fundamental difference between (1) and (2) is that the former only concerns a single agent, while the latter concerns more than one agent. With a naive implementation, gathering all measurements at one point, performing the processing, and distributing the results, this would make no difference. Unfortunately, this would give an unacceptable sensitivity to varying connectivity and delays that would be too large for the 3D audio. Instead, just like what is outlined in [9], (1) is implemented locally (on each agent); while (2) is implemented centrally.

The foot-mounted inertial navigation has been implemented with OpenShoe units on both feet [16]. The OpenShoe units perform the basic step-wise inertial navigation, and supply displacement and heading changes. Currently, the inter-agent ranging is provided as synthetic range measurements from a Ubisense radio localization system. However, our ambition is to replace this with in-house developed UWB-radios [17]. The processing and communication platforms are Android smartphones and tablets (Samsung Galaxy SIII and Samsung Tab 2) relying on a WiFi network for communication. The hardware for a single agent is shown in Fig. 4.

The processing and supportive softwares have been implemented as Android applications and services. The local processing, performing (1) and interfacing the OpenShoe units, is implemented as an Android service. This service provides a local controller and a graphical user interface (GUI) with pose estimates of the agent. It also forwards the basic data to a central processing server that implements (2), receiving back corrections to its own pose estimate. The central processing server is also implemented as an Android application

together with a global system controller, which distributes the results of the server, providing all agents with estimates of their colleagues. The processing has been implemented with IT++ linked with CLAPACK compiled for ARM/Android, and wrapped in a Java service. The C/C++ code is interfaced via the Java Native Interface (JNI). The system has been implemented to handle an arbitrary number of agents. In Fig. 5, the software architecture and components of a running system is illustrated. The audio part receives pose estimates of all agents, as well as information on whom to make contact with in the system, via the local controller/GUI.

III. 3D AUDIO AND VOICE RADIO

Binaural hearing allows individuals to obtain a sense of direction for the origins of sounds. Primarily, the intensity difference and the interaural time delay (delay between the ears) allow for this sense of direction; although frequency cues caused by the diffraction, reflection, and scattering from ears, pinnae, torso, and head also contribute [18]. The sense of distance is mainly created by the sound intensity, reverberation, and the direct-to-reverberant energy ratio [19]. The directional and distance cues are primarily determined by the pose of the agent and the position of the sound source. Consequently, by exploiting the pose estimates from the localization system, spatial cues can be (re-)introduced in the voice communication.

A. 3D audio principles

Directional cues can be modeled by Head Related Transfer functions (HRTF) [20,21], which are obtained by measuring (preferably in an anechoic chamber to eliminate environmental impingements) the impulse responses from a nearby sound

source, at different bearings (azimuth and elevation), to the respective ears of a subject. In practice, the measurements of individualized HRTFs are not always feasible, and it has been shown that the use of non-individualized HRTFs, obtained by measurements on a representative subject or a dummy-head [21], indeed enables fairly accurate localization of virtual sound sources at the price of somewhat increased up-down and front-back confusion [20,22].

Given a virtual sound source in the form of a sending agent b and a receiving agent a, the distance, azimuth, and elevation of the sound source relative to the receiver are

$$d_{b,a} = \|\mathbf{p}_b - \mathbf{p}_a\|$$
$$\psi_{b,a} = \arctan2(x_b - x_a, y_b - y_a) - \theta_a \quad (3)$$
$$\chi_{b,a} = \arcsin((z_b - z_a)/d_{b,a}).$$

The spatial relations are illustrated in Fig. 6. The HRTFs

$$h^{\mathrm{R}}(\psi_{b,a}, \chi_{b,a}) \quad \text{and} \quad h^{\mathrm{L}}(\psi_{b,a}, \chi_{b,a})$$

for the right (R) and left (L) ears can then be used to (re-)create 3D audio effects, which when presented through stereo head-phones worn by agent a, are perceived as originating from the direction of agent b.

Using only HRTFs for 3D auralization can give a "flat" sound that is not well externalized, i.e. the sound is perceived as originating from inside the head. By adding artificial reverberation to the signal, the sensation of externalization can be significantly improved. It appears that it is mainly the early reflections (up to about $80 - 100$ ms) that contribute to this effect [22,23]. Late reflections (>100 ms) seem to reduce the localization accuracy [22].

