A Wearable System for the Wireless Experience of Extended Range Telepresence

Ferdinand Packi, Antonia Pérez Arias, Frederik Beutler, and Uwe D. Hanebeck

Abstract— Extended range telepresence aims at enabling a user to experience virtual or remote environments, taking his own body movements as an input to define walking speed and viewing direction. Therefore, localization and tracking of the user's pose (position and orientation) is necessary to perform a body-centered scene rendering. Visual and acoustic feedback is provided to the user by a head mounted display (HMD). To allow for free movement within the user environment, the tracking system is supposed to be user-wearable and entirely wireless. Consequently, a lightweight design is presented featuring small dimensions to fit into a conventional 13" laptop backpack, which satisfies the above stated demands for highly immersive extended range telepresence scenarios. Dedicated embedded hardware combined with off-the-shelf components is employed to form a robust, low-cost telepresence system that can be easily installed in any living room.

I. INTRODUCTION

In a telepresence scenario the user experiences real and virtual environments, which are commonly presented to him on a head mounted display (HMD). At the Intelligent Sensor-Actuator-Systems (ISAS) lab at the Karlsruhe Institute of Technology (KIT) an experimental setup of $5\,\mathrm{m} \cdot 5\,\mathrm{m} \cdot 2,5\,\mathrm{m}$ has been established to form a telepresence environment. Figure 1 shows a user wearing the mobile components that are mounted onto a wearable backplane.

A special feature of the current system is the integrated motion compression [1], which allows users to explore extended range target environments given only limited user space. The basic principle is to transform local motion one-to-one into a target environment of variable dimension. Distances covered in the target environment are projected onto circular arcs in the user environment, while preserving lengths. Target environments are not necessarily virtual worlds, but can as well be remote environments explored by a mobile teleoperator such as the Omnibase [2]. In the latter case, the user controls the teleoperator's motion and receives its camera images directly projected onto the HMD screens. Either way, the user is put in place to explore target environments as visual and acoustic feedback are provided via an HMD. Gaming scenarios like Quake™ and Pacman™ clones have been implemented (as presented in [3]) as well as virtual museum visits. Essential for the performance of a telepresence environment is the underlying tracking system, as it feeds movements of the user's head or hand continuously into the visualization engine. A variety of methods ranging from optical, acoustic, electromagnetic, and mechanical setups are available to accomplish the task in a more or less satisfying way. Every approach has its own specific drawbacks and benefits concerning robustness, achievable precision, hardware requirements, or the way of dealing with occlusions. As all of them measure physical phenomena, they are naturally subject to disturbances and noise. The concept to obtain pose estimation is to measure time-of-flight (TOF) of acoustic signals between a set of stationary speakers and user worn microphones in the "just above audible" spectrum.
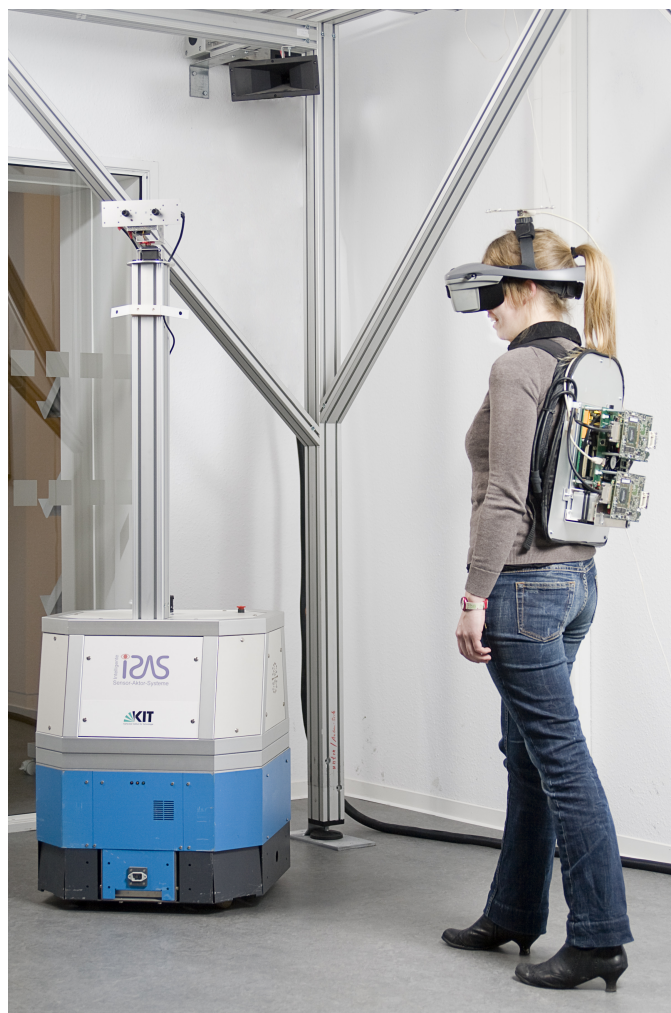

