Assessment of Binaural–Proprioceptive Interaction in Human-Machine Interfaces

M. Stamm and M. E. Altinsoy

1 Introduction

The human auditory system is capable of extracting different types of information from acoustic waves. One type of information which can be assessed from acoustic waves that impinge upon the ears, is the perceived direction of the sound source. For more than 100 years, scientists investigate how the auditory system determines this direction, that is, how we localize an acoustic event. The first knowledge about the auditory localization process stemmed from listening tests and was used to develop localization models trying to mimic human hearing. Pioneering modeling work was done by Jeffress in 1948 and Blauert in 1969. Jeffress proposed a lateralization model [22], which uses interaural differences to explain the localization in lateral direction within the horizontal plane. Blauert proposed a model analyzing monaural cues to explain human localization in the vertical direction within the median plane [5]. Based on these approaches, new and further developed binaural models have been implemented and published over the years—see [8] for a more detailed overview. These models did not only help to understand human auditory localization, they have also been used as a basic requirement for different technical applications, for example, hearing aids, aural virtual environments, assessment of product-sound quality, room acoustics and acoustic surveillance [7].

With the help of binaural models, human localization can be predicted under the assumption that a corresponding localization process is based on acoustic signals, thus, on unimodal information. However, what happens if this localization process is realized in an environment with available bimodal or even multimodal sensory input? This is an important question, because we are encountering such situations in our daily life. For example, we normally look and listen from which direction and at which distance a car is approaching to estimate whether it is safe to cross the street—

M. Stamm · M. E. Altinsoy (✉)
Chair of Communication Acoustics, Technische Universität Dresden,
Dresden, Germany
e-mail: ercan.altinsoy@tu-dresden.de

a case of audio–visual interaction. In close physical proximity to the human body, a further perceptual sense, the haptic sense, becomes important. Everyone has already tried to localize the position of the ringing alarm clock in the morning by listening and simultaneously grabbing after it, that is, audio–proprioceptive interaction. But what role does auditory information play in such situations? Do we consider the auditory modality in the localization process or is it dominated mainly by vision or haptics? These questions have to be investigated within suitable experiments to gain a deeper insight into human intermodal integration. Thus, such experiments also help to understand whether and to which extent binaural models allow to predict human localization in bimodal or multimodal scenes. The gained knowledge could play an important role for further technical developments.

1.1 Binaural–Visual Localization

Many studies have been conducted in the field of audio-visual interaction and perceptual integration. For example, Alais et al. [1] revealed that vision only dominates and captures sound when visual localization functions well. As soon as visual stimuli are blurred, hearing dominates vision. If the visual stimuli are slightly blurred, neither vision nor audition dominates. In this case, the event is localized by determining a mean position based on an intermodal estimation that is more accurate than the corresponding unimodal estimates [1]. Thus, the weighting factor of the binaural signals is chosen indirectly proportional to the perceiving person's confidence in visual input. However, the weighting factor does not only depend on the quality of visual information but also on the corresponding task. For example, if the localization process requires sensitivity to strong temporal variations, binaural sensory input dominates the localization. This is because of the superior temporal resolution of the auditory system over vision [38].

In conclusion, binaural models are definitely required for mimicking human localization even when visual information is available. The sensory input of the auditory and visual modality is integrated in an optimal manner and, thus, the resulting localization error is minimized. These insights can be used for mimicking human audio-visual localization computationally with the help of existing binaural modeling algorithms, for example, to improve technical speaker tracking systems. Equipping such systems with a human-like configuration of acoustical and optical sensors, which feed binaural and visual tracking algorithms, is advantageous because of two reasons. First, the simultaneous analysis of auditory and visual information is superior than the analysis of unimodal information alone. When auditory tracking is weak due to acoustic background noise or acoustic reverberation, the visual modality may contribute to a reliable estimation. The converse is also true. When visual tracking is not reliable due to visual occlusions or varying illumination conditions, the auditory modality may compensate for this confusion. Thus, the integration of auditory and visual information increases the robustness of tracking accuracy. Second, a human-like configuration of sensors means using only two microphones and two

cameras simultaneously. Such a configuration is quite effective, because applying only four sensors corresponds to the minimum requirement for spatial auditory and spatial visual localization. Furthermore, such a configuration enables sufficiently high tracking accuracy as we know from our daily life. An approach concerning the application of a human-like configuration of sensors and the computational integration of bimodal sensory input for multiple speaker localization with a robot head is described in [25]. Further details on binaural systems in robotics can be found in [2], this volume.

1.2 Binaural–Proprioceptive Localization

There is a further human sense which is capable of localization, namely, the haptic sense. Within the haptic modality, localization is realized by tactile perception and proprioception. Tactile perception enables the localization of a stimulated position on the skin surface. In this chapter, the focus is on proprioception which enables to determine the absolute position of the own arm and to track its movements. Proprioception is realized unconsciously with the help of the body's own sensory signals that are provided by *cutaneous mechanoreceptors* in the skin, by *muscle spindles* in the muscles, by *golgi tendon organs* in the tendons, and by *golgi and ruffini endings* in the joints of the arm.

As mentioned before, binaural models are required for modeling human localization even when visual information is available. However, what happens if the localization process is realized with proprioceptive and auditory signals? Is the auditory modality considered in the localization process or is it dominated mainly by proprioception, that is, by evaluating the body's own sensory signals about the upper limb position? These questions are not only of theoretical interest. If the auditory modality would be considered and could probably even guide the localization process, practical relevance exists especially for virtual environments in which the reproduction of acoustic signals can be directly controlled. Computational models mimicking human binaural-proprioceptive localization could then be used to simulate how an audio reproduction system has to be designed to optimize localization accuracy within a specific workspace size directly in front of the human body. In addition, such simulations may help to reproduce suitable auditory signals to diminish systematically oriented errors in proprioceptive space perception, for example, the radial-tangential illusion [26], and thus to further sharpen human precision. Two questions arise in this context.

- Which applications in virtual environments depend on proprioception?
- Should proprioceptive localization be improved with additional auditory signals?

Proprioceptive localization performance plays an important role for all applications in which a haptic device is involved. Such a device is controlled by the user's hand-arm system within the device-specific workspace and enables haptic interaction in virtual environments. One specific application, for which human interaction with

a haptic device is essential, refers to virtual haptic object identification. This field of research will be introduced subsequently. Furthermore, two studies revealing the necessity of improving users' proprioceptive localization precision will be presented in this context.

Virtual Object Identification

The haptic sense is of increasing importance in virtual environments, because it provides high functionality due to its active and bi-directional nature. The haptic sense can be utilized for solving a variety of tasks in the virtual world. One important task is the exploration and identification of virtual shapes and objects. In particular, blind users, who cannot study graphical illustrations in books, benefit immensely from employing their sense of touch. Creating digital models based on graphical illustrations or models of real physical items allows these users to explore virtual representations and, therefore, to gain information effectively also without the visual sense. A similar idea was followed by the PURE-FORM project [21], which aimed to enable blind users to touch and explore digital models of three-dimensional art forms and sculptures. Providing haptic access to mathematical functions is another exemplary application [42].

The haptic identification of virtual shapes and objects is of great importance for sighted users as well. Studies have shown that memory performance can be increased significantly using multimodal learning methods [37]. Thus, haptic identification has great potential in the field of education, for example, if digitized models or anatomical shapes are explored multimodally rather than solely visually. Another important application refers to medical training or teleoperation in minimally invasive surgery. Because of the poor camera view and the sparsely available visual cues, surgeons must use their long medical instruments to identify anatomical shapes during surgery, for instance, during the removal of a gallbladder [24]. This task is quite challenging, which is why medical students must receive training to perform it [20]. Furthermore, utilizing haptic feedback to identify anatomical shapes is of vital importance for teleoperating surgeons [18, 32]. Finally, another promising application refers to the "haptification" of data, for example, scientific ones [15, 35].

These examples represent only a small portion of the entire spectrum of possible applications, but they demonstrate the importance of haptic virtual shape and object identification for various groups of users.

Enabling the user to touch, explore and identify virtual shapes and objects requires a haptic feedback device that serves as an interface between the user and the application. First of all, this device must be capable of delivering geometrical cues to the user because such cues are of primary importance for creating a mental image of the object. Stamm et al. investigated whether a state-of-the-art haptic force-feedback device providing one point of contact can be successfully applied by test persons in geometry identification experiments [39]. Exploring a virtual geometry with one point of contact means imitating the contour-following exploratory procedure that is intuitively used in daily life to determine specific geometric details of a real object

[27]. Stamm et al. revealed that test persons experience various difficulties during the exploration and recognition process. One such observed difficulty refers to participants' insufficient spatial orientation in the haptic virtual scene. They often reached the boundaries of the force-feedback workspace unconsciously and misinterpreted the mechanical limits as a virtual object. An additional problem occurred if the participants explored the surface of a virtual object and approached an edge or a corner at one of its sides. Often, the haptic interaction point, HIP, which is comparable to the mouse cursor, slipped off of the object and got lost in the virtual space. Thus, participants lost orientation and could not locate their position in relation to the object. They required a considerable amount of time to regain contact with the object and typically continued the exploration process at a completely different position on the surface. This considerable problem was also observed by Colwell et al. [12] and makes it quite difficult and time consuming to explore and identify virtual objects effectively.

