A RADAR-BASED NAVIGATION ASSISTANCE DEVICE WITH BINAURAL SOUND INTERFACE FOR VISION-IMPAIRED PEOPLE

Christoph Urbanietz[1], *Gerald Enzner*[1], *Alexander Orth*[2], *Patrick Kwiatkowski*[2], *Nils Pohl*[2]

Ruhr University Bochum, Department of Electrical Engineering and Information Technology
Institute of Communication Acoustics[1], Institute of Integrated Systems[2]
Bochum, 44801, Germany
christoph.urbanietz@rub.de[1], alexander.orth@rub.de[2]

ABSTRACT

Sound is extremely important to our daily navigation, while sometimes slightly underestimated relative to the simultaneous presence of the visual sense. Indeed, the spatial sense of sound can immediately identify the direction of danger far beyond the restricted sense of vision. The sound is then rapidly and unconsciously interpreted by assigning a meaning to it. In this paper, we therefore propose an assisted-living device that deliberately stimulates the sense of hearing in order to assist vision-impaired people in navigation and orientation tasks. The sense of vision in this framework is replaced with a sensing capability based on radar, and a comprehensive radar profile of the environment is translated into a dedicated sound representation, for instance, to indicate the distances and directions of obstacles. The concept thus resembles a bionic adaptation of the echolocation system of bats, which can provide successful navigation entirely in the dark. The process of translating radar data into sound in this context is termed "sonification". An advantage of radar sensing over optical cameras is the independence from environmental lighting conditions. Thus, the envisioned system can operate as a range extender of the conventional white cane. The paper technically reports the radar and binaural sound engine of our system and, specifically, describes the link between otherwise asynchronous radar circuitry and the binaural audio output to headphones.

1. SYSTEM OVERVIEW

The goal of this work is to design a tool to support blind or visually impaired people in navigation and orientation tasks. As a general concept, we collect information about the environment that is usually recognized by the visual sense and therefore missing by a technical sensor and convert this collected and processed information to a sensation the user is able to recognize.

There are many products on the market that also use this principle. The app "The vOICe" translates a camera image into an audio signal, whereas the "Orcam" line of products analyze the camera image and translate it to meaningful spoken words. Other tools such as the "UltraCane" or "Live Braille" use ultrasonic detectors and translate object distances into vibrations. A variation of haptic output in the form of pressure on the head in combination with an ultrasonic sensor input is realized by the "Proximity Hat". Additionally, a combination of camera input and vibrating output has been investigated, e.g., by the University of Southern California. Common to all these tools is that they do not need exact maps of the environment. Instead, they rely only on the sensor input, and so they can also be used in unknown environments. For the technical sensing part of our system, we use a radar sensor, and for the output, we choose the binaural acoustic modality.

Radar sensing is a common tool in exploring unknown environments. Most modern cars are equipped with one or more radar sensors to scan across the driving direction, either to identify preceding cars and follow them in an autonomous or half-autonomous driving configuration or to identify dangerous situations such as a rapidly braking car ahead to initiate automatic emergency braking. Radar sensing brings some advantages over competing technologies such as light detection and ranging (lidar) or ultrasonic detection. Ultrasonic detection has a limited distance range and a wide detection beam. Lidar has the opposite characteristics. It can accommodate long distances and has a very pointlike steering region. Radar systems present a compromise between these two approaches. The beam can be focused, and the distance range can extend for several tens of meters. The distance range of a radar system is optimal for our purpose to extend the explorable area compared to that of a white cane. Although the lidar operational wavelength is most similar to the wavelengths used by the human visual sense, radar can extend the exploration of the environment. It can look through fog and even detect glass doors, which can be very challenging with lidar or, in some cases, even with human eyes. Furthermore, we protect the environment from harmful laser emission when using radar instead of lidar. Although the lasers used in lidar devices are claimed to be harmless to human eyes, they can destroy camera sensors as present in many devices such as cameras, smart phones, security cameras or autonomous cars.

For the output part, we use the acoustic modality; therefore, we speak of sonification. For navigation and orientation tasks, an audio channel is often used, but without the presentation of binaural cues. A car navigation system is a good example of this configuration. Although the navigation information is presented visually on a screen, there is the commonly used option to give additional acoustic navigation advice. The main reason for this apparently redundant presentation is to let the eyes focus on the street without the need to look at the screen. Although there are approaches to reduce the time of visual distraction from the street (e.g., head-up displays), the acoustic modality overcomes this issue more rigorously since it can be sensed in parallel. In addition to the factor of convenience and security over the pure visual display, in our case

https://doi.org/10.21785/icad2019.023

of blind or vision-impaired people, the sound modality is more essential to deliver cues for navigation and orientation.

A further step in convenience is the use of natural cues such as binaural localization for indicating directional information, e.g., perceiving sound from the direction to drive or walk instead of explaining the direction in words. Thus, by indicating the directional information by attributes of the sound and not the verbal articulation, we wish to reduce the sound to a more subliminal representation without verbal content for directional information. For simple instructions such as "left" or "right", this approach provides a more subconscious access and avoids unnecessary long speech and therefore might be perceived as a more pleasant hint. For more complex navigation instructions such as "slightly left" or "half right", the use of binaural sound can provide additional cues for more accurate indication of routes or obstacles. Here, the additional binaural cue will be more essential when more degrees of freedom exist for navigation, for instance, for moving in free space. As in most scenarios, the mobility of the subject is restricted to a 2-D plane (e.g., the street level); thus, we restrict the locations of the virtual sound sources to the horizontal plane, i.e., indicating only the azimuth. In this field, there is already some research, e.g., by Geranazzo et al. [1, 2], who investigated human performance in auditory navigation on virtual maps.