B. Implementation

To allow easy access to the audio streams, the voice radio communication is implemented as an Android service, inter-facing the localization subsystem as described in the previous section. The underlying voice communication engine is PJSIP, which is an open source multimedia communication library written in C. PJSIP was compiled for ARM/Android and inter-faced from Java using JNI. The audio communication between agents is done over WiFi using the G.722 speech codec and point-to-point connections. Point-to-point communication is of course inefficient and imposes limited scalability, but is adequate for initial small group (<5 agents) tests.

The 3D auralization was based on the KEMAR HRTF database [21], where artificial reverberation was added for improved externalization. A room of $4 \times 4 \times 3$[m] (width $\times$ depth $\times$ height) with a wall reflection coefficient of 0.65 was simulated using the room image method [24], where two arti-ficial receivers were positioned at $2 \pm 0.1 \times 2 \times 1.75$[m] (using a left-handed Cartesian coordinate system with the origin in the bottom left corner of the virtual room) and the source was positioned on a sphere at a distance of approximately 1[m] from the receivers. Transfer functions in the form of 512-tap Finite Impulse Response (FIR) filters, corresponding to 32 ms at a sample rate of 16 kHz, were calculated for a total of 1022 different positions on the sphere with the azimuth
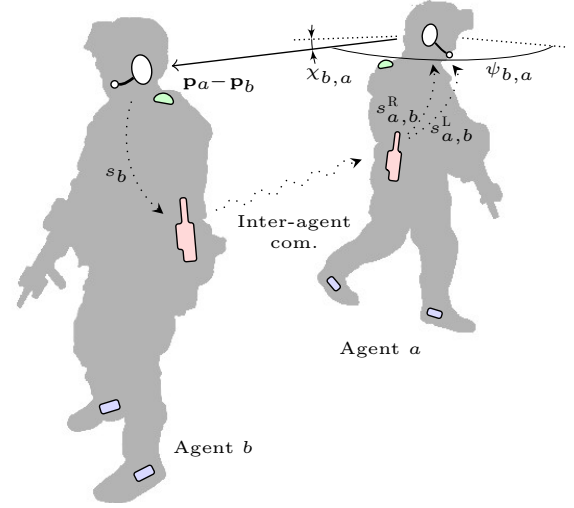

Fig. 6. Illustration of the voice radio communication and the spatial relations between agents. The 3D audio effects are created by the receiving agents, by convolving the audio stream with the appropriate FIR filters (for left and right ear); selected based on the azimuth and elevation of the sending agent.

angle varying between -180 and 180 degrees in steps of 5 degrees and the elevation angle varying between -40 to 90 degrees in steps of 10 degrees; as of that of the KEMAR HRTF database. All FIR filters were pre-calculated and stored in two tables residing in the memory of the Android device. Hence, the only real-time processing required is finding a set of filters that corresponds to the relative position in the pre-calculated tables, together with the actual filtering. With the received audio stream s_b from agent b and the table indices $i = \lfloor \psi_{a,b}/5 \rfloor$ and $j = \lfloor \chi_{a,b}/10 \rfloor$, the left and right ears of agent a are presented with

$$s_{a,b}^{\mathrm{R}} = h_{\mathrm{FIR}}^{\mathrm{R}}(i,j) * s_b$$
$$s_{a,b}^{\mathrm{L}} = h_{\mathrm{FIR}}^{\mathrm{L}}(i,j) * s_b$$

where $h_{\mathrm{FIR}}^{\mathrm{R/L}}(i,j)$ are the FIR filters for the left and right ear. The whole chain of information transfer is illustrated in Fig. 6. When changing the relative position (and corresponding FIR filters), linear cross-fading over a duration of 20 ms (the length of an audio frame) is used.

In the current implementation, only *directional* information is used, i.e. leaving out *distance* information, which in this case would be in the form of a varying direct-to-reverberant energy ratio. This is done since speech intelligibility is prioritized over positional information; naturally, it would be more difficult to communicate when the direct-to-reverberant energy ratio is low rather than high.

IV. Tactical Usage and Implications

The two direct advantages of combined 3D voice radio communication and pedestrian localization are improved voice communication and location perception of communicating group members. Improved voice communication is achieved through the use of spatial cues, aiding the so-called *cocktail party effect* [25], i.e. the ability to focus on one voice in a noisy

environment full of ongoing conversations. Applications using 3D audio in the field of tele- and videoconferencing have been proposed [26], and it has been shown that it improves memory, focal assurance, and perceived comprehension [4,5]. The tactical advantages of improved voice communication means improved coordination, group agility, and context awareness.