In general, studies as those mentioned above have been conducted with blind-folded participants, because the focus is set on investigating the capability of the haptic sense during interaction with the corresponding devices. However, the observed orientation-specific difficulties cannot be solved easily by providing visual information, because the orientation in a workspace directly in front of the human body is not only controlled by vision. In our daily life, the position of the hand is determined by integrating proprioceptive and visual information [14, 41]. Thus, proprioception plays an important role in the localization process. This role may be even more important in virtual workspaces, where the weighting factor of proprioceptive information may often be considerably higher than that of visual sensory input. How can this be explained? Here are five reasons.

1. With a two-dimensional presentation of a three-dimensional scene on a computer monitor, depth cannot be estimated easily using vision. This problem is also often described in medical disciplines [24]
2. Due to varying illumation conditions or often occurring visual occlusions in the virtual scene, visual information cannot be used reliably.
3. The same holds true when the visual perspective is not appropriate or the camera is moving
4. The visual channel is often overloaded during the interaction in virtual environments. This is why information cannot be processed appropriately [9, 31]
5. Visual attention has to be focused on the localization task and, for example, not on specific graphs on the computer monitor, otherwise vision cannot contribute to a reliable estimation

These examples demonstrate that confusion arises even if visual sensory input is available. When the visual information is not used or cannot be used reliably during virtual interaction, the haptic and *auditory* modality has to compensate for this confusion.

1.3 Outline of the Chapter

Motivated by the orientation-specific difficulties observed in the aforementioned object identification experiments, first, the authors of the present chapter developed an experimental design to quantitatively measure proprioceptive localization accuracy within three-dimensional haptic workspaces directly in front of the human body. The experimental design and the results of a corresponding study with test persons are described in Sect. 2. In a second step, the influence of binaural signals on proprioception is investigated to reveal whether synthetically generated spatial sound is considered and can probably even improve human localization performance. In this context, a hapto–audio system was developed to couple the generated aural and the haptic virtual environment. This approach and the corresponding experimental results are described in Sect. 3. On the basis of the results, finally, conclusions will be drawn in Sect. 4 concerning the importance of binaural models for haptic virtual environments.

2 Proprioceptive Localization

To describe the accuracy of proprioception quantitatively, studies were conducted by scientists across a range of disciplines, for example, computer science, electrical engineering, mechanical engineering, medicine, neuroscience and psychology. These studies focused primarily on two issues. First, the ability to detect joint rotations was investigated to determine the corresponding absolute threshold of the different joints [10, 11, 19]. Second, the ability to distinguish between two similar joint angles was investigated to determine the corresponding differential threshold of the joints and thus the just noticeable difference, JND, [23, 40]. In addition to the physiological limits, the so-called *haptic illusions* must also be considered. Haptic illusions, and perceptual illusions in general, are systematically occuring errors resulting from an unexpected "discrepancy between a physical stimulus and its corresponding percept" [26]. These illusions enable us to obtain a greater understanding of the higher cognitive processes that people use to mentally represent their environments [26]. A haptic illusion that distorts the proprioceptive space perception is the radial-tangential illusion. This illusion describes the observation that the extent of radial motions away and toward the body is consistently overestimated in comparison to the same extent of motions that are tangentially aligned to a circle around the body. Different explanations have been offered, but the factors that cause this illusion to arise are not yet understood—see [26] for more details.

Proprioceptive accuracy and haptic illusions were investigated within corresponding experiments by restricting the test persons to specific joint rotations, movement directions, movement velocities, and so on. However, how do blindfolded test persons actually perform in a localization task if they are allowed to freely explore the haptic space? To the best of our knowledge, no study has investigated freely exploring test

persons' localization performance in a three-dimensional haptic virtual environment. However, this is an important issue for two reasons. First, free exploration is typically used in real world interaction and, thus, should not be restricted in virtual environments. Otherwise, the usability of a haptic system is considerably reduced. Second, the abovementioned observations of the haptic identification experiment indicated that the participants experienced various orientation-specific difficulties. These difficulties should be investigated quantitatively to obtain a cohesive understanding of the proprioceptive orientation capabilities.

2.1 Experimental Design and Procedure

The challenge of an experiment that investigates the abovementioned relation is the *guiding* of the test person's index finger, which freely interacts with a haptic device, to a specific target position. If this target position is reached, the test person can be asked to specify the position of the index finger. The difference between the actual and the specified position corresponds to the localization error. The details of this method are explained below.

An impedance-controlled PHANToM–Omni haptic force-feedback device[1] [36] was used in the present experiment. It provides six degrees-of-freedom, 6–DOF, positional sensing and 3–DOF force-feedback. The small and desk-grounded device consists of a robotic arm with three revolute joints. Each of the joints is connected to a computer-controlled electric DC motor. When interacting with the device, the user holds a stylus that is attached to the tip of the robot arm. The current position of the tip of the stylus is measured with an accuracy of approximately 0.06 mm.

For the experiment, a maximally-sized cuboid was integrated into the available physical workspace of the PHANToM. This cuboid defines the virtual workspace and is shown in Fig. 1. Its width is 25 cm, its height is 17 cm, and its depth is 9 cm. The entire cuboid can be constructed with 3825 small cubes whose sides measure 1 cm. The cubes that are positioned on the three spatial axes are shown in Fig. 1. However, the cubes cannot be touched because they are not present as haptic virtual objects in the scene. Rather, they serve to illustrate a specific position inside the virtual workspace.

The test persons were seated on a chair without armrests. At the beginning of each trial of the experiment, the target—a sphere with a diameter of 1 cm—was randomly positioned inside the virtual workspace and thus inside one of the cubes. Then, the experimenter moved the cursor-like HIP—which corresponds to the tip of the stylus—to the reference position, namely, the center of the virtual workspace. Because of a magnetic effect that was implemented directly on the central point, the stylus remained in this position. The blindfolded test persons grasped the stylus and held it parallel to the y-axis in such a manner that the extended index finger corresponded to the tip of the stylus. The test persons were asked to search for the
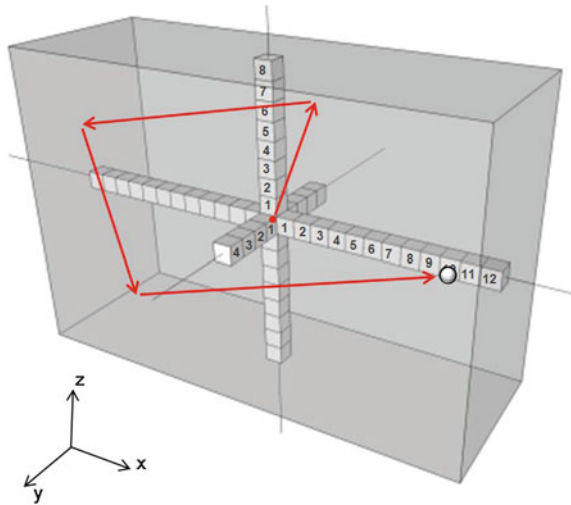
---

[1] Manufactured by SensAble Technologies

Fig. 1 The virtual workspace is shaped like a cuboid. The cuboid can be constructed with 3825 small cubes whose sides measure 1 cm. The cubes that are positioned on the three spatial axes are shown here. However, the virtual cubes cannot be touched. Rather, they serve to illustrate a specific position inside the virtual workspace

hidden sphere inside the virtual workspace with random movements starting from the reference position. Once the HIP was in close proximity to the target, the magnetic effect on the surface of the touchable sphere attracted the HIP. Thus, the target position was reached. In the next step, the test persons were asked to specify the current position of the HIP without conducting additional movements. They could use words such as *left/right*, *up/down* and *forward/backward*. In addition, they were asked to specify the exact position relative to the reference position with numbers in centimeters for each axis. The numbers are exemplarily indicated for each axis in Fig. 1.