As an example of navigation advice, we can use short "ping" tones [3] originating from the direction of interest instead of using full words such as "30 degrees left". Theoretically, the directional information can also be coded in features other than in the natural binaural sound direction. For example, we can generate a beep-tone with the frequency of this tone or its repetition rate coding the direction, but this would not be a natural code on which decoding the brain has trained over its whole life. Therefore, it is assumed that the decoding of the frequency modulation into information specifying a direction might be a relatively difficult task. The repetition frequency is rather used to indicate the object distance, which is difficult to encode in binaural format.

With this binaural technology, we can translate directional information from the radar scan in a natural way to the acoustic sense as we virtually position sound events in a virtual acoustic environment. More precisely, we build an augmented acoustic environment, where the augmented sounds are meant to deliver additional information that is not directly accessible by the user due to the loss of the visual sense. The blind or vision-impaired user relies more than others on his or her hearing sense to manage everyday tasks. Therefore, it is important not to impede this acoustic sense by our tool. Thus, we cannot use simple headphones that would occlude the ears. Instead, we use open-fitting hearing aid technology to supply acoustic information. Moreover, we aim to extract only necessary information from the radar data to create a sparse acoustic output since we want to avoid excessive distraction.

As an interface between the radar input and audio output, we rely on a sparse representation of the data that are of interest. From the point of view of the user, only the first obstacle in each direction is of interest since it is the only object the user would run into if he or she were to head in this direction. The obstacle behind the first obstacle in a given direction will never be reached with straight motion since the user will be stopped by the first obstacle before reaching the one behind it. Therefore, we compress the radar data to a low-dimensional representation, coding only the distance to the first obstacle in every considered direction, which we term the "radar distance profile" and is constrained to only one value per azimuth direction. For this constraint, there are various

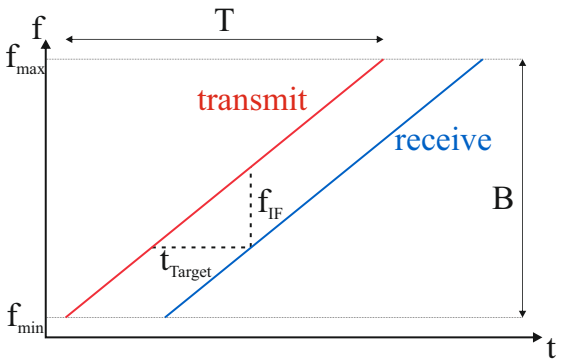


Figure 1: FMCW principle.

reasons. On the one hand, this approach immediately reduces the sound potentially delivered to the user to a more restrained level. On the other hand, enough information is provided to guarantee safe navigation when only the nearest obstacle per azimuth is indicated. The user has to avoid an obstacle no matter where it is located in terms of height. It is our task to warn the user about an obstacle in all cases, whether he or she will hit the obstacle with a foot, the torso or the head. In further implementations, there is a possibility to further indicate various heights to allow a more customized reaction depending on the type of obstacle: In the case of a low door, the user can simply cower, but in the case of a wall, the user knows that there is no way to go through the obstacle.

The remainder of this paper is organized according to the signal flow in the presented assistance tool. In Sec. 2 the utilized radar technology is presented, and the extraction of the radar distance profile is explained. Sec. 3 then presents the extraction of meaningful sound events from this distance profile, followed by Sec. 4, which gives a deeper insight into the technology of binaural rendering. Sec. 5 reports fundamental aspects of the actual implementation of the system.

2. RADAR SYSTEM

2.1. FCMW Radar Concept

Frequency-modulated continuous wave (FMCW) radar systems use a continuously radiated signal to spread the output power over a period of time, which makes FMCW radar systems easier and cheaper to manufacture than classical pulsed radar systems. The use of a linear frequency ramp enables precise target detection and localization [4, 5].

The FMCW principle is visualized in Fig. 1. A single-frequency radio frequency (RF) signal is generated and radiated by the radar device. This signal's frequency is raised with a linear frequency sweep over a frame of time T. The frequency range covered by this sweep is called the bandwidth B. This radiated RF signal is reflected by one or multiple targets, and the reflection is picked up by the sensor. By mixing the signal (multiplying the momentary signal amplitudes) that is being sent with the reflected one, a so-called difference signal of intermediate frequency f_{IF} is generated. This intermediate frequency relates directly to the distance between the sensor and the reflecting target via the bandwidth and sweep duration.

In a real scenario, the generated intermediate-frequency signal

is sampled with an analog-digital converter. On these data, a fast Fourier transform is performed to compute the amplitude of the signal's frequency components and therefore the target reflections. Therefore, every peak location in this frequency domain representation corresponds to a real target reflection. The distance to a target for a corresponding intermediate frequency f_{IF} is given by:

$$R = \frac{c \cdot t_{\mathrm{Target}}}{2} = \frac{T}{B} \cdot \frac{c \cdot f_{\mathrm{IF}}}{2} \tag{1}$$

where c is the speed of light in the relevant medium.