The tactical advantages of the ability to perceive the location of communicating group members have, to our knowledge, not been studied specifically. A study of the perception of nearby pedestrians can be found in [27]. However, some qualitative benefits, apart from the obviously improved context awareness, can be pointed out. It makes common relative (implicit or explicit) spatial specifiers such as *here* and *there*, easier to use and understand in a conversation; e.g., the phrase 'Come here!' gets a clear meaning, without knowledge of the location of the utterer. Further, in a conversation with multiple agents, a participant can directly determine wether any of the other participants may be collocated. Moreover, it gives coarse 360° X-ray vision in the way that an agent can perceive the position of another agent outside his normal field of vision, through walls, and in the dark. This could potentially help greatly with coordination in low visibility environments and in situations where an agent should not look away from a scene.

The tactical usage of 3D audio in scenarios without voice communication has previously been studied. Such systems also rely on localization and 3D audio rendering. Consequently, the presented setup gives a basis for such applications. In such cases the virtual sound source(s) can be in the form of e.g. a continuously pulsating ringing sound [28], pure tones [29], various kinds of white/colored noise [3,29,30], or synthetic/recorded speech [31] such as e.g. 'Doorway here' when approaching a doorway. The most common application is probably navigation [28,30–32], where the agent follows virtual audio cues. The combined localization and 3D audio functionality presented here makes it possible to: dynamically guide agents by providing a sound source (trajectory) on a commanding display, provide automatic acoustic homing, and allow agents to retreat along their own path. Similar applications that have also been studied are target acquisition [3] and threat alert and threat cueing [6,33]. With the presented system, agents could potentially alert each other based on their relative location to the threat.

A. Limitations

The presented system comes with a number of limitations. The most fundamental one is probably the requirement of stereo headphones. This could have the implication that the perception of the near acoustic environment is lessened, due to the sound insulating properties of the headphones. Further, the 3D audio adds a system dependence between the localization system and the voice communication system; if the positioning systems fails, the 3D audio presentation will be inaccurate. Moreover, our specific implementation comes with a number of limitations. The TCP/IP based voice communication adds time delays in the communication, which at near distances causes a mismatch between the direct audio component and
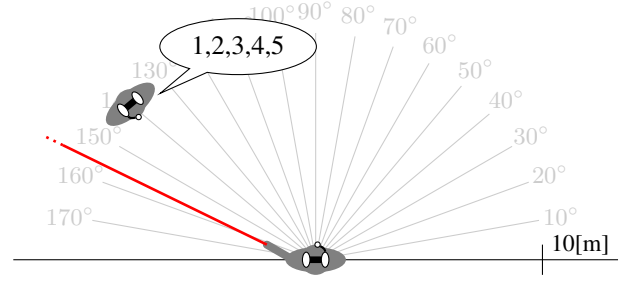

Fig. 7. Experimental setup: a test subject is pointing out, with a laser pointer, the perceived location of a test object after it has counted to five.
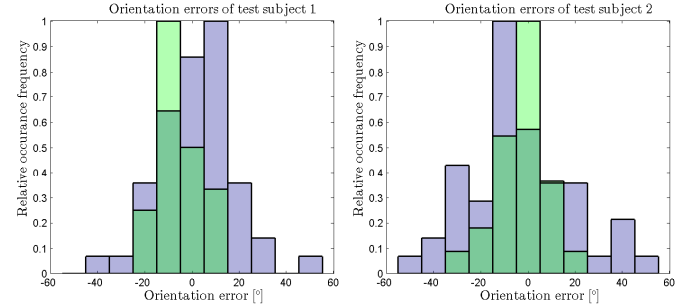

Fig. 8. Histograms of azimuth deviations between the actual and pointed location of the two different test subject. Blue and green corresponds to the results with the experimental system and with no equipment, respectively.

the component presented by the system. However, the system is intended for basic research and not operational use; and currently, efficient communication strategies is not our main focus. Finally, at present no head tracking is implemented and therefore, the audio rendering rely on the orientation of the body rather than the orientation of the head.