Participants received training on the entire procedure prior to the test conditions. In the training session, first, the test persons were introduced to the device and the dimensions of the virtual workspace. For this purpose, they used the visualization that is shown in Fig. 2. This visualization helped the participants to imagine what occurs when they move their arms. Because of the arm movement, the index finger respectively the tip of the stylus was displaced. The HIP was displaced in the same direction inside the virtual workspace. Once the HIP left a cube and entered another, the highlighted block moved and visualized the change of position. Furthermore, the test persons were introduced to the mechanical limits of the physical workspace of the device, which were somewhat rounded and slightly outside of the virtual workspace. In the second step, the participants were blindfolded and administered the experimental task in four exemplary trials. Each time they found the hidden sphere and estimated its position, the experimenter gave feedback about their localization error. However, this kind of feedback was only provided in these exemplary trials. After
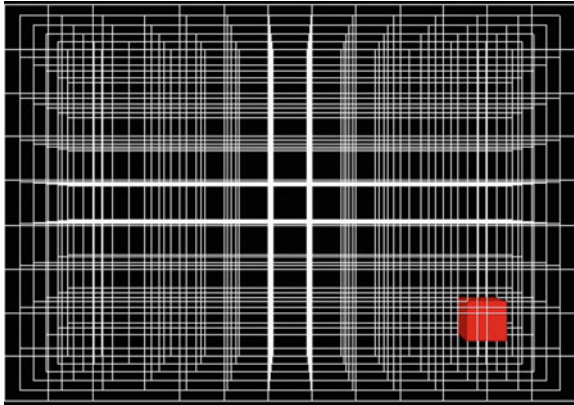
Fig. 2 Visualization used at the beginning of the training session. It helps the participants to imagine what occurs when they move their arms. Because of the arm movement, the index finger respectively the tip of the stylus is displaced. The HIP is displaced in the same direction inside the virtual workspace. Once the HIP leaves a cube and enters another, the highlighted block moves and visualizes the change of position

the training procedure, subjects indicated no qualitative difficulties in the estimation process. They completed this task in a few seconds.

2.2 Condition #1

To investigate the influence of differently positioned haptic workspaces on localization performance, a construction was built such that the position of the PHANToM could be easily changed. In the first condition, the device was placed at the height of a table—Fig. 3a. This position is quite comfortable and familiar in daily life from writing or typing on a computer keyboard.

Twelve test persons, two female and ten male, voluntarily participated in the first experimental condition. Their ages ranged from 21–30 years with a mean of 24 years. All participants indicated that they had no arm disorders. They were students or employees of Dresden University of Technology and had little to no experience using a haptic force-feedback device. The participants each completed 10 trials.

The results of the proprioceptive measurements are outlined in Fig. 4. The average of the absolute localization error vector in centimeters in the x-, y- and z-direction is shown. The grey bars refer to the overall localization errors, that is, the localization errors averaged over all hidden target spheres. The error increased from 1.4 cm in the forward-backward direction, that is, the y-axis, to 1.8 cm in the vertical direction, z-axis, and to 2.4 cm in the lateral direction, x-axis. The error increased because of the different side lengths of the cuboid—its length in the forward-backward direction is 9 cm, in the vertical direction 17 cm and in the lateral direction 25 cm. This
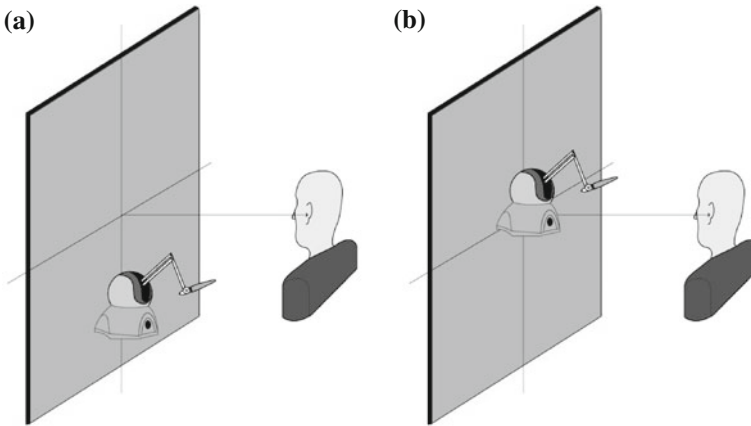
Fig. 3 A construction was built such that the position of the PHANToM could be easily changed. By varying the position of the device, it is possible to investigate the influence of differently positioned haptic workspaces on localization performance. **a** In the first experimental condition, the device was placed at the height of a table. **b** In the second experimental condition, the device was placed at the height of test persons' head

fact was verified by calculating the localization errors solely for those target spheres that were randomly hidden in the same cuboid within a centered cubic volume with side lengths of 9 cm. In this case, the localization errors were almost the same in all directions. The corresponding error was 1.4 cm in the forward-backward direction and 1.7 cm in the vertical and the lateral directions. Thus, the localization accuracy of the proprioceptive sense did not depend on movement direction within this small workspace, although different types of movements were used, nameley radial movements along the y-axis and tangential movements along the x- and z-axis. Rather, the localization accuracy depended on the distance between the hidden sphere and the reference position. On average, the sphere was hidden further away from the reference point in the lateral direction than in the vertical and forward-backward directions. Therefore, the resulting error was greatest along the x-axis. This dependency is verified by the dark bars that are shown in Fig. 4. These bars refer to the localization errors for the spheres that were randomly hidden in the border area of the virtual workspace. The border area is defined as follows.

$$
\begin{aligned}
x_{\mathrm{pos}} &> |6\,\mathrm{cm}| \mathrm{OR} \\
y_{\mathrm{pos}} &> |2\,\mathrm{cm}| \mathrm{OR} \\
z_{\mathrm{pos}} &> |4\,\mathrm{cm}|.
\end{aligned}
\tag{1}
$$

The mean localization error increased for all three directions. Therefore, if the sphere was hidden further away from the reference point in a specific direction, the resulting error was greater. An ANOVA for repeated measurements revealed a significant
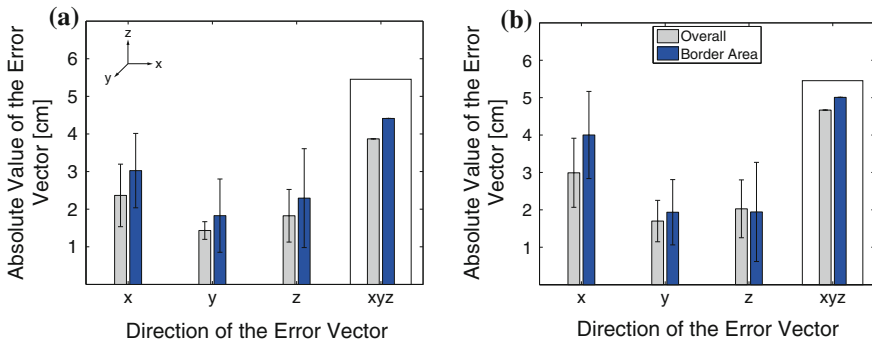
Fig. 4 Experimental results. **a** Condition #1, PHANToM at table level. **b** Condition #2, PHANToM at head level. The average of the absolute localization error vector ± standard deviations in centimeters in the x-, y- and z-direction is shown. The calculated localization error inside the three-dimensional virtual workspace is also outlined. The *grey bars* refer to the overall localization errors, that is, the localization errors averaged over all hidden spheres. The *dark bars* refer the localization errors for those spheres that were randomly hidden in the border area of the virtual workspace. The border area is defined in (1)

difference between the localization errors indicated by the grey and dark bars for the x-axis, namely, $F = 19.164$ and $p < 0.01$. In addition, the standard deviations also increased for all three directions because of the broader spreading of test persons' estimations in these seemingly more complex trials.

However, what is the reason for this increasing error and for the increasing instability of test persons' estimations? Previous works observed a diminished accuracy of the proprioceptive sense once whole arm movements were involved in the exploration process [26]. In the present experiment, participants were required to move their whole arms to reach the border area of the virtual workspace, which may have affected their estimations. As a result, their arms were not in contact with their body; therefore, their body could not be used as a reference point. To investigate whether the described effect can be verified or even strengthened if whole arm movements are provoked, a second experimental condition was conducted.

2.3 Condition #2

In the second condition, the PHANToM was placed at the height of the test persons' head—see Fig. 3b. Thus, even if the HIP was located near the reference position, the arm was not in contact with the body anymore.

Twelve different subjects, three female and nine male, voluntarily participated in the second experimental condition. Their ages ranged from 21–49 years with a mean of 29 years. All participants indicated that they had no arm disorders. They were students or employees of Dresden University of Technology and had little to no

experience using a haptic force-feedback device. The participants each completed 10 trials.

The results are presented in Fig. 4b. In comparison to condition #1, the overall localization error increased from 1.7 cm in the forward-backward direction, y-axis, to 2.0 cm in the vertical direction, z-axis, and to 3.0 cm in the lateral direction, x-axis. Again, no dependency between the localization accuracy and the movement direction was observed. This was verified as described above in Sect. 2.2. Rather, the error increased because of the different side lengths of the cuboid-shaped virtual workspace. On average, the sphere was hidden further away from the reference point in the lateral direction than in the vertical and forward-backward directions. Therefore, the resulting error was greatest along the x-axis. The dependency between the localization performance and the average distance to the target was verified through the calculation of the localization error for the border area of the virtual workspace. This error increased in comparison to the overall localization error for the x- and y-axis but not for the z-axis. An ANOVA for repeated measurements identified significant differences on the x-axis with $F = 9.692$ and $p < 0.05$.