The maximum distance R_{max} from which a target can be detected is given by (1) with a maximum f_{IF} where the Shannon Nyquist theorem is valid with the sampling rate f_{S} of the analog-digital converter used:

$$R_{\mathrm{max}} = \frac{T}{B} \cdot \frac{c \cdot f_{\mathrm{S}}}{4} \tag{2}$$

The resolution ΔR of an FMCW radar system is solely defined by the bandwidth B used by the system. The resolution is defined as the minimum distance between two equally strong target reflections. If these two targets are closer than the given minimum distance, their peaks in the frequency-domain representation merge into one peak, which makes the two reflections indistinguishable. This distance is based on the 3 dB beamwidth of a target reflection in the frequency domain and is given by:

$$\Delta R = \frac{c \cdot A_{\mathrm{W}}}{2B} \tag{3}$$

where A_{W} is a widening factor based on the windowing function used before FFT computation.

2.2. 2D Scanner

The radar sensor used in this work is an 80 GHz FMCW radar sensor developed at Ruhr-Universität Bochum in collaboration with Fraunhofer FHR. This sensor is capable of extremely precise measurements [6, 7] with a very high bandwidth of up to 25.6 GHz. Its ellipsoid PTFE lens gives the sensor a 3 dB beamwidth of 5°.

To expand this linear measurement system, a rotating metallic mirror is used to steer the radar beam in the azimuthal direction. The mirror is angled 45° to the rotational axis as shown in Fig. 2. This operation deflects the radar beam orthogonally to the rotational axis. The deflecting mirror is rotated by a Trinamic PD42-1141 stepper motor with an integrated controller. For a stable and reproducible measurement with each revolution, a hardware-controlled trigger for the radar system was chosen. The trigger is based on a Hall sensor with an integrated comparator circuit that generates a trigger signal each time a magnet fixed to the rotating mirror enters a specific distance to the Hall sensor. For the compensation of movement of the scanner system between measurements, an inertial measurement unit (IMU) (Bosch BNO055) has been used to correct possible rotational changes. Because of the positioning of system components, the azimuthal range covered is reduced to 270°.

The scanner system is controlled by a Raspberry Pi 3B, which is connected to an external PC by a WiFi connection. The PC is used to control the system configuration parameters and to collect and process the scanner data. In the scanner system, radar and IMU data are collected and sent to the PC as UDP data packages via the Raspberry Pi and WiFi. This process is controlled with a Python script, and data types are conserved over the transmission.

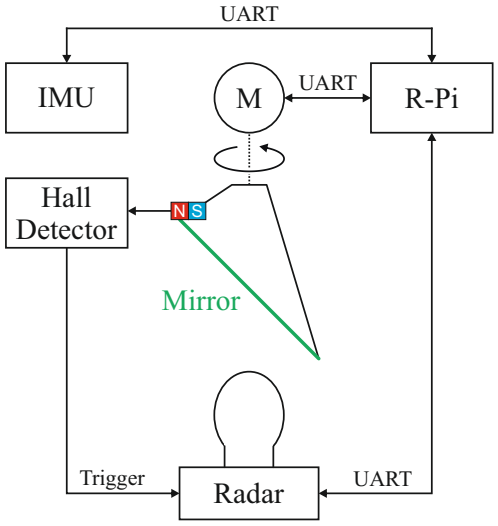


Figure 2: 2D Scanner block diagram.

2.3. Processing

The radar sensor sends its measurement data as a consecutive byte stream once per revolution. This data stream is converted to a two-dimensional data matrix in 16 bit integer format, where one dimension represents the frequency sweeps and therefore the azimuth dimension and the other the captured IF signal. On the IF dimension, a Hann window is applied to suppress sidelobes in the range. On the windowed data, an FFT is performed to represent the data in the frequency domain and therefore the range domain. Because of the high dynamic range of the signal in terms of amplitude, the data are then converted to their logarithmic magnitude.

On these data, target detection can be performed. A static threshold detection algorithm is not suitable for radar measurements because of the high dynamic range of targets as well as clutter. Dynamic threshold algorithms are used called CFAR (constant false alarm rate) [8]. The specific form of algorithm used is OS-CFAR (ordered statistics, Fig. 3), which has been shown to be very robust [9]. This algorithm is used to detect the closest target in each direction. If no target can be detected in a direction, the maximum detectable range of the radar measurement is assumed. This range profile data is then corrected by the IMU data to provide range profile data corresponding to the world coordinate system.

Examples of these radar measurements and the extracted radar distance profile (red) are shown in Fig. 4a. Additionally, we plotted the ground truth positions of the walls (black) into the figure. The measurement was performed during a stepwise walk through a hallway. The discrete positions where a measurement was taken are denoted by R1 to R8. Fig. 4a shows the measurement at positions R3, R6 and R8. In addition to the walls, we put two metal stands as additional obstacles, O1 and O2, into the hallway.