V. EXPERIMENT

The foundation for all tactical advantages provided by the 3D audio is the ability to perceive the direction of the source of a voice. The limitation of this ability has been studied for a long time [34]. To test the effectiveness of the presented setup, a similar experiment to that described in [34] was carried out with and without the experimental system. The experimental setup is illustrated in Fig. 7. Two test subjects were used. The blind-folded test subject stood in the KTH R1 reactor hall with a test object walking around him in a half-circle with a radius of 10[m]. The test object stopped at a uniformly distributed azimuth angle (pre-generated) relative to the test subject, after which he counted to five. Following this, the test subject pointed out the perceived direction of the voice with a laser pointer and the angle was noted down. This was repeated 50 times for each test subject, with the experimental system, and 25 times for each test subject without any equipment. The resulting orientation errors are shown in Fig. 8. To prevent any sound cues apart from those provide by the system, the headphones were replaced with in-ear type headphones (earbuds), which were covered with class II (35 dB damping) acoustic earmuffs. It can be seen that a fairly

accurate direction was perceived with the system, even though it was slightly less accurate than without any equipment. This can be explained by a number of factors such as e.g. non-individualized HRTFs, limited frequency response of the system (as mentioned previously, the G.722 speech codec was used, resulting in a bandwidth of 50–7000 Hz), and imperfect location information. Nevertheless, it was demonstrated that the system is able to provide a fairly accurate sense of direction, and a natural next step in our research is to utilize this in more elaborate experiments.

VI. CONCLUSION AND OUTLOOK

In this article, 3D audio has been used as a linking technology between voice radio communication and pedestrian localization. We have described a complete system implementation with pedestrian localization integrated with spatial audio based voice communication. The effectiveness of the setup has been demonstrated with a basic test of the perception of the source direction of an agent's voice.

There are two paths for further research: a technical path and a psychological path. For the technical part, head tracking and improved voice communication should be implemented. The head tracking is most likely implemented by adding an extra sensor to the head and including the sensor data provided into the local agent's pose processing. Further, individualized HRTFs adapted to the headphones should be used. The psychological part relates to quantitative testing of spatial awareness and coordination with and without 3D audio. The impact of the artifical reverberation on the percieved direction should be studied, as well as the possibility to use other distance cues besides the direct-to-reverberant energy ratio.

REFERENCES

[1] O. V. Prinzo and T. W. Britton, "ATC/Pilot voice communications – A survey of the literature," Federal Aviation Administration Washington DC Office of Aviation Medicine, Tech. Rep. ADA274457, Nov. 1993.

[2] R. M. Phelps, "Tactical radio communications survey – A pilot study," in *Proc. Human Factors Society – 29th Annu Mtg*, 1995.

[3] P. Andersson, O. Carlander, U. Höberg, H. Jander, M. Kindström, and S. Sandberg, "Target acquisition using free speech, bearing, and 3D-audio," FOI, Tech. Rep. FOI-R–1786–SE, 2005.

[4] O. Carlander and M. Kindström, "Intelligibility of simultaneous radio-communication," FOI, Tech. Rep. FOI-R–1525–SE, 2004.

[5] J. J. Baldis, "Effects of spatial audio on memory, comprehension, and preference during desktop conferences," in *Proc. of the SIGCHI Conf. on Human Factors in Computing systems*. ACM, 2001, pp. 166–173.

[6] P.-A. Oskarsson, L. Eriksson, and O. Carlander, "Enhanced perception and performance by multimodal threat cueing in simulated combat vehicle," *Human Factors*, vol. 54, no. 1, pp. 122–137, 2012.

[7] J. Rantakokko, J. Rydell, P. Strömbäck, P. Händel, J. Callmer, D. Törnqvist, F. Gustafsson, M. Jobs, and M. Grudén, "Accurate and reliable soldier and first responder indoor positioning: multisensor systems and cooperative localization," *IEEE Wireless Communication*, vol. 18, pp. 10 –18, april 2011.

[8] C. Fuchs, N. Aschenbruck, P. Martini, and M. Wieneke, "Indoor tracking for mission critical scenarios: a survey," *PMC*, vol. 7, pp. 1 – 15, 2011.

[9] J.-O. Nilsson, D. Zachariah, I. Skog, and P. Händel, "Cooperative localization by dual foot-mounted inertial sensors and inter-agent ranging," *CoRR*, vol. abs/1304.3663, 2013.

[10] W. Hawkinson, P. Samanant, R. McCroskey, R. Ingvalson, A. Kulkarni, L. Haas, and B. English, "GLANSER: Geospatial Location, Accountability, and Navigation System for Emergency Responders - system concept and performance assessment," in *Proc. IEEE/ION PLANS*, Myrtle Beach, SC, USA, 24-26 Apr. 2012.