When comparing these results with those of condition #1, both the overall localization accuracy and the localization accuracy for selected spheres in the border area decreased, especially for the x- and the y-axis. Therefore, the resulting overall error in the three-dimensional space increased considerably from 3.9 cm in condition #1 to 4.7 cm in condition #2. The resulting error in the three-dimensional workspace is an important issue concerning virtual haptic interaction. The values of 3.9 and 4.7 cm are considerable amounts, if the length of the virtual objects is limited, for example, to 10 cm. This size was used in the aforementioned identification experiments [39] in which the orientation-specific difficulties were originally observed.

Finally, it is important to note that the outlined results were obtained in a workspace that was slightly smaller than a shoe box. Because it is was found that movement distance directly influences the accuracy of proprioception, the mean-percentage localization errors might be greater in larger workspaces.

3 Audio–Proprioceptive Localization

3.1 Pre-Study

During the localization experiment described in Sect. 2, the test persons only used their body's own proprioceptive signals and the physical boundaries of the device-specific workspace for orientation. Thus, the experimental conditions were identical to those of the object identification experiments in which the orientation-specific problems were originally observed [12, 39]. This was a crucial requirement for identifying the cause of the difficulties. However, the physical boundaries of the device-specific workspace located slightly outside of the virtual cuboid were somewhat rounded and irregularly shaped due to the construction of the robot arm. This irreg-

ular shape could probably impede users' localization performance when interacting with haptic devices. It should be investigated in the pre-study whether such a negative influence actually exists. The experimental procedure was the same as described in Sect. 2.1 except that the test persons utilized their body's own proprioceptive signals and a pink-noise monophonic sound, which was switched off when the HIP touched or moved beyond the boundaries of the cuboid-shaped virtual workspace. These virtual boundaries were straight and thus regularly shaped—Fig. 1.

Two independent groups voluntarily participated in the pre-study. The first group consisted of twelve test personss, two female and ten male, aged from 21–34 years with a mean of 27 years. They took part in the first experimental condition—Fig. 3a. The second group also consisted of twelve test personss, two female and ten male, aged from 21–49 years and a mean of 29 years. They participated in the second experimental condition—Fig. 3b. All participants indicated that they had no arm disorders. They were students or employees of Dresden University of Technology and had little to no experience using a haptic force-feedback device. The participants each completed 20 trials. The trials were divided into two halves whose presentation order was randomized. In one half of the trials, the test persons used proprioception and a monophonic sound. The corresponding results are presented subsequently. The other half of the trials is detailed in the following section.

The results of the pre-study are illustrated with the help of the dark bars in Fig. 5. The grey bars refer to the results of the proprioceptive-only measurements that were already discussed in Sect. 2. They are shown again to aid comparison. Within the first experimental condition, test persons' localization performance in lateral direc-
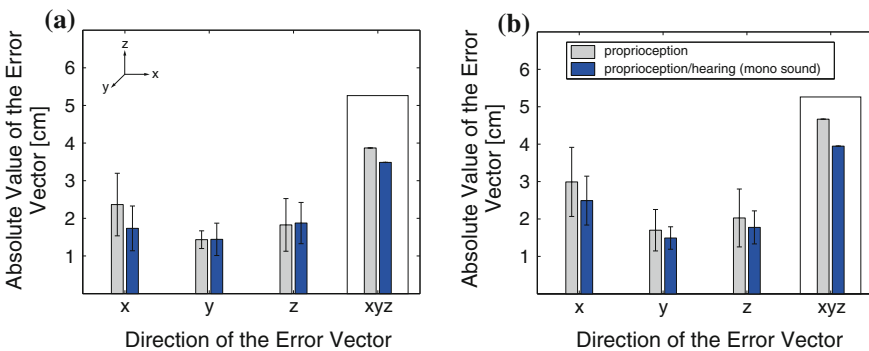

Fig. 5 Experimental results. **a** Condition #1, PHANToM at table level. **b** Condition #2, PHANToM at head level. The average of the absolute localization error vectors ± standard deviations in centimeters are depicted for the *x*-, *y*- and *z*-direction as well as for the three-dimensional workspace. The *grey bars* refer to the results of the proprioceptive measurements, that is, the test persons only used their body's own proprioceptive signals and the physical boundaries of the device-specific workspace for localization. The *dark bars* refer to the results of the pre-study in which the test persons utilized their bodies' own proprioceptive signals and a monophonic sound that was switched off when the HIP touched or moved beyond the boundaries of the cuboid. The *grey bars* have been previously depicted in Fig. 4; they are shown again for comparison

tion increased due to the acoustically defined, regularly shaped borders of the virtual workspace. An ANOVA for independent samples revealed a statistically significant effect with $F = 4.607$ and $p < 0.05$. However, no effect was observed concerning the localization accuracy in forward-backward and in vertical direction. Within the second experimental condition, test persons' localization performance increased slightly in x-, y- and z-direction, but no statistically significant effect was found.

In conclusion, the pre-study can be summarized as follows. First, the acoustically defined, regularly shaped borders of the virtual workspace helped to prevent users from misinterpreting the mechanical limits of the haptic device as a virtual object. Second, the localization errors and the standard deviations could yet only be partly decreased.

3.2 Main Study

It is investigated within the main study how both proprioception and spatialized sound are used in combination to localize the HIP. Thus, the influence of binaural signals on proprioception is studied to reveal whether synthetically generated spatial sound is considered and might even improve human proprioceptive localization performance. Employing the hearing system to extend proprioceptive perception auditorily seems to be a valuable approach. For example, individuals utilize their hearing system on the streets daily to estimate the direction and distance from which a car approaches without being forced to look at the car. Because spatial audible information is intuitively used in the real world, it should also be incorporated in virtual environments.

The usefulness of auditory localization cues for haptic virtual environments was verified in several studies. Such cues were successfully used in hapto-audio navigational tasks, for example, when users attempted to explore, learn and manage a route in a virtual traffic environment [29]. They were also quite helpful for locating objects in the virtual space [28]. In these studies, the haptic device was used to move the virtual representation of oneself and, thus, one's own ears freely in the virtual scene—that is, as an *avatar*. Therefore, if an object emitted a sound, the user heard this sound from a specific direction depending on the position and the orientation of the hand-controlled avatar. Thus, this method was called the *ears-in-hand* interaction technique [29].

However, the present work aims to investigate how proprioceptive signals and auditory localization cues provided by spatialized sound are used in combination to localize the absolute position of the HIP. The present study's approach is not comparable to the abovementioned approaches. In the current study, the HIP does not correspond to a virtual representation of oneself including one's own ears. Rather, it corresponds to a virtual representation of the fingertip. Furthermore, this study does not aim to localize a sound-emitting object or something similar in the virtual scene. Rather, the study aims to trace the movements of the virtual fingertip auditorily to increase the localization resolution. To investigate whether localization performance can be improved, it is essential to develop a hapto-audio system that auralizes each

movement of the user and, thus, the movement of the haptic device. Therefore, each arrow depicted in Fig. 1 must cause a corresponding variation in the reproduced spatial sound, as if the user moves a sound source with his/her hand in the same direction directly in front of the his/her head. The authors of the present chapter call this method the *sound source-in-hand* interaction technique.

Development of a Hapto–Audio System

There are two main methods for reproducing spatial sound. On the one hand, head-related transfer functions, HRTFs, can be used for binaural reproduction via headphones [6]. In this case, the free-field transfer function from a sound-emitting point in space to the listener's ear canal is used to filter the sound. On the other hand, spatial sound can be reproduced by various techniques using loudspeaker arrays, for example, ambisonics [30], vector-base amplitude panning, VBAP, [34] and wave field synthesis, WFS, [4].