Fig. 1. Telepresence environment at the ISAS lab: Teleoperator "Omnibase" and user wearing the tracking system. Sound signals travel from 4 speakers (mounted in the corners of the room) to 4 microphones (mounted on top of the HMD).

F. Packi, A. Perez, F. Beutler and U. D. Hanebeck are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany `packi@kit.edu, perez@kit.edu, beutler@ieee.org, uwe.hanebeck@ieee.org`

A. Requirements

To allow for a deep feeling of immersion into the remote/virtual environment, the wearable system needs to operate entirely wireless. Power supply, control and data cables must be avoided to ensure the highest possible degree of ergonomics and to increase safety, which encourages especially unexperienced users. Common wired designs use the same hardware to emit and receive packets of audio signals, so the time-of-flight can be easily retrieved by subtracting timestamps of sending and receiving time. The challenge that evolves in wireless setups is therefore to provide an accurate synchronization between signal generator and receiving units. As multiple sources emit signal sequences, a method of sharing the common medium must be chosen. Another requirement is a lightweight and small outline design that's also available at low cost. More technical requirements are a working volume the size of a usual living room, capability of locating extended objects in 6 degrees of freedom (DOF), update rate of at least 20 Hz, and the possibilty to track multiple objects, such as a user's head and hand. For position accuracies, < 1 cm is desirable.

B. Related work

In [4], the principal idea of stationary transmitters placed in the corners of the ceiling and body worn receivers is introduced, as well as the use of spread spectrum technologies, but aimed at radio frequency tracking systems. Among the acoustic tracking systems, a special regard is payed to Whisper [5], where several aspects of the present design are anticipated, such as the spread spectrum methods for CDMA or the considerations about occlusion. More precisely, direct sequence spread spectrum (DSSS) was implemented in order to allow for a unique identification of emitters. The bandspreaded signal lies within the audible spectrum. The advantage of lower frequencies is their ability to diffract around occluding objects to a certain extent. Approaches in wireless acoustic tracking for sensor networks have been made in [6]. The synchronization problem is dealt with using reference-broadcast synchronization (RBS). In [7], a 3 DOF acoustic localization system is presented that processes TOF-measurements of narrowband transmitting signals in the range between 25 kHz and 40 kHz using transducers. For signal processing, a Texas Instruments TMS320™ DSP is deployed. Position estimation is then achieved by trilateration. Due to the high frequencies, losing line-of-sight leads to signal loss. The general idea of using TOF between arrays of signal emitters and receivers was also used in Constellation© [8]. Here, 40 kHz ultrasonic transducers emit coded signals to identify the different sources. Additionally, an intertial measurement unit (IMU) is used, involving state estimation techniques. Noticeable is also the previous design of the acoustic tracking system currently installed within the ISAS lab [9], which represents a wired 6 DOF acoustic tracking system realized in standard hardware.

C. Main Contributions

Innovations within the present development of an acoustic tracking system are significant with respect to ergonomics and user-friendliness. The whole design is totally wireless: The mobile tracking unit runs on battery for at least 1 hour. It is controlled by WLAN, whereas the time-critical position and orientation values are sent via embedded radio modules (XBEE). A major effort in a wireless realization is to keep sending and receiving units closely synchronized, which is achieved by periodic emission of synchronization pulses. Combined use of standard and embedded components keeps the balance between computation power and efficiency. The tracking unit is easily extensible by expansion slots and several communication interfaces. DVI, VGA, and Displayport outputs are available to directly connect to various HMDs, whereas video control units are mounted in the actual setup to support high resolution SXGA microdisplays used in NVIS HMDs. The overall concept is a user-wearable, low-cost tracking device for telepresence environments, which fits in a regular 13" Laptop backpack and can easily be installed in almost all surroundings.