We see that the walls in most cases are well recognized, although they seem to consist of single points. Indeed, the walls in this environment are built in a lightweight construction, and predominantly the girders are seen by the radar system. Caused by the azimuthal spread of the radar beam, the girders are smeared in this direction. Nevertheless, the essential contours and obstacles can be recognized. The middle dataset at R6 gives an example of

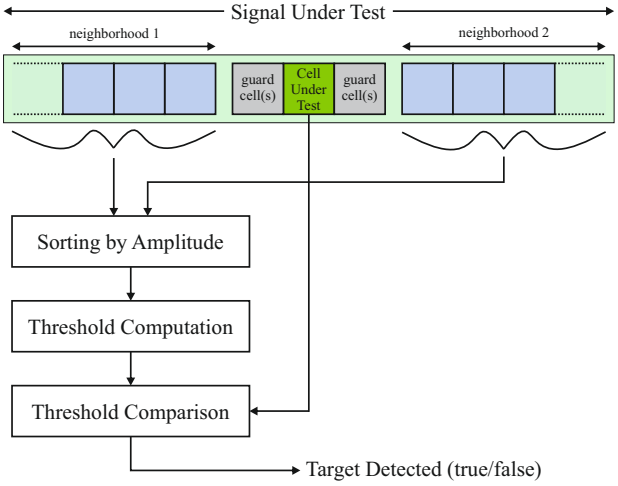


Figure 3: In OS-CFAR, each cell value in a signal under test is iteratively tested if it surpasses a threshold value computed specifically for this cell under test. A neighborhood window above and below the cell under test is selected, with optional guard cells between the cell and the neighborhood. The cell values from the neighborhood windows are then sorted by amplitude, and the value of a chosen rank, for example, the second largest value, is selected as the threshold. Upon this value, scaling factors and a bias can be applied to customize the sensitivity to the given scenario The value of the cell under test is then compared to the computed threshold.

misrecognition. Although there is a solid wall on the left side of the user, the radar sees the nearest obstacle far behind the wall. In this particular case, there was a poster wall made from plain metal at this position. This is a nearly perfect mirror for the radar beam, and therefore, the mirrored wall at the opposite site was recognized as the next obstacle in this direction.

3. FEATURE EXTRACTION

Now that we have the radar distance profile extracted from the measurement, in this section, we describe our approaches to select the information that should be sonified out of the radar distance profile. For the current implementation, we use simple approaches with the aim of a sparse output that can be easily interpreted by the user. In these modes of operation, we do not claim to deliver a comprehensive picture of the environment but only a sparse and helpful augmentation of the natural acoustic environment.

3.1. "Nearest Obstacle" Mode

The main purpose of the first implemented mode is to prevent the user from running into an obstacle. Therefore, the "nearest obstacle" is sonified but only in the case that the "nearest obstacle" is closer than a certain limit. Therefore, the sonification is silent if there is no need for the user to react. More precisely, the radar distance profile is weighted by the viewing direction, and then the weighted minimum is sought, and sound is created from that direction if the target is closer than this certain limit. The distance weighting is applied since both the average and the maximum velocity of human subjects are larger in the viewing direction than in the lateral directions. Therefore, an obstacle at a distance of one meter in front of the user poses more danger than an obstacle at a distance of one meter on the side of the user. The weighting of the distance is performed by

$$\widetilde{d}(\phi) = d(\phi) \cdot (1 + \alpha \sin(|\phi|)) \tag{4}$$

where $d(\phi)$ is the unweighted radar distance and $\widetilde{d}(\phi)$ is the weighted distance in the direction ϕ. Here, $\phi = 0$ is the viewing direction. Therefore, the distance in the viewing direction is not affected by the weighting. Distances on the side are suppressed from sonification since they are projected to longer distances up to a factor $(1 + \alpha)$. The strength of the suppression can be adjusted by the parameter α between 0 and $+\infty$, where 0 means no suppression and ∞ means complete suppression of side obstacles.

In addition to the direction of the nearest obstacle, which is coded by the natural binaural cue, the distance to the obstacle is coded by the repetition rate of the sound, a short "ping" tone as it is commonly used, for example, in parking assistants in cars. A faster repetition means a nearer obstacle and therefore more danger. This association is very natural since the more frequent sound is more imposing and therefore draws more attention as the obstacle grows closer and the danger becomes greater.

Another use case of this very simple mode is a movement along a wall. Sometimes there is the need to walk in parallel along a wall, e.g., if the user is walking down a long hallway. Looking parallel to the wall, the distance to the wall has its minimum at plus or minus 90 degrees. Therefore, we have to hold the orientation in a way that renders the sound laterally to move along the wall without hitting it. The distance to the wall can be easily controlled by the repetition rate of the sound. Although the weighting of the distance will change the effective distances in such a way that the distance at $\pm$ 90 degrees becomes larger, the minimum of the weighted distance will still be at $\pm$ 90 degrees in this scenario. This condition is assured by our choice of the weighting function. To prove this claim, let us assume that the wall is at the left of the user without loss of generality. Then, the minimum of $d(\phi)$ is at +90 degrees looking parallel along the wall. Let us denote this minimum distance by d_1, and hence, the weighted distance in the direction of +90 degrees is