In the present study, generalized HRTFs are used to generate auditory localization cues. An extensive set of HRTF measurements of a KEMAR dummy-head microphone is provided by Gardner et al. [17]. These measurements consist of the left and right ear impulse responses from a total of 710 different positions distributed on a spherical surface with 360° azimuth and −40° to 90° elevation. In a first step, the set of shortened 128-point impulse responses was selected for this investigation. Second, this set was reduced to impulse responses in the range between −30° and +30° in azimuth and elevation. For the real-time binaural simulation of a continuously moving sound source according to the sound source-in-hand interaction technique, an algorithm capable of interpolating between the HRTF measurement points is required. This algorithm should smoothly handle also very fast movements without interruptions. That is the reason why the convincing time-domain-convolution algorithm of Xiang et al. [43] was selected. This algorithm approximates the HRTF at the target position of the current signal block by linearly interpolating the four nearest measurement points. An exemplary case is depicted in Fig. 6. The point T corresponds to the target position which is surrounded by the points P_1, P_2, P_3 and

Fig. 6 The HRTF at target position, T, of the current signal block is computed by linearly interpolating the HRTFs at the four nearest measurement points, P_1, P_2, P_3 and P_4. This figure is only for demonstration purposes. The four measurement points were chosen arbitrarily—after [43]
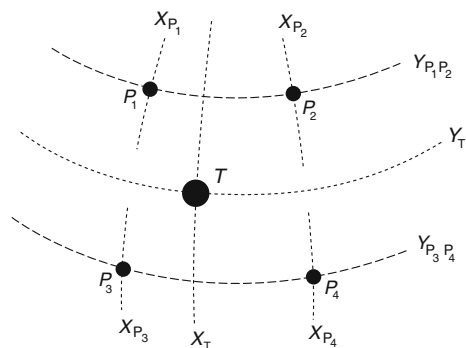
P_4. The measured HRTFs at these four points, F_{P_1}, F_{P_2}, F_{P_3} and F_{P_4}, can now be used to calculate the HRTF, F_T, at the target position, T, according to the following equation [43],

$$F_T = \frac{Y_T - Y_{P_1 P_2}}{Y_{P_3 P_4} - Y_{P_1 P_2}} \left(\frac{X_T - X_{P_1}}{X_{P_2} - X_{P_1}} F_{P_1} - \frac{X_{P_2} - X_T}{X_{P_2} - X_{P_1}} F_{P_2} \right)$$
$$+ \frac{Y_{P_3 P_4} - Y_T}{Y_{P_3 P_4} - Y_{P_1 P_2}} \left(\frac{X_T - X_{P_3}}{X_{P_4} - X_{P_3}} F_{P_3} - \frac{X_{P_4} - X_T}{X_{P_4} - X_{P_3}} F_{P_4} \right) . \tag{2}$$

The next step refers to the transformation of the monophonic signal block, x, to the spatialized signal block, y. Both blocks can be specified in their lengths by the block-size of $b = 64$ samples. The transformation process is realized by the time-domain convolution of the input signal with a 128-tap filter. This filter is again the result of a linear interpolation between the HRTFs F_T and F_{T_0}. F_T was computed for the current signal block and the target position, T. F_{T_0} was computed for the previous signal block and corresponding previous target position T_0. In conclusion, each sample, k, of the output signal is calculated according to the following equation [43],

$$y(k) = \sum_{n=0}^{127} x(k - n) \cdot \left(\frac{k}{b} \cdot F_T(n) + \frac{b - k}{b} \cdot F_{T_0}(n) \right) , \tag{3}$$

$$\text{with} \quad k = 0, 1, \ldots b - 1 .$$

Using the abovementioned selection of head-related impulse responses, the sound source reproducing broadband noise could be virtually positioned anywhere in the range between $-30°$ and $+30°$ in azimuth and elevation. However, at this point, the generated aural environment was only two-dimensional. To auralize movements in the forward-backward direction, auditory distance cues must be provided. An intuitive way to achieve a high resolution for the localization in the forward-backward direction is to vary the sound pressure level of the corresponding signal. It is well-known from our daily experiences that a distanced sound source is perceived as quieter than a proximal sound source. This fact can be easily utilized in virtual environments. However, because of the small virtual workspace, the sound pressure level variation must be exaggerated in comparison to the real physical world. This exaggeration allows to profit considerably of the high resolution and, thus, the low differential thresholds of the hearing system as concerns level variation, namely, $\Delta L \approx 1$ dB, [16]. In the current study, a level range of $L = 60 \pm 12$ dB(A) was applied. Thus, the localization signal became louder when the user moved the HIP forward, and it became quieter when the user moved the HIP backward, that is, away from the body.

Finally, the aural and haptic environments must be linked with each other. The aim is to generate a hapto–audio space by auditorily extending the haptic environment. This extension is performed linearly in depth but also in height and width. Thereby, the height is enlarged more than the width to qualitatively account for the lower
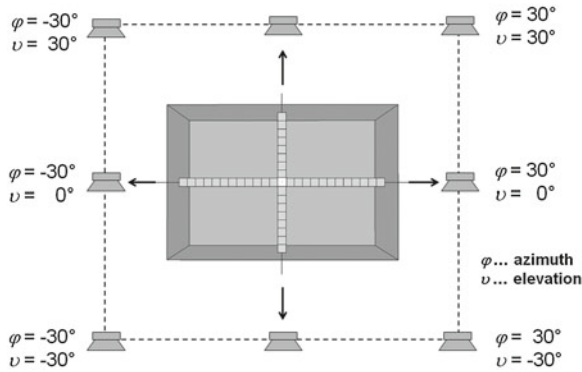
Fig. 7 The hapto–audio workspace is generated by auditorily extending the haptic environment. This extension is shown here schematically for the height and width
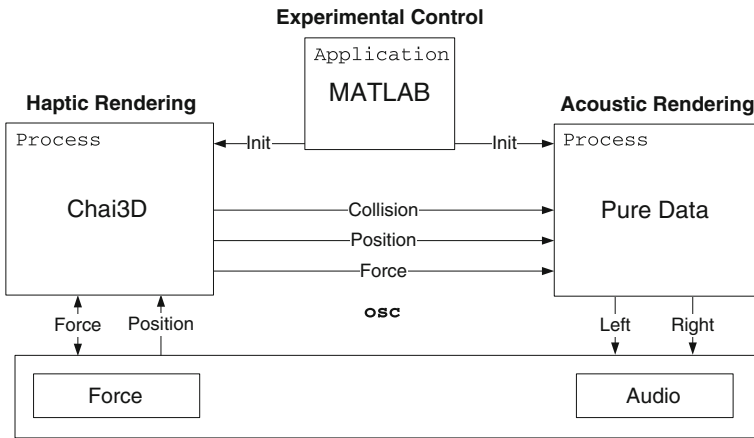

Fig. 8 Experimental software framework

auditory localization resolution in elevation [6]. The extension is illustrated in Fig. 7 and mathematically detailed in the following section.

The developed software framework incorporating haptic and acoustic rendering processes as well as the experimental control is depicted in Fig. 8. Matlab was used to automate the experimental procedure and question the test subjects. Matlab also initiated the processes for haptic rendering and acoustic rendering.[2] To reproduce the corresponding sound signals, the acoustic renderer requires information from the PHANToM, which communicates with the haptic renderer via a software interface. This information, for example, about the position of the HIP or a detected collision

---

[2] Chai3D, Stanford University, and Pure Data, Open-Source Project, respectively. For details on the C++ haptic-rendering framework, particularly, the algorithms for collision detection, force control and force response, the reader is referred to [13]. The details of Pure Data are outlined in [33]

with the hidden sphere, is steadily transferred to the acoustic renderer via Open Sound Control, OSC. The information is then processed for subsequent signal output, that is, to reproduce spatialized broadband noise via headphones and to provide force-feedback, once the target position is reached.

Experimental Procedure

The test persons were seated on a chair without armrests. At the beginning of each trial of the experiment, a sphere with a diameter of 1 cm was randomly positioned inside the virtual workspace, thus, inside one of the 3825 virtual cubes—see Fig. 1. Then, the experimenter moved the cursor-like HIP, which corresponds to the tip of the stylus, to the reference position, that is, the center of the virtual workspace. Because of a magnetic effect that was implemented directly on the central point, the stylus remained in this position. The generated virtual sound source was positioned at 0° in azimuth and elevation. The sound pressure level of the broadband noise was 60 dB(A). The blindfolded test persons grasped the stylus and held it parallel to the y-axis in such a manner that the extended index finger corresponded to the tip of the stylus. The test persons were asked to search for the hidden sphere with random movements starting from the reference position. When moving the stylus of the haptic device, the HIP and the sound source were also displaced. The direction of the sound source, φ in azimuth and υ in elevation, depended linearly on the position, x and z, of the HIP according to the following equations,

$$\varphi(x) = \frac{\varphi_{max}}{x_{max}} \cdot x \quad \text{with} \quad x = [-x_{max}, \ x_{max}], \tag{4}$$

$$\upsilon(z) = \frac{\upsilon_{max}}{z_{max}} \cdot z \quad \text{with} \quad z = [-z_{max}, \ z_{max}]. \tag{5}$$

The right border of the cuboid, $x_{max} = 12.5$ cm, and the left border of the cuboid, $-x_{max} = -12.5$ cm, corresponded to a maximal displacement of the sound source at $\varphi_{max} = 30°$ and $-\varphi_{max} = -30°$. Similarly, the upper border, $z_{max} = 8.5$ cm, and the lower border, $-z_{max} = -8.5$ cm, corresponded to a maximal displacement of the sound source at $\upsilon_{max} = 30°$ and $-\upsilon_{max} = -30°$. When the HIP moved beyond the limits of the cuboid-shaped virtual workspace, the sound was immediately switched off. The sound-pressure level of the broadband noise depended on the position of the HIP in y-direction. It was calculated according to the following equation with $L_{init} = 60$ dB(A), $\Delta L = 12$ dB(A) and $|y_{max}| = 4.5$ cm,

$$L(y) = L_{init} + \frac{\Delta L}{y_{max}} \cdot y \quad \text{with} \quad y = [-y_{max}, \ y_{max}]. \tag{6}$$

Once the HIP was in close proximity to the target, the magnetic effect that was implemented on the surface of the touchable sphere attracted the HIP. Thus, the

target position was reached. In the next step, the test persons were asked to specify the current position of the HIP without conducting further movements. This method was previously described in Sect. 2.1.