II. Hardware Design

Great emphasis was placed on designing the tracking system's hardware using state-of-the-art dedicated components. Optimal signal treatment throughout the whole audio chain is assured by low-noise electronic parts and adequate printed circuit board (PCB) layout considerations.
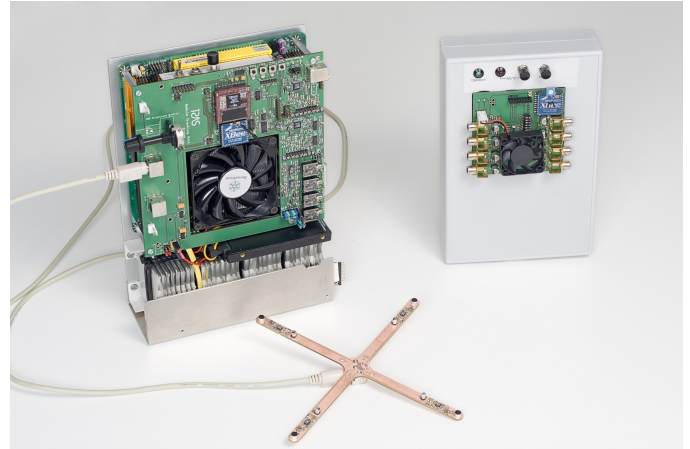

Fig. 2. Components of the wireless tracking system: Tracking unit, microphone array, and signal generator unit.

The tracking system (see Figure 2) is a combination of embedded and regular desktop components. Distributed and loosely coupled, they constitute two principal units: a stationary unit for signal generation and amplification, and a mobile unit for signal recording and processing.

The design is modular and extensible. Both amplifier and tracking unit are supported by Analog Devices Blackfin™ DSPs.

A. Signal Generator

Based on the Analog Devices EZ-Kit$^{\text{TM}}$ development board, an expansion daughterboard was designed featuring two 4-channel digital amplifiers TAS5704 by Texas Instruments. Together, this yields a full digital, programmable 8-channel amplifier unit with 8x oversampling at 48 kHz data (384 kHz switching rate). It delivers 10 W output power per channel. The mode of operation is as follows: A set of characteristic signal sequences is generated according to the chosen spread spectrum method and stored in a lookup table within the DSP. These sequences are then concurrently cycled into the amplifier chips using the digital I^{2}S bus, synchronously to the preset update rate. Piezo-based tweeters emit the signals to be captured by the body-worn microphones.

B. Microphone Array

4 omnidirectional microphones are placed on a PCB carrier, where preamplification is directly performed to prevent loss in signal quality. The parts supplied feature constant frequency response up to 24 kHz and high S/N-ratio thanks to low-noise operational amplifiers.

C. Mobile Tracking Unit

The mobile tracking unit comprises 8 (expandable to 16) input channels sampling at 96 kHz. Hence, in the current configuration two microphone carriers can be plugged in, enabling concurrent head and hand tracking, for instance. The input signals are preconditioned to fit the full range of the 24 bit ADC contained in the audio codecs of type AD1938 by Analog Devices. The samples are then concurrently transferred into the DSP-Stamp of type Blackfin$^{\text{TM}}$ BF533 inside a TDM data frame for further processing. Figure 3 illustrates the different areas on the PCB and their functions.

Underlying, a mini-ITX form factor computer running Ubuntu Linux is attached to the mobile tracking unit, concerned with generating and rendering scenes to be visualized for the user on the HMD (see Figure 4). The deployed HMD by NVIS requires a special video format to be fed to the microdisplays, therefore the mobile tracking unit carries on top two video control units, one for each eye, to enable high resolution stereo vision. A 32-cell lithium-polymer battery concludes the wireless mobile design, powering all components for at least one hour. The 14.8 V, 8 Ah battery pack is organized in 4 independently chargeable segments, allowing flexibility in recharging either serially or in parallel. A spare battery pack can be easily installed, and mains operation can be switched in during operation.