[11] P. Strömbäck, J. Rantakokko, S.-L. Wirkander, M. Alexandersson, K. Fors, I. Skog, and P. Händel, "Foot-mounted inertial navigation and cooperative sensor fusion for indoor positioning," in *ION ITM*, San Diego, CA, USA, 25-27 Jan. 2010.

[12] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," *IEEE Commun. Surv. Tutor.*, vol. PP, no. 99, pp. 1–13, 2013.

[13] M. Laverne, M. George, D. Lord, A. Kelly, and T. Mukherjee, "Experimental validation of foot to foot range measurements in pedestrian tracking," in *ION GNSS*, Portland, OR, USA, 19-23 Sept. 2011.

[14] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *IEEE Comput. Graph. Appl.*, vol. 25, pp. 38 –46, 2005.

[15] J.-O. Nilsson and P. Händel, "Recursive Bayesian initialization of cooperative localization based on ranging and dead reckoning," in *Proc. IEEE/RSJ IROS*, Tokyo, Japan, 3-7 Nov. 2013.

[16] J.-O. Nilsson, I. Skog, P. Händel, and K. Hari, "Foot-mounted inertial navigation for everybody – An open-source embedded implementation," in *Proc. IEEE/ION PLANS*, Myrtle Beach, SC, USA, 23-26 Apr. 2012.

[17] A. De Angelis, S. Dwivedi, and P. Händel, "Characterization of a flexible UWB sensor for indoor localization," *Instrumentation and Measurement, IEEE Transactions on*, vol. 62, no. 5, pp. 905–913, 2013.

[18] D. R. Begault *et al.*, *3-D sound for virtual reality and multimedia*. Ap Professional Boston etc, 1994, vol. 955.

[19] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio," *J. Acoust. Soc. Am.*, vol. 124, pp. 450–461, 2008.

[20] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, pp. 111–123, 1993.

[21] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am.*, vol. 97, pp. 3907–3908, 1995.

[22] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.*, vol. 49, no. 10, pp. 904–916, 2001.

[23] F. Völk, "Externalization in data-based binaural synthesis: effects of impulse response length," in *Proc. of Intern. Conf. on Acoustics NAG/DAGA*, Rotterdam, Netherlands, 23-26 Mar. 2009.

[24] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, 1979.

[25] A. R. Conway, N. Cowan, and M. F. Bunting, "The cocktail party phenomenon revisited: The importance of working memory capacity," *Psychonomic Bulletin & Review*, vol. 8, no. 2, pp. 331–335, 2001.

[26] S. L. Gay and J. Benesty, *Acoustic signal processing for telecommunication*. Kluwer Academic, 2000.

[27] M. Ardavan and F. Chen, "Listen! Somebody is walking towards your car (Introducing the awareness-3D sound system into the driver to increase the pedestrian's safety)," in *Proc. HCI (10)'11*, 2011, pp. 89–98.

[28] L. Eriksson, A. Berglund, B. Willén, J. Svensson, M. Pettersted, O. Carlander, B. Lindahl, and G. Allerbo, "On visual, vibrotactile, and 3d audio directional cues for dismounted soldier waypoint navigation," in *Proc. Human Factors and Ergonomics Society, 52nd Annu Mtg*, New York, NY, USA, 22-26 Sept. 2008.

[29] B. N. Walker and J. Lindsay, "Effect of beacon sounds on navigation performance in a virtual reality environment," in *Proc. ICAD, Boston, MA, USA*, 2003.

[30] B. D. Simpson, D. S. Brungart, R. H. Gilkey, and R. L. McKinley, "Spatial audio displays for improving safety and enhancing situation awareness in general aviation environments," in *New Directions for Improving Audio Effectiveness, Mtg Proc.*, Neuilly-sur-Seine, France, 11-13 Apr. 2005.

[31] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, "Navigation system for the blind: Auditory display modes and guidance," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 2, pp. 193–203, 1998.

[32] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert, "Swan: System for wearable audio navigation," in *Wearable Computers, 11th IEEE International Symposium on*, 2007, pp. 91–98.

[33] O. Carlander and L. Eriksson, "Uni- and bimodal threat cueing with vibrotactile and 3D audio technologies in a combat vehicle," in *Proc. Human Factors and Ergonomics Society 50nd Annu Mtg*, San Francisco, CA, USA, 16-20 Oct. 2006.

[34] L. Rayleigh, "Our perception of the direction of a source of sound," *Nature*, vol. 14, pp. 32–33, 1876.