Participants received training on the entire procedure prior to the test conditions. In the training session, first, the test persons were introduced to the device and the hapto-audio workspace. They used the visualization that is depicted in Fig. 2. In the second step, they were blindfolded and listened to consecutively presented sound sources that moved on a specific axis. The corresponding number that specified the actual position of the sound source was provided. This was conducted separately for each axis. Therefore, the test persons experienced the maximal displacements of the sound source and the entire range in between, which is essential for making estimations. Subsequently, the participants conducted the experimental task in four exemplary trials. Each time they found the hidden sphere and estimated its position, the experimenter gave feedback about their localization error. However, this kind of feedback was only provided in these exemplary trials.

Listening Test

In a first step, a listening test was conducted to measure test persons' achievable auditory localization accuracy if the aforementioned spatialization technique is applied. Thus, test persons did not interact with the haptic device. Rather, they only listened to a continuously moving virtual sound source. The path of this source was selected automatically in a random manner and also its motion speed varied randomly. After approximately 10 sec the sound source reached the target position, which could then be specified as described above.

Eight test persons, two female and six male, voluntarily participated in the listening test. Their ages ranged from 24–49 years with a mean of 33 years. All participants indicated that they did not have any hearing or spinal damage. They were students or employees of Dresden University of Technology. The participants each completed 10 trials.

The results of the listening test are shown in Fig. 9 with the white bars. These results enable to determine whether it will be at least theoretically possible to further improve localization accuracy with the spatialization technique. The average auditory localization error in azimuth direction was measured to be $4°-5°$. This value corresponds to a localization error of 1.8 cm inside the virtual workspace. In the first experimental condition, test persons already achieved a localization error of 1.8 cm when using the monophonic sound. Thus, they will not be able to further improve their localization performance with spatial auditory cues. However, test persons should be able to clearly decrease the localization error in lateral direction from 2.5 to 1.8 cm in the second experimental condition, in which whole arm movements are provoked. In this case, the spatial auditory cues should be helpful. Furthermore, the localization error in forward-backward direction might even be halved from 1.4 cm in the first condition and 1.7 cm in the second condition to 0.7 cm with available spatial sound. In contrast, no improvement should be achievable in vertical direction, because the
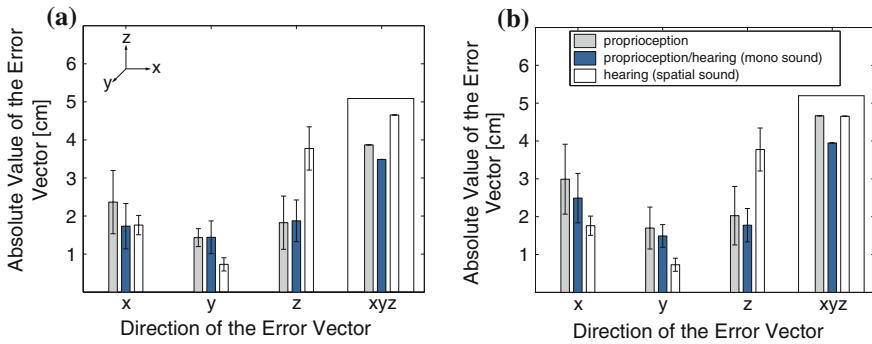
Fig. 9 Experimental results of the listening test. The average of the absolute localization error vector ± standard deviations in centimeters in *x*-, *y*- and *z*-direction is shown. The calculated localization errors inside the three-dimensional virtual workspace are also outlined. The *grey bars* refer to the results of the proprioceptive measurements, that is, no acoustic signals were available. The *dark bars* refer to the results of the pre-study in which the test persons utilized proprioception and a monophonic sound that was switched off when the HIP touched or moved beyond the boundaries of the cuboid. The *white bars* refer to the results of the listening test. They are depicted in the *left* (**a**) and *right* graphs (**b**) for comparison

average auditory localization error in elevation was measured to be $13° - 14°$. This value corresponds to a localization error of 3.8 cm inside the virtual workspace. The proprioceptive localization errors of 1.8 cm in conditions #1 and 2 cm in condition #2 are clearly smaller.

Subsequently, results will be presented that reveal how test persons perform if both proprioceptive and spatial auditory signals are used in combination. These results will help to understand whether binaural signals are considered during localization or whether the localization process is mainly dominated by proprioception. In audio–visual interaction, the bimodal sensory input is integrated in an optimal manner. If participants would also integrate proprioceptive and binaural signals optimally, the resulting localization performance could be predicted. According to such a prediction in the context of this investigation, test persons would reject binaural signals for estimating the target position in vertical direction. They would rather trust in proprioception. However, the binaural signals would be considered to estimate the position in forward-backward direction and in lateral direction in condition #2.

Condition #1

In the first condition, the device was placed at the height of a table, as illustrated in Fig. 3a. The group of participants was already introduced in Sect. 3.1. The participants each completed 20 trials. The trials were divided into two halves, and the presentation order was randomized. During one half of the trials, the test persons used proprioception and a monophonic sound which helped to clearly define the
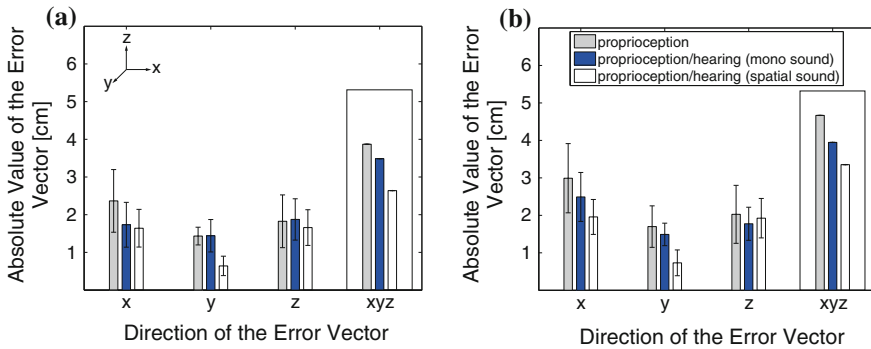
Fig. 10 Experimental results. **a** Condition #1 – PHANToM at table level. **b** Condition #2—PHANToM at head level. The average of the absolute localization error vector ± standard deviations in centimeters in x-, y- and z-direction is shown. The calculated localization errors inside the three-dimensional virtual workspace are also outlined. The *grey bars* refer to the results of the proprioceptive-only measurements. The *dark bars* refer to the results of the pre-study in which the test persons used their body's own proprioceptive signals and a monophonic sound that helped to clearly define the boundaries of the virtual workspace. The *white bars* refer to the results of the main study in which test persons used both proprioceptive sensory input and spatial auditory cues according to the sound source-in-hand interaction technique

boundaries of the virtual workspace. This part of the experiment was denoted as pre-study. During the other half of the trials, the test persons used proprioception and additionally perceived spatial auditory cues according to the sound source-in-hand interaction technique. This part is analyzed subsequently.

The results are shown in Fig. 10a. The grey bars refer to the results of the proprioceptive measurements that were previously illustrated in Fig. 4. The dark blue bars refer to the results of the pre-study. The white bars show the localization errors that occurred during binaural-proprioceptive interaction according to the sound source-in-hand interaction technique. The results indicate that test persons integrate proprioceptive and binaural signals in such a manner that the resulting localization error is minimized. They achieved a localization accuracy of 1.7 cm in vertical direction. Thus, they rejected binaural signals and rather trusted in proprioception. However, they used binaural signals to estimate the position in forward-backward direction which is why the localization error was significantly reduced from 1.4 cm to approximately 0.6 cm. An ANOVA for repeated measurements rendered $F = 57.863$ and $p < 0.001$. No further improvement was observable concerning the localization performance in lateral direction, as predicted before.

In conclusion, test persons could actually differentiate between the more reliable proprioceptive signals for the localization in vertical direction and the more reliable auditory cues for the localization in forward-backward direction. The resulting error in the three-dimensional workspace was overall reduced by approximately 30 % from 3.9 to 2.8 cm.

Condition #2

In the second condition, the PHANToM was placed at the height of the test person's head—Fig. 3b. Thus, whole arm movements were provoked. The group of participants was already introduced in Sect. 3.1. The corresponding results are analyzed below.