All components were chosen according to low power considerations: SSD memory, onboard AMD780 graphics, AMD4850e CPU in cool'n'quiet$^{©}$ mode, and a picoPSU$^{\text{TM}}$ DC-DC converter contribute to longer operating times and little cooling efforts.

The total dimensions of the mobile tracking unit including the NVIS video control units are $25\,\text{cm} \cdot 18\,\text{cm} \cdot 11\,\text{cm}$ at a total weight of $\approx 4,2\,\text{kg}$.
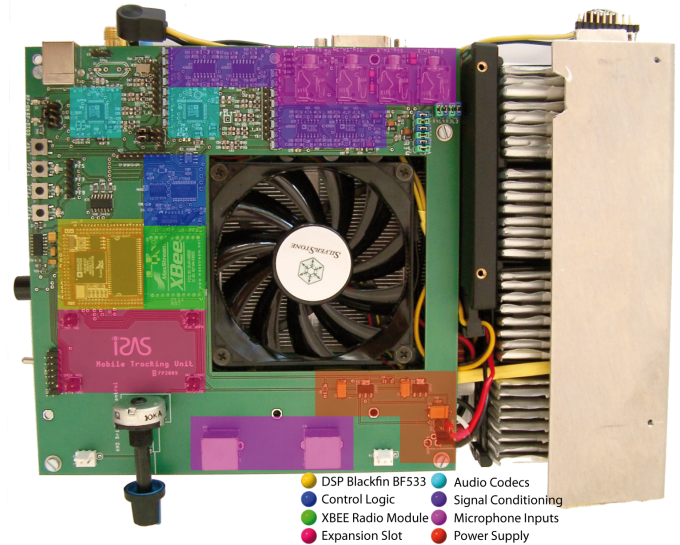

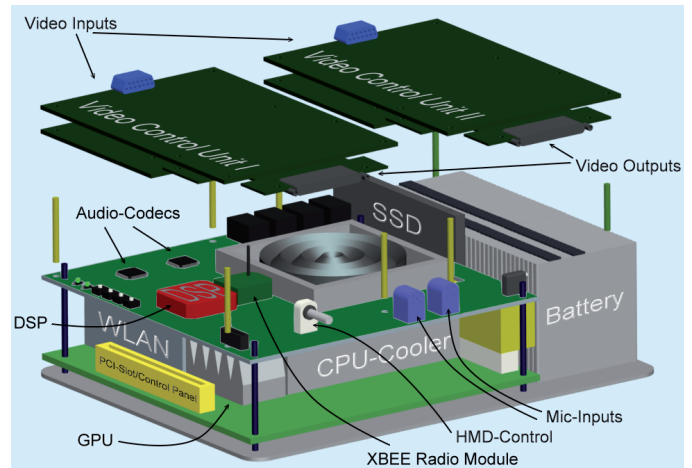Fig. 3. Overview of the mobile tracking unit's PCB and the placement of the components used.

Fig. 4. 3D model of the mobile tracking unit. The video control units can be detached if not needed (only required for NVIS HMDs).

D. Communication

All units involved in the wireless tracking system are interconnected for the exchange of operational data or for control purposes as illustrated in Figure 5. High-level interaction between stationary computer systems (such as the motion compression server) and the mobile tracking unit is performed over WLAN. Time-critical data transfers like pose estimates use XBEE radio modules, which provide lower latency. Internal communication between the mobile tracking unit's components utilizes the built-in USB/RS232 interface.