$$\widetilde{d}_1 = d_1 \cdot (1 + \alpha). \tag{5}$$

The raw distance to the wall in any other direction is

$$d_2(\phi) = d_1 / \sin(\phi). \tag{6}$$

The weighted distance in the ϕ direction is then given by

$$\widetilde{d}_2(\phi) = d_2(\phi) \cdot (1 + \alpha \sin(\phi)) \tag{7}$$

$$= d_1 \frac{1 + \alpha \sin(\phi)}{\sin(\phi)} \tag{8}$$

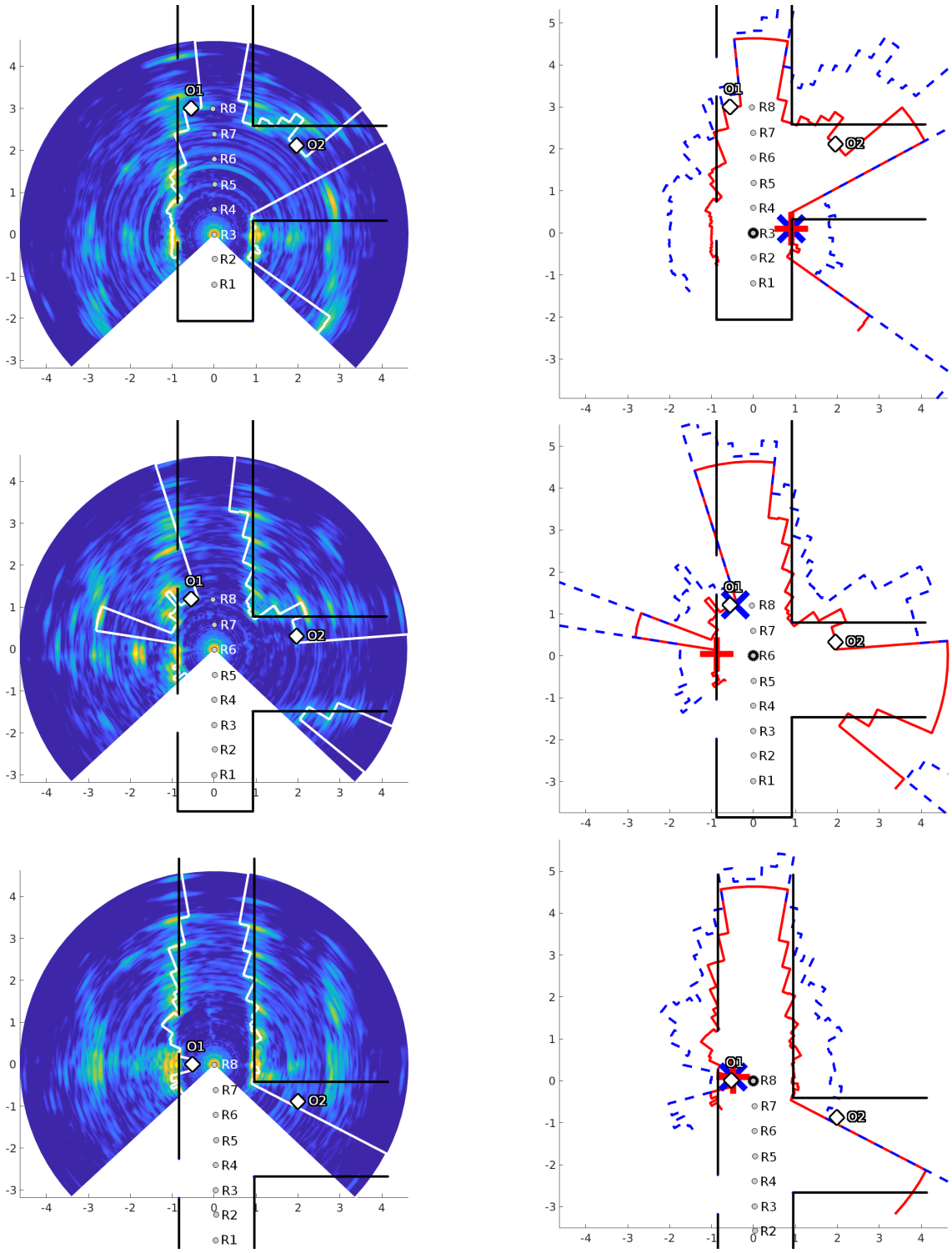
$$= \widetilde{d}_1 \frac{1 + \alpha \sin(\phi)}{(1 + \alpha) \sin(\phi)} \tag{9}$$

and the ratio

$$\frac{\widetilde{d}_2(\phi)}{\widetilde{d}_1} = \frac{1 + \alpha \sin(\phi)}{(1 + \alpha) \sin(\phi)} \tag{10}$$

is always larger or equal to 1 in the area $0^\circ < \phi < 90^\circ$.

Fig. 4b shows the weighted distance with $\alpha = 1$ together with the unweighted radar distance profile. The nearest obstacle based



(a) Radar data and extracted distance profile (white) with ground-truth walls (black) and obstacles (O1, O2)

(b) Weighted (blue dashed) and unweighted (red) distance profile, weighted (blue x) and unweighted (red +) "nearest obstacle"

Figure 4: Examples of measured and processed radar data and user positions from top to bottom: R3, R6, and R8

on the weighted distance profile is denoted by a blue **x**, while the nearest obstacle without weighting is denoted by a red **+**. In many cases, they coincide, but in the case where the user is located at R6, the weighted mode sonifies the obstacle O1, which stands in front of the user and probably represents more danger than the wall at the side where the unweighted distance has its minimum.