In Fig. 10b, the average of the absolute localization error vectors are depicted for the x-, y- and z-direction as well as the three-dimensional workspace. As mentioned before, the grey bars refer to the results of the proprioceptive measurements that were previously illustrated in Fig. 4. The dark bars refer to the results of the pre-study. The white bars show the localization errors that occurred during binaural-proprioceptive interaction according to the sound source-in-hand interaction technique. The results verify that test persons integrate proprioceptive and binaural signals in such a manner that the resulting localization error is minimized. Again, they rejected binaural signals and rather trusted in proprioception when specifying the target position in vertical direction. They achieved a localization accuracy of 1.9 cm. However, they used binaural signals to estimate the position in forward-backward direction which is why the localization error was more than halved from 1.5 to 0.7 cm. An ANOVA for repeated measurements rendered $F = 25.226$ and $p < 0.001$. Furthermore, as predicted before in case of optimal integration, the localization error in lateral direction could be further reduced from 2.5 to 1.9 cm. An ANOVA for repeated measurements revealed a statistically significant effect, namely, $F = 5.532$ and $p < 0.05$.

In conclusion, test persons could again differentiate between the more reliable proprioceptive signals for the localization in vertical direction and the more reliable auditory cues for the localization in forward-backward and lateral direction. The resulting error in the three-dimensional workspace was overall reduced by approximately 30 % from 4.7 to 3.4 cm.

4 Summary and Conclusions

With the help of binaural models, human localization can be predicted under the assumption that a corresponding localization process is based on acoustic signals, thus, on unimodal information. However, what happens if this localization process is realized in an environment with available bimodal or even multimodal sensory input? This chapter sets focus on investigating the influence of binaural signals on proprioception to reveal whether synthetically generated spatial auditory signals are considered during localization and might even improve human proprioceptive localization performance in virtual workspaces. Quantitative results were gained with the help of corresponding experiments in which freely exploring test persons' unimodal and bimodal localization performance was measured. To obtain a greater understanding of such complex experiments with actively interacting test persons, the experimental variables, concurrent experimental processes, results and potential errors, it is quite helpful to develop a formalization of the corresponding experimental
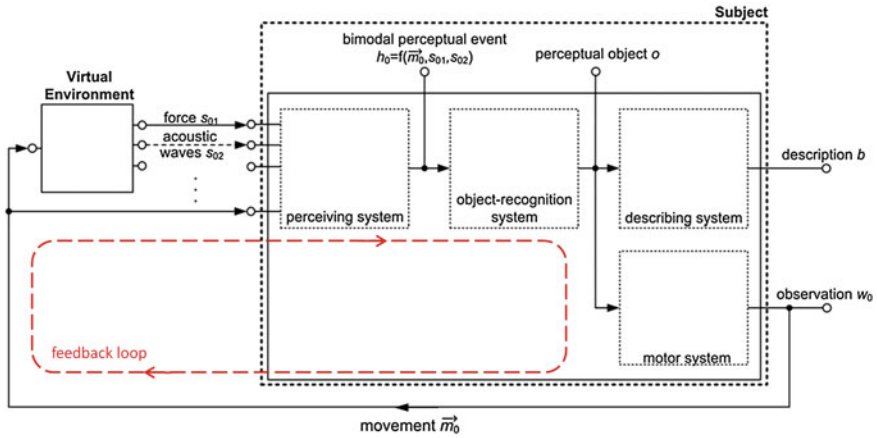
Fig. 11 Formalization. Basic schematic representation of a test person, (*right block*), that haptically interacts in a virtual environment (*left block*) during a haptic experiment

type. Blauert's system analysis of the auditory experiment serves as an example [6]. However, because of the active and bi-directional nature of the haptic human sense, Blauert's approach must be adapted and extended to describe experiments or situations in which subjects haptically interact in a virtual environment. Therefore, a further developed formalization based on his model is introduced subsequently.

This formalization is depicted in Fig. 11. It contains the basic schematic representation of the active subject—right block—who interacts in the virtual environment— left block—by controlling a specific device. The inputs and outputs of the blocks and the interaction between them can be explained with reference to the main study. During the experiment, the test person actively moved his/her own arm through commands of the motor system. The corresponding movement, $\vec{m}_0$, at the time t_0 was transmitted to the haptic device, which was controlled by the test person's index finger. Thus, the device served as an interface and provided input to the virtual environment. According to the sound source-in-hand interaction technique, the subject heard a spatial sound that varied depending on the current movement direction. The corresponding acoustic waves, s_{02}, were output ideally without delay at the time t_0 by the virtual environment via headphones. At the same time, the subject also directly perceived his/her own joint movements through proprioception. As a result of the proprioceptive signals and auditory cues, the test person perceived a bimodal perceptual event, h_0, which can be described. Of course, both unimodal perceptual events, h_{01} and h_{02}, can also be described separately. By integrating the preceding bimodal events and the current event, a perceptual object develops. A perceptual object is the result of mental processing. In the present case, it corresponded to a mental image of the virtual workspace and the corresponding position of the HIP in the workspace. A special characteristic and, at the same time, a distinguishing characteristic in comparison to auditory experiments is indicated by the feedback loop depicted in Fig. 11. To find the hidden sphere, the subject explored the virtual workspace continuously and

individually by moving the stylus of the PHANToM. Thus, the perceived bimodal events, $h_0 ... h_n$, differed across subjects because of the individual movements $\vec{m}_0 ... \vec{m}_n$ and the resulting acoustic signals $s_{02} ... s_{n2}$. Once the sphere was found at time t_n, a collision between the HIP and the sphere was detected. The virtual environment exerts a force, s_{n1}, to the test person via the haptic device. The shortly preceding bimodal events and h_n, the current event, were integrated to specify the perceptual object, o. As a result, the experimenter obtained a description, b, of the subject concerning the internal perceptual object, o. This description contained the assumed position of the target. With the help of these descriptions, collected from a group of participants, quantitative relations can be obtained, for instance, Figs. 4, 5, 9 and 10. However, the experimenter can also profit from experimental observations, for example, by watching each movement, watching the involved joints and recording the required time. These observations can help in understanding how subjects interact with the virtual environment and in determining possible explanations for their responses.

In the main part of this chapter, first, the localization accuracy of proprioception was investigated. The test persons only perceived their own movements, $\vec{m}_0 ... \vec{m}_n$, with the help of the proprioceptive receptors. The resulting unimodal perceptual events, $h_0 ... h_n$, served to develop a perceptual object, that is, a mental representation of the virtual workspace and the HIP in it. The virtual environment provided no additional information. The experimental results help to explain the orientation-specific difficulties that were originally observed in the object identification experiments in Stamm et al. [39] and Colwell et al. [12]. In those experiments, subjects required a substantial amount of time to regain a virtual object after they lost contact with it because they could not easily locate the HIP in relation to the object. Furthermore, they often reached the boundaries of the physical workspace unconsciously and misinterpreted the mechanical borders as an object. These difficulties indicate that their mental representation of the virtual workspace and the position of the HIP in it deviated from reality. The present study found that, indeed, subjects considerably misjudge the actual position of the HIP inside the three-dimensional workspace by approximately 4–5 cm. This is a remarkable amount, for example, if the length of the virtual objects is limited to 10 cm, as it was the case in the aforementioned identification experiments. The current study further found that the localization accuracy of proprioception depends on the distance between the current position of the HIP and a corresponding reference point. If whole arm movements are used to overcome this distance, the localization error and the standard deviations increase considerably. This increase was observed especially for the lateral direction and to a small, non-significant extent for the forward-backward direction.

In the second step, it was investigated whether the localization accuracy improves if proprioception and the hearing sense are used in combination. Thus, the test persons perceived their own movements, $\vec{m}_0 ... \vec{m}_n$, but also the acoustic signals $s_{02} ... s_{n2}$. As a result, they perceived the bimodal events $h_0 ... h_n$. The experimental results of the main study demonstrated that the abovementioned proprioceptive localization inaccuracy was reduced significantly. Proprioception can be guided if additional spatial auditory cues are provided and, thus, the localization performance can be improved. Significant effects were found concerning the localization errors

in forward-backward direction and also in lateral direction when whole arm movements were provoked. The resulting errors in the three-dimensional workspace were overall decreased by approximately 30 %.

It is important to mention that the localization accuracy was investigated in a haptic workspace that was slightly smaller than a shoe box. Because it is generally accepted that proprioceptive accuracy depends on the degree to which whole arm movements are involved in the exploration process, proprioceptive localization performance might be worse in larger workspaces. In such workspaces, the positive influence of auditory localization cues may be even stronger. This hypothesis should be investigated in future studies. Furthermore, it would be quite useful to investigate to which extent the localization performance in vertical direction can be improved, for example, if individualized HRTFs are used for the spatialization.

In conclusion, the experimental results clearly show that synthetically generated auditory localization signals are considered in the localization process and can even guide human proprioceptive localization within workspaces directly in front of the human body. The auditory and proprioceptive information is combined in such a way that the resulting localization error is minimized. As described in the introductory part, a similar effect was also observed during audio–visual localization. However, audio–visual interaction involves hearing and vision that both belong to exteroception by which one perceives the outside world. During binaural–proprioceptive localization, hearing and proprioception are involved and, thus, exteroception and interoception are combined. This combination was not investigated before as far as efficient integration of bimodal sensory signals within the given context is concerned.