III. WIRELESS ACOUSTIC TRACKING

In the present setup, tracking of extended objects is based on distance measurements between stationary signal sources (loudspeakers) and receivers (microphones) utilizing the signal's TOF. The mode of operation is to estimate a
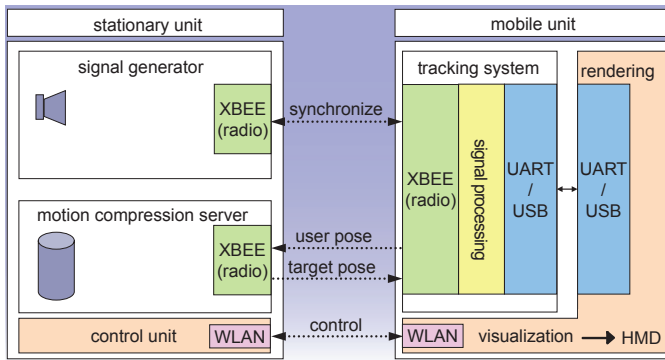
Fig. 5. Communication between distributed components of the telepresence system. The stationary unit consists of the sound signal generator and a server to transform posture information from user environment into target environment. The mobile unit includes sound signal receiving and processing as well as generating the video data to be passed to the HMD

unique pose from these distance measurements given the known loudspeaker positions in world coordinates as well as the geometry of the user-worn microphone array. Pose estimation of extended objects in three-dimensional space requires a configuration of at least 3 loudspeakers and 3 microphones to uniquely define a pose, otherwise ambiguities remain. On the other hand, overdetermined systems with redundant speakers/microphones can lead to more accuracy and fault tolerance. In the proposed design, the user's pose is estimated from distance measurements using a closed-form range-based pose estimation algorithm presented in [10]. As several signal sources share the same medium (often referred to as channel), modulation is necessary. Spread spectrum techniques are well suited to increase diversity and noise immunity. In addition to the tasks needed for conducting conventional acoustic tracking, wireless distance measurements require accurate synchronizing between emitter and receiver units, as TOF estimation within the mobile tracking unit needs exact knowledge about the signals' sending timestamp.

A. Spread Spectrum Techniques

As the permitted bandwidth for audio signals is limited, a narrowband channel partitioning like FDM is not recommended for the purpose of concurrently emitting signal sequences from different sources. Simple sine waves are also easily corrupted by noise and yield little discrimination in the matched filter.

A more convenient approach is the use of Code Division Multiple Access (CDMA), where a choice has to be made among the various methods: Time hopping, frequency hopping, direct sequence, or multi-carrier spread spectrum. As the signals shall remain inaudible, the frequency hopping approach appeared to be optimally suited. Adjusted to a slow chipping rate, the wanted signal "jumps" between several carrier frequencies according to a pseudonoise code, exhibiting the unique transmitting signal.

B. Synchronization

A robust method of keeping the distributed components' time-base accurate is presented in the following. The first

step is to equalize the DSP's operating frequencies, which can be managed by setting the appropriate prescalers and dividers within the Blackfin$^{\text{TM}}$ processors. In our example they are both set to run at 675 MHz core clock rate. Derived from this rate is the update cycle rate to perform the distance measurements. Unfortunately, the clocks are subject to drift and run off from each other. Therefore we define a clock master (in our case the digital amplifier unit) and one or more clock slaves (the tracking unit), and emit periodic synchronization pulses to adjust the slaves' time base. By estimating the latency within the transmission over radio, the time bases of all nodes involved can be aligned. Smoothing the incoming pulses avoids leaps in time bases. Using cyclic timers (as illustrated in Figure 6), failing pulses are automatically compensated, ensuring that remote and local times are always kept synchronous.
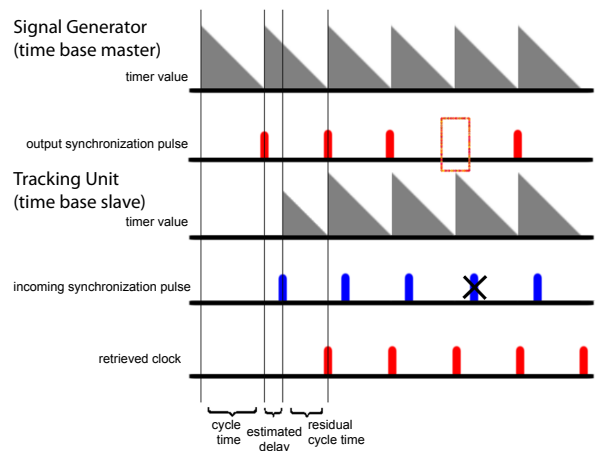

Fig. 6. Scheme of Synchronization - The incoming synchronization pulse sets the receiving timer value according to the difference of full cycle time and estimated system-dependent delay. The original timestamp is retrieved as the cyclic countdown timer reaches zero (reset). The method is tolerant to missing sending pulses as they are compensated by the underlying timer activity.