3.2. "Ahead Distance" Mode

As a second mode in the current state of development, we implemented a sonification process that gives the user the distance in the gaze direction. Thus, the user is able to scan the room on his or her own volition. The benefit of this "self-operated" mode compared to a comprehensive presentation of the whole environment is the following:

- The user is able to scan the environment at a speed that he or she is able to process the sound stimulation.
- The speed of scanning can vary depending on the complexity of the part of the environment being considered.
- The user can easily select which area is of interest.
- The sound is rather sparse and easy to interpret.
- The distance can be coded in the same way as in the sparse "Nearest Obstacle" mode.

Although the directional coding of the sound is not as essential as in the "Nearest Obstacle" case, we will also use the binaural rendering engine for this mode. In particular, we do so to assure that one single sound is locked to the virtual acoustic environment and does not move in the static environment when the user turns his or her head. This is one aspect of making the virtual acoustic augmentation more realistic.

Since the sound is continuously playing in this mode, the mode is meant only for active exploration purposes. It can be manually switched on by the user if he or she decides to actively look around. The "Nearest Obstacle" mode instead is intended to be an always-on tool that automatically turns silent if not needed but appears automatically if something of interest is happening, i.e., if an obstacle comes too close to the user.

4. BINAURAL RENDERING

In this section, we give a short introduction to binaural rendering using headphones. As an approximation of the sound propagation from a real sound source to the ear of a human, a linear time-invariant (LTI) system can be assumed as long as the sound source and the head have static positions. For slow motions of a walking listener, the approximation is quite precise. The LTI system from the sound source to the left and right ear is called the head-related transfer function (HRTF) or, in the time-domain, the head-related impulse response (HRIR). The influence of the room as reflections of the sound from walls is not part of the HRTF. In particular, the HRTF is often defined as the difference between the sound pressure at the ears and a hypothetical pointlike pressure receiver located at the center of the head [10, 11, 12]. As this is in general a noncausal system, because one ear is almost always closer to the source than the center of the head, an additional delay to the HRIR is appended in technical use to ensure causality. Below, we use the terms HRTF and HRIR for the causal form of the transfer function.

The HRTF can be measured from either a human being or a dummy head [12, 13, 14] or calculated from a model [15, 16, 17, 18]. The accuracy of the HRTF is essential to ensure localization precision. Obviously, the HRTF depends on the relative orientation of the head to the sound source. The impact of the distance can be neglected in the far field, except for an additional delay. Therefore, we need two coordinates to describe the orientation of the head relative to the sound source. Usually, a coordinate system of azimuth ϕ and elevation θ is used, as shown in Fig. 5. In this work, we constrain our discussion to the horizontal plane only with fixed elevation due to the given sonification task.

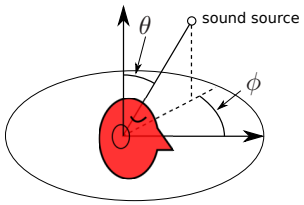


Figure 5: Head-related coordinate system

The HRTF can be utilized to create virtual positioned sound sources using headphones [12, 19, 20, 21, 22] or, as intended in our project, hearing aids, simulating the sound pressure of the virtual sounds at the ears. To position a sound source signal s virtually at a desired position of interest, we pass the signal through the corresponding transfer function to create the output signals y_l and y_r for the left and right ear, respectively. In the time domain, this operation can be performed by a convolution with the corresponding HRIR, i.e., $y_l = s * h_l(\phi, \theta)$ and $y_r = s * h_r(\phi, \theta)$, where ϕ and θ designate the position of the sound source relative to the head. In a real-time environment, the convolution is usually accomplished bufferwise and often performed by fast FFT convolution.

In many practical use cases and in our own use case, the relative sound source position to the head is dynamic, either because a source is moving or, as in our case, the head is rotating. For the latter, head tracking is needed [23] to unlock the virtual sound field from head rotation. Therefore, we no longer have an LTI system. Common practice, however, is to assume a piecewise, i.e., a bufferwise, LTI system. These buffers have to employ varying HRTF filters, and their outputs are crossfaded [24, 25]. This approach can lead to artifacts, especially if the effectively crossfaded HRTF filters differ massively from each other, e.g., when the head orientation changes strongly within one buffer (fast movements) or if the spatial resolution of the HRTF is too coarse in general. Additionally, the total system latency (TSL) for the compensation of head rotation is important for realistic perception [26, 27]. An approach to overcome some of these issues is to render the binaural sound samplewise as described, e.g., in [28, 29].

5. PROCESSING IMPLEMENTATION

As all parts of our assisting device are now known in theory, we present some important aspects of the end-to-end implementation. The main process of creating binaural audio from the radar profile is split into two parts. The first part, the actual sonification algorithm, analyses the radar profile and creates an acoustic scene description from it. In particular, the algorithm creates monaural data containing the sound source signal together with the desired angular position of that sound. The second part is the binaural renderer that produces binaural output from these elements.