If auditory localization signals can guide human proprioceptive localization, as it is described in this chapter, then binaural models can also help to model the corresponding bimodal integration process, for example, the model mimicking the localization of an elevated sound source out of [3], this volume. To build a reliable computational model, of course, deep knowledge is required. That is the reason why the complex proprioceptive localization process and bimodal integration have to be studied in more detail. Furthermore, existing binaural models have to be extended, for example, to handle distance cues. New fields of applications will profit of a binaural-proprioceptive localization model. For example, such a model might help to simulate how an audio reproduction system has to be designed to guide proprioception and, thus, optimize bimodal localization accuracy within an arbitrary haptic workspace size directly in front of the human body. Thus, practical relevance exists especially for virtual environments in which the reproduction of acoustic signals can be directly controlled and the localization accuracy can be consciously influenced. Furthermore, simulations of such a model may help to reproduce suitable auditory localization signals to diminish erroneous proprioceptive space perception, for example, the radial-tangential illusion, and thus to further sharpen human precision by auditorily calibrating proprioception.

These examples make obvious, finally, that the field of possible applications of binaural models is not limited to audio-only or audio-visual scenes. The increasingly important field of virtual haptic interaction will also profit from binaural modeling algorithms.

References

1. D. Alais and D. Burr. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*, 14:257–262, 2004.
2. S. Argentieri, A. Portello, M. Bernard, P. Danès, and B. Gas. Binaural systems in robotics. In J. Blauert, editor, *The technology of binaural listening,* chapter 9. Springer, Berlin-Heidelberg-New York NY, 2013.
3. R. Baumgartner, P. Majdak, and B. Laback. Assessment of sagittal-plane sound-localization performance in spatial-audio applications. In J. Blauert, editor, *The technology of binaural listening,* chapter 4. Springer, Berlin-Heidelberg-New York NY, 2013.
4. A. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *J. Audio Eng. Soc.,* 93:2764–2778, 1993.
5. J. Blauert. Sound localization in the median plane. *Acustica,* 22:205–213, 1969.
6. J. Blauert. *Spatial hearing.* Revised Edition. The MIT Press, Cambridge, London, 1997.
7. J. Blauert and J. Braasch. Binaural signal processing. In *Proc. Intl. Conf. Digital Signal Processing,* pages 1–11, 2011.
8. J. Braasch. Modelling of binaural hearing. In J. Blauert, editor, *Communication Acoustics,* chapter 4, pages 75–108. Springer, 2005.
9. S. Brewster. Using non-speech sound to overcome information overload. *Displays,* 17:179–189, 1997.
10. F. Clark and K. Horch. Kinesthesia. In L. K. K. Boff and J. Thomas, editors, *Handbook of Perception and Human Performance,* chapter 13, pages 1–62. Willey-Interscience, 1986.
11. F. J. Clark. How accurately can we perceive the position of our limbs? *Behav. Brain Sci.,* 15:725–726, 1992.
12. C. Colwell, H. Petrie, and D. Kornbrot. Use of a haptic device by blind and sighted people: Perception of virtual textures and objects. In I. Placencia and E. Porrero, editors, *Improving the Quality of Life for the European Citizen: Technology for Inclusive Design and Equality,* pages 243–250. IOS Press, Amsterdam, 1998.
13. F. Conti, F. Barbagli, D. Morris, and C. Sewell. CHAI 3D - Documentation, 2012. last viewed on 12–09-29.
14. M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature,* 415:429–433, 2002.
15. A. Faeth, M. Oren, and C. Harding. Combining 3-D geovisualization with force feedback driven user interaction. In *Proc. Intl. Conf. Advances in Geographic Information Systems,* pages 1–9, Irvine, California, USA, 2008.
16. E. Fastl and H. Zwicker. *Psychoacoustics - Facts and Models.* Springer, 2007.
17. B. Gardner and K. Martin. HRTF Measurements of a KEMAR Dummy-Head Microphone. *MIT Media Lab Perceptual Computing,* (280):1–7, 1994.
18. T. Haidegger, J. Sándor, and Z. Benyó. Surgery in space: the future of robotic telesurgery. *Surg. Endosc.,* 25:681–690, 2011.
19. L. A. Hall and D. I. McCloskey. Detections of movements imposed on finger, elbow and shoulder joints. *J. Physiol.,* 335:519–533, 1983.
20. K. L. Holland, R. L. Williams II, R. R. Conatser Jr., J. N. Howell, and D. L. Cade. The implementation and evaluation of a virtual haptic back. *Virtual Reality,* 7:94–102, 2004.
21. G. Jansson, M. Bergamasco, and A. Frisoli. A new option for the visually impaired to experience 3D art at museums: Manual exploration of virtual copies. *Vis. Impair. Res.,* 5:1–12, 2003.

22. L. A. Jeffress. A place theory of sound localization. *J. Comp. Physiol. Psychol.,* 41:35–39, 1948.

23. L. Jones and I. Hunter. Differential thresholds for limb movement measured using adaptive techniques. *Percept. Psychophys.,* 52:529–535, 1992.

24. M. Keehner and R. K. Lowe. Seeing with the hands and with the eyes: The contributions of haptic cues to anatomical shape recognition in surgery. In *Proc. Symposium Cognitive Shape Processing,* pages 8–14, 2009.

25. V. Khalidov, F. Forbes, M. Hansard, E. Arnaud, and R. Horaud. Audio-visual clustering for multiple speaker localization. In *Proc. Intl. Worksh. Machine Learning for Multimodal Interaction,* pages 86–97, Utrecht, Netherlands, 2008. Springer-Verlag.

26. S. J. Lederman and L. A. Jones. Tactile and haptic illusions. *IEEE Trans. Haptics,* 4:273–294, 2011.

27. S. J. Lederman and R. L. Klatzky. Haptic identification of common objects: Effects of constraining the manual exploration process. Percept. *Psychophys.,* 66:618–628, 2004.

28. C. Magnusson and K. Rassmus-Grohn. Audio haptic tools for navigation in non visual environments. In *Proc. Intl. Conf. Enactive Interfaces,* pages 17–18, Genoa, Italy, 2005.

29. C. Magnusson and K. Rassmus-Grohn. A virtual traffic environment for people with visual impairment. *Vis. Impair. Res.,* 7:1–12, 2005.

30. D. Malham and A. Myatt. 3D sound spatialization using ambisonic techniques. *Comp. Music J.,* 19:58–70, 1995.

31. I. Oakley, M. R. McGee, S. Brewster, and P. Gray. Putting the feel in look and feel. In *Proc. Intl. Conf. Human Factors in Computing Systems,* pages 415–422, Den Haag, Niederlande, 2000.

32. A. M. Okamura. Methods for haptic feedback in teleoperated robot-assisted surgery. *Industrial Robot,* 31:499–508, 2004.

33. Open Source Project. Pure Data - Documentation, 2012. last viewed on 12–09-29.

34. V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.,* 45:456–466, 1997.

35. W. Qi. Geometry based haptic interaction with scientific data. In *Proc. Intl. Conf. Virtual Reality Continuum and its Applications,* pages 401–404, Hong Kong, 2006.

36. SensAble Technologies. Specifications for the PHANTOM Omni Ⓡ haptic device, 2012. last viewed on 12–09-29.

37. G. Sepulveda-Cervantes, V. Parra-Vega, and O. Dominguez-Ramirez. Haptic cues for effective learning in 3d maze navigation. In *Proc. Intl. Worksh. Haptic Audio Visual Environments and Games,* pages 93–98, Ottawa, Canada, 2008.

38. L. Shams, Y. Kamitani, and S. Shimojo. Illusions: What you see is what you hear. *Nature,* 408:788, 2000.

39. M. Stamm, M. Altinsoy, and S. Merchel. Identification accuracy and efficiency of haptic virtual objects using force-feedback. In *Proc. Intl. Worksh. Perceptual Quality of Systems,* Bautzen, Germany, 2010.

40. H. Z. Tan, M. A. Srinivasan, B. Eberman, and B. Cheng. Human factors for the design of force-reflecting haptic interfaces. *Control,* 55:353–359, 1994.

41. R. J. van Beers, D. M. Wolpert, and P. Haggard. When feeling is more important than seeing in sensorimotor adaptation. *Curr Biol,* 12:834–837, 2002.

42. F. L. Van Scoy, T. Kawai, M. Darrah, and C. Rash. Haptic display of mathematical functions for teaching mathematics to students with vision disabilities: Design and proof of concept. In S. Brewster and R. Murray-Smith, editors, Proc. *Intl. Worksh. Haptic Human Computer Interaction,* volume 2058, pages 31–40, Glasgow, UK, 2001. Springer-Verlag.

43. P. Xiang, D. Camargo, and M. Puckette. Experiments on spatial gestures in binaural sound display. In *Proc. Intl. Conf. Auditory Display,* pages 1–5, Limerick, Ireland, 2005.