C. Distance Measurements

A method to measure distances is to emit a characteristic signal sequence of a certain length at a known time, and to count the time elapsed until the signal can be detected within a receiving unit. Given that both emitter and receiver are synchronous, this can be done by sampling input values over a window length adequate to the maximum expected time (constrained by the room dimensions) and afterwards correlating this input buffer with all the (reversed) signal sequences registered. The maximum correlation-value yields the time passed between sending and receiving timestamp in terms of sample periods. Dividing by the sampling rate (in the present design 96 kHz) and multiplying by the speed of sound yields the required distance from one distinct speaker to the evaluated microphone input buffer. A means to ameliorate the results is passband filtering of the input signal as well as outlier detection, after all the channel can be

exposed to broadband ambient noise or massive narrowband perturbations.

Pose estimation can then be performed using the closed-form algorithm discussed in [10].

IV. EXPERIMENTAL RESULTS

A. Synchronization

The core clocks of the digital amplifier unit (clock master) and the mobile tracking unit (clock slave) are synchronized by periodically emitted radio signals and running at 675 MHz. If the synchronization pulse is stalled, the drift (as seen in Figure 7) increases without limit with a mean value of $\approx 0,08$ ppm/s at room temperature.

In normal operation the synchronization signal is sent every 100 ms by the clock master. The latency of the pulse arrival is subject to variability with a standard deviation of $\sigma = 70$ µs as shown in Figure 8. To prevent fluctuations in distance measurements, smoothing needs to be applied over the incoming synchronization pulses.
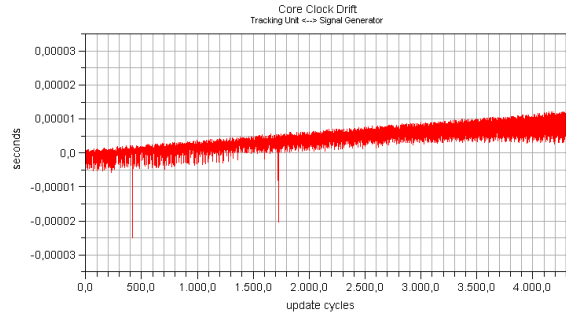

Fig. 7. Drift between signal generator (clock master) and mobile tracking unit (clock slave).
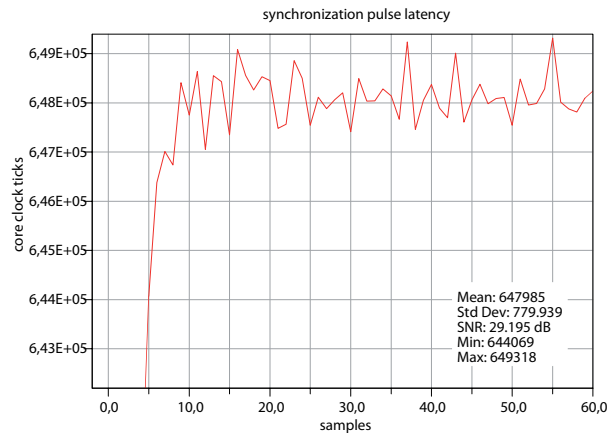

Fig. 8. Transient oscillation of received synchronization pulses. The 5 first values are due to initial tuning of the synchronization routine.