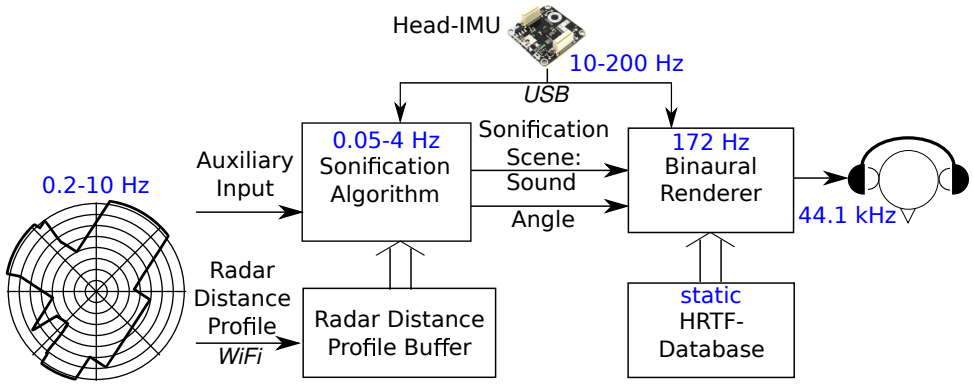


Figure 6: Sonification processing scheme

5.1. Multirate System

Since the sonification process involves many system components, we have to address the various time bases. Fig. 6 shows the blockwise implementation of the audio part and its interface with the other modules of the system. The whole system consists of asynchronous blocks. Whenever radar scanning is available in the form of a radar distance profile, this profile is delivered to the sonification engine. Depending on the type of radar sensor and the mode of operation, the update rate can be between 0.2 Hz and 10 Hz. The IMU delivers updates of the head orientation with an update frequency between 10 Hz and 200 Hz, depending on the IMU sensor. Both the radar profile input and the IMU input are not designed to deliver new data at a constant update rate; rather, the update rate can vary substantially. At the output, we need an audio output stream with a constant sampling frequency, in our case 44100 Hz. At least in our implementation on a Raspberry Pi computer, we apply blockwise audio processing since the platform does not offer enough performance for samplewise real-time processing of binaural audio. On the Raspberry Pi, we use a block length of 256 samples for the audio processing. Thus, an audio block is processed with a repetition rate of 172 Hz, which is fixed due to the fixed audio output sampling frequency. Instead, the sonification block can create acoustic scene descriptions with various durations. A sonification scene can last for a very short time, e.g., if we have a single beep tone that denotes only the nearest obstacle, or can last for very long time, e.g., if we have an algorithm that describes the whole environment. Hence, the sonification block has a variable rate between 0.05 Hz and 4 Hz. Also, in the current implementation, where we have the “Ahead Distance” mode, for instance, there are various sonification scene durations within one sonification mode. One sonification scene, in this case, consists of the “ping” tone with a constant length and a pause that varies in its length depending on the distance to the obstacle.

5.2. Dealing with Asynchronicity

To accommodate these various sampling frequencies in a single system, we use buffers in every connection between the blocks. The radar profile buffer plays a special role because it is the connection between the sonification and radar acquisition. This buffer is used to store the last received radar distance profile to deliver this profile to the subsequent function blocks at any arbitrary time and in the presence of a link failure. The connection between the IMU and binaural rendering engine is similar. The buffer between the sonification and binaural rendering is different since the buffer is a first-in-first-out (FIFO) buffer. Every sound sample is delivered only once through the buffer, and the buffer has two main tasks. On the one hand, it compensates for the difference between the scene length coming out of the sonification block and the audio buffer size used by the binaural renderer. On the other hand, it can deliver data on demand to the binaural renderer while the sonification block is calculating a new sonification scene. This feature is important since the analysis of the radar data and the creation of the sonification scene may take longer than the duration of one audio buffer. Every output scene of the sonification block should be output by the binaural rendering, and the sonification block creates a new sonification scene whenever the delivery of the previous scene to the binaural renderer has started and is paused after creating this scene until it is triggered again. The two blocks operate in independent time bases and are synchronized by this procedure. Indeed, it would be possible to run them completely asynchronously and let the sonification block produce as many scenes as it can. The buffer would then have the task to deliver only whole scenes to the binaural renderer and would always start with the most recent scene. This approach would prevent the buffer from underrunning if the processing of a sonification scene would take longer than the duration of the previous sonification scene. Nevertheless, we did not use this fully asynchronous approach since it would increase the processing costs to a maximum. In our synchronous approach, we simply add silence in the case of a buffer underrun.

6. CONCLUSION

In this paper, we presented a concept of an assisting device for vision-impaired people for navigation and orientation. The device is based on radar input and binaural audio output and was implemented as a research study. The quality of the radar data acquisition was shown in examples. We demonstrated two sonification modes, representing the essence of first subjective preferences from blind and seeing people, who demand a simple interpretability and a sparse sound for an easy-to-access utility. The latter aspect is addressed by restricting the acoustic indications to just the horizontal plane in both presented modes and by paying particular attention to the walking direction of the user in the “nearest obstacle” mode using the weighted radar distance profile. The device is

meant to provide orientation cues additional to, e.g., a white cane and support the user in everyday orientation and navigation tasks. Further end-to-end investigations of the presented system have to be performed with various users to evaluate the helpfulness in realistic scenarios. Depending on these results, we can further tune the modes and algorithms to deliver a more satisfying experience.

7. ACKNOWLEDGMENT

This work is supported by the European Regional Development Fund Nr. EFRE-0800372 NRW grant, LS-1-1-044d as part of the project "Ravis 3D".

8. REFERENCES

[1] M. Geronazzo, A. Bedin, L. Brayda, C. Campus, and F. Avanzini, "Interactive spatial sonification for non-visual exploration of virtual maps," *Int. J. of Human-Computer Studies*, vol. 85, pp. 4–15, 2016.