B. Audio signal processing

At frequencies above 20 kHz the loudspeakers suffer from higher directivity, which results in weaker amplitudes captured by the microphones, if they are placed off the main emitting direction. Figure 9 shows a constellation, where

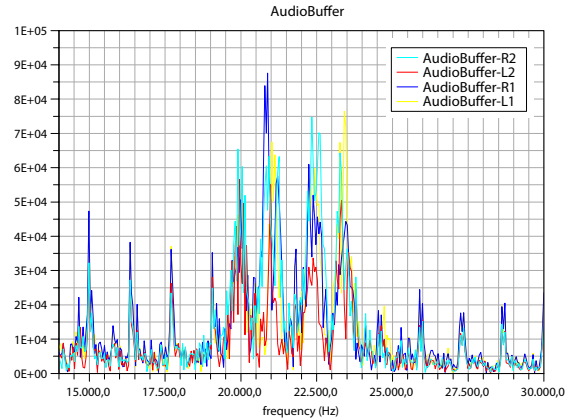the desired signals are only minimally stronger than the undesired noise.

Fig. 9. Spectrum snapshot of 4 overlayed input audio buffers after a full window of 2000 samples has been recorded. The signals were emitted by 4 speakers within the inaudible spectrum between 20 kHz and 24 kHz.

C. Distance measurements

In Figure 10 a static distance measurement was recorded over 40 update cycles. Standard deviation in quiet surroundings without obstacles is 2 cm. In the presence of ambient noise, outliers may occur, which can be handled by filtering (median filtering with variable step count is implemented). It is obvious, that deviations within the imcoming synchronization pulse latency strike through to the distance measurements, and thus to the position and orientation estimates. Therefore, smoothing of the synchronization timestamps can lead to significantly improved range and posture estimates.
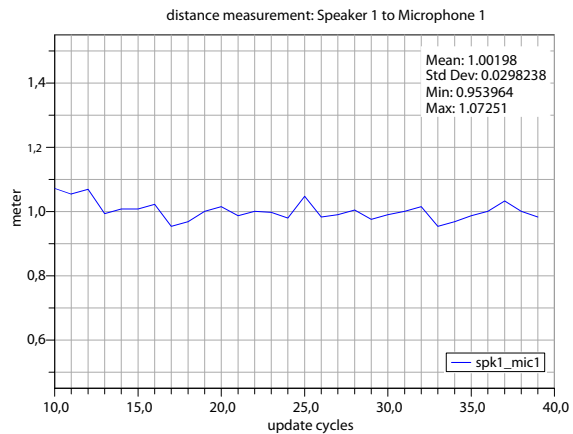

Fig. 10. Distance measurement in a static scene (unfiltered).

Figure 11 shows static pose estimation recorded over 30 seconds (300 cycles at 10 Hz update rate). Again, the first 5 values are influenced by the initial tuning process of the synchronization routine. Figure 12 shows a closeup with statistics.
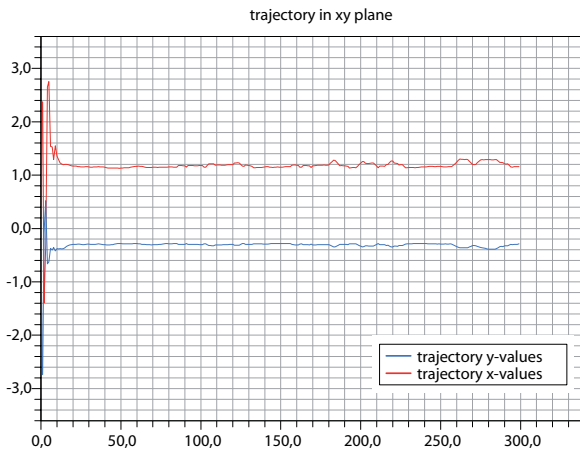
Fig. 11. Static position measurement in the xy-plane measured over 30 seconds (unfiltered).
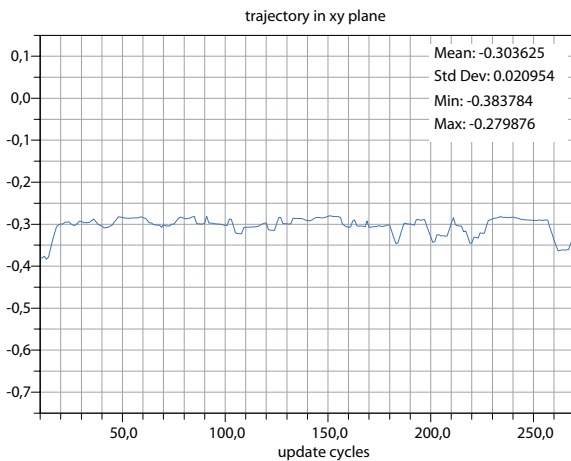

Fig. 12. Closeup to static position x-value measurement in the xy-plane measured over 30 seconds (unfiltered) included statistics.

V. CONCLUSION

A wireless realization of an embedded acoustic tracking system has been presented, which is suitable for indoor tracking tasks like extended range telepresence scenarios. It enables users to freely move in extended telepresence environments, wearing only a lightweight mobile tracking unit less the size of a shoe box.

The underlying hardware features a fully digital signal generation and amplification, avoiding loss in signal quality encountered in most common assemblies. On the recording side, multi channel processing is available at high sampling rates. Multiple communication interfaces yield easy expandability. Thanks to signal propagation in the inaudible spectrum, noise pollution is minimized.

The overall assembly to be installed comprises merely of the mobile tracking unit, a set of speakers and the digital amplifier unit to support those speakers, thus it embodies an easy-to-install, low cost telepresence system to be set up within minutes in any mid-size room.

Future work will combine the acoustic tracking with inertial measurement units in order to increase update rates required e.g. for fast changes in heading. With the update rate of 10 Hz to 20 Hz achieved so far, quick head movements would tend to overshoot. Another approach for optimization is the limited computation power of the DSP, which reaches its limits on the hard task of evaluating 8 or 16 audio channels. Integration of several parallel DSPs (in grids of 2 or 4 units) could help to reduce workload, increase the update rate and allow for a higher number of channels.

State estimation could be applied to deal with underdetermined configurations (less than the required number of speakers/microphones), unknown speaker locations, or unknown sending times.

REFERENCES

[1] P. Rößler and U. D. Hanebeck, "Simultaneous Motion Compression for Multi–User Extended Range Telepresence," in *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, Beijing, China, Oct. 2006, pp. 5189–5194.

[2] P. Rößler, F. Beutler, U. D. Hanebeck, and N. Nitzsche, "Motion Compression Applied to Guidance of a Mobile Teleoperator," in *Proceedings of the 2005 IEEE International Conference on Intelligent Robots and Systems (IROS 2005)*, Edmonton, Canada, Aug. 2005, pp. 2495–2500.

[3] P. Rößler, F. Beutler, and U. D. Hanebeck, "A Framework for Telepresent Game-Play in Large Virtual Environments," in *Proceedings of the 2nd International Conference on Informatics in Control, Automation and Robotics (ICINCO 2005)*, vol. 3, Barcelona, Spain, Sept. 2005, pp. 150–155.

[4] S. R. Bible, M. Zyda, and D. Brutzman, "Using spread-spectrum ranging techniques for position tracking in a virtual environment," in *Proceedings of the IEEE Conference on Networked Realities*, Boston, MA, 1995.

[5] N. M. Vallidis, "WHISPER: A spread spectrum approach to occlusion in acoustic tracking," Ph.D. dissertation, University of North Carolina, Chapel Hill, 2002.

[6] Q. Wang, W.-P. Chen, R. Zheng, K. Lee, and L. Sha, *Scalable and Low-Cost Acoustic Source Localization for Wireless Sensor Networks*. Department of Computer Science University of Illinois at Urbana-Champaign Urbana, IL 61801, 2006, ch. Track 4: Sensor Networks, pp. 517–526.

[7] I. Karaseitanidis and A. Amditis, *A Novel Acoustic Tracking System for Virtual Reality Systems*. Institute of Computer and Communications Systems, Athens, Greece, 2008, ch. 1, pp. 99–122.

[8] E. Foxlin, M. Harrington, and G. Pfeifer, "Constellation: A wide-range wireless motiontracking system for augmented reality and virtual set applications," 1998.

[9] F. Beutler and U. D. Hanebeck, "The Probabilistic Instantaneous Matching Algorithm," in *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Heidelberg, Germany, Sept. 2006, pp. 311–316.

[10] ——, "Closed-Form Range-Based Posture Estimation Based on Decoupling Translation and Orientation," in *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005)*, vol. 4, Philadelphia, Pennsylvania, Mar. 2005, pp. 989–992.