[2] M. Geronazzo, F. Avanzini, and F. Fontana, "Auditory navigation with a tubular acoustic model for interactive distance cues and personalized head-related transfer functions," *J. on Multimodal User Interfaces*, vol. 10, no. 3, pp. 273–284, Sep 2016.

[3] C. Urbanietz and G. Enzner, "Binaural Rendering for Sound Navigation and Orientation," in *2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, Mar. 2018, pp. 1–5.

[4] M. A. Richards, *Fundamentals of Radar Signal Processing*. McGraw-Hill Education, 2014.

[5] M. I. Skolnik, *Radar Handbook*. McGraw-Hill Education, 1990.

[6] N. Pohl, T. Jaeschke, and K. Aufinger, "An Ultra-Wideband 80 GHz FMCW Radar System Using a SiGe Bipolar Transceiver Chip Stabilized by a Fractional-N PLL Synthesizer," in *2012 IEEE Transactions on Microwave Theory and Techniques*, vol. 3, Mar. 2012, pp. 757–765.

[7] N. Pohl, T. Jaeschke, S. Scherr, S. Ayhan, M. Pauli, T. Zwick, and T. Musch, " Radar measurements with micrometer accuracy and nanometer stability using an ultra-wideband 80 GHz radar system," in *2013 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet)*, Jan. 2013.

[8] H. M. Finn and R. S. Johnson, "Adaptive detection mode with threshold control as a function of spacially sampled clutter-level estimates," in *RCA Review*, vol. 29, Sept. 1968, pp. 141–464.

[9] H. Rohling, "Ordered statistic CFAR technique - an overview," in *Radar Symposium (IRS), 2011 Proceedings International*, vol. 7, Sept. 2011, pp. 631–638.

[10] V. Pulkki and M. Karjalainen, *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. Wiley, 2015.

[11] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display: Second Edition*. J. Ross Publishing, July 2013.

[12] J. Blauert, *Spatial Hearing - The Psychophysics of Human Sound Localization*, rev. ed. Cambridge: MIT Press, 1997.

[13] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Inter-laboratory round robin HRTF measurement comparison," *IEEE J. of Selected Topics in Signal Process.*, vol. 9, no. 5, pp. 895–906, Aug 2015.

[14] G. Enzner, C. Antweiler, and S. Spors, "Trends in acquisition of individual head-related transfer functions," in *The Technology of Binaural Listening*, J. Blauert, Ed. Springer, 2013, pp. 57–92.

[15] K. Young, T. Tew, and G. Kearney, "Boundary element method modelling of KEMAR for binaural rendering: Mesh production and validation," in *Interactive Audio Systems Symposium*, September 2016.

[16] L. Bonacina, A. Canalini, F. Antonacci, M. Marcon, A. Sarti, and S. Tubaro, "A low-cost solution to 3D pinna modeling for HRTF prediction," in *IEEE Int. Conf. Acoust., Speech and Signal Process.*, March 2016, pp. 301–305.

[17] C. T. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, A. I. Tew, C. Hetherington, and J. Thorpe, "Creating the Sydney York morphological and acoustic recordings of ears database," *IEEE Trans. on Multimedia*, vol. 16, no. 1, pp. 37–46, Jan 2014.

[18] F. Brinkmann, A. Lindau, S. Weinzierl, S. v. d. Par, M. Müller-Trapet, R. Opdam, and M. Vorländer, "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations," *J. Audio Eng. Soc*, vol. 65, no. 10, pp. 841–848, 2017.

[19] D. R. Begault, *3D Sound for Virtual Reality and Multimedia*. San Diego, CA, USA: Academic Press Professional, Inc., 1994.

[20] K. Sunder, J. He, E. L. Tan, and W. S. Gan, "Natural sound rendering for headphones: Integration of signal processing techniques," *IEEE Sig. Process. Mag.*, vol. 32, no. 2, pp. 100–113, March 2015.

[21] F. Rumsey, *Spatial Audio*. Focal Press, 2001.

[22] S. Carlile, *Virtual Auditory Space: Generation and Applications*. Landes Bioscience, 1996.

[23] V. R. Algazi and R. O. Duda, "Headphone-based spatial sound," *IEEE Signal Processing Mag.*, vol. 28, no. 1, pp. 33–42, Jan 2011.

[24] A. Kudo, H. Hokari, and S. Shimada, "A study on switching of the transfer functions focusing on sound quality," *Acoustical Sci. and Techn.*, vol. 26, no. 3, pp. 267–278, 2005.

[25] M. Vorländer, *Auralization (RWTHedition)*. Springer, 2007.

[26] D. S. Brungart, B. D. Simpson, and A. J. Kordik, "The detectability of headtracker latency in virtual audio displays," in *Int. Conf. Auditory Display (ICAD)*, 2005, pp. 37–42.

[27] A. Lindau, "The perception of system latency in dynamic binaural synthesis," *Proc. of 35th DAGA*, pp. 1063–1066, 2009.

[28] J. W. Scarpaci, H. S. Solburn, and J. A. White, "A system for real-time virtual auditory space," in *Int. Conf. on Auditory Display*, July 2005.

[29] C. Urbanietz and G. Enzner, "Binaural Rendering of Dynamic Head and Sound Source Orientation Using High-Resolution HRTF and Retarded Time," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Apr. 2018, pp. 566–570.