ISSN 2464-4617 (print)
ISSN 2464-4625 (online)
Computer Science Research Notes - CSRN 3301
http://www.wscg.eu
WSCG 2023 Proceedings

Error-Robust Indoor Augmented Reality Navigation: Evaluation Criteria and a New Approach

Oliver Scheibert
TH Köln
Steinmüllerallee 1
51643, Gummersbach, Germany
scheibert.oliver@gmail.com

Jannis Möller Steve Grogorick Martin Eisemann
TU Braunschweig
Mühlenpfordtstr. 23
38106, Braunschweig, Germany
{moeller,grogorick,eisemann}@cg.cs.tu-bs.de

ABSTRACT

Tracking errors severely impact the effectiveness of augmented reality display techniques for indoor navigation. In this work we take a look at the sources of error and accuracy of existing tracking technologies. We derive important design criteria for robust display techniques and present objective criteria that can be used to evaluate indoor navigation techniques without or in preparation of quantitative user studies. Based on these criteria we propose a new error tolerant display technique called Bending Words, where words move along the navigation path guiding the user. Bending Words outranks the other evaluated display techniques in many of the tested criteria and provides a robust, error-tolerant alternative to established augmented reality indoor navigation display techniques.

Keywords

Augmented Reality; Robust Indoor Navigation; Display Techniques; Perceived accuracy; Tracking errors;

1 INTRODUCTION

Augmented Reality (AR) device tracking systems are not perfect and errors can accumulate over time enforcing cognitive compensation of the user [MMC00]. In this work we compare the robustness of AR display techniques (interfaces) for indoor navigation to provide useful navigation information even in the presence of tracking errors.

AR provides location-aware user experience by overlaying spatially registered, digital information on a screen for real-time interaction with the physical and virtual environments [BCL15]. An important application scenario is pedestrian navigation. With recent advances in user tracking technologies and sufficient processing power of modern smartphones, the more challenging indoor navigation has become feasible [MKH$^+$12].

However, the technology is still in its infancy and reliability for an extended amount of time has not been achieved [YNA$^+$17,FPS$^+$20]. McIntyre et al. stated that the problem of accumulation errors within a tracking system will not be solved in the near future [MCJ02]. And 20 years later he is still right. Relying on such a faulty system results in digital objects being far off their supposed position. One way to tackle this problem is to increase the user's awareness of the tracking imperfections using different display techniques [PDCK13], e.g., by using 3D arrows that can change in color and shape, or similar feedback. Error visualization generally improves AR navigation systems but it is challenging to design suitable visualizations [PDCK13]. Even worse the system might not be aware of the tracking errors, lulling the user into false security.

In order to reduce the *impact* of tracking errors on indoor navigation instead of only making users aware of it, we contribute the following: We first define the problem of uncertain tracking errors in the context of tracking systems (section 2). Next, we classify AR navigation display techniques (section 3), and investigate typical representatives. We present objective criteria to evaluate AR display techniques with regard to error-robustness (section 4) and evaluate the representatives (section 5). In section 6, we propose our own display technique Bending Words and discuss (section 7) its advantages.

2 RELATED WORK

Pedestrian navigation services have gained attention for several years now [MRBT03,RC17]. They evolved from 2D paper maps to digital maps on mobile devices to location-based turn-by-turn instructions [Kim10]. Modern positioning systems enable AR systems to guide the user in real-time. These systems require a very high accuracy to correctly display information and avoid confusion and misguidance of the user. Yet, there is still no generally accepted solution for localization systems [MKH$^+$12], e.g., Adler *et al.* who analyzed and categorized 183 indoor localization techniques published between 2010 and 2014 [ASWK15].

A user generally expects accuracy to be perfect or at least good enough to rely on the displayed information which is often a mistake and may lead to navigation failures. Visualization techniques that adapt to the amount of tracking errors have the power to decrease the impact of inaccurate tracking solutions [PDCK13]. To understand why tracking solutions are always imperfect it is important to take a look at the underlying types and sources of error.

Localization

Localization is the process of tracking a user's position or more precisely the position of the device used for navigation. Applications for localization include pedestrian navigation, robotics, dynamic personalized pricing, product placement, advertisement, fleet management or intelligent spaces [YNA+17, LLY+15].

Localization systems rely on physical properties of signals, e.g. speed of light, or other measurable forces, such as earth's magnetic field. Sensor and information fusing may improve the overall performance, e.g. Wi-Fi and magnetic signals [SBS+15] or incorporating a priori knowledge, such as a map of the environment, to make the localization systems more robust and accurate [HFH04].

Despite constant improvements in localization techniques, a perfect solution seems almost impossible. Development times for indoor navigation systems are often several years and currently might not result in a widely accepted solution which has the high precision required to accurately display AR content within the camera feed [LLY+15, MSTSP+21, GFW21].

The cognitive load posed on the user of indoor navigation systems correlates with the accuracy of localization. This effect is especially prevalent in AR applications where the wrong positioning of visual media within the camera feed becomes distracting at best or misleading at worst [MCJ02].

Generally, two types of error exist: Rotational error, the angular difference between the direction to the assumed position of the next waypoint and the direction to the actual next waypoint; and translational error, the distance between the current position and the assumed position of the tracking device (Figure 1).

Tracking Systems

Different tracking systems have different sources and degrees of errors. Signal-based localization techniques (Wi-Fi, GPS, Bluetooth, etc.) are more prone to errors caused by changes in the environment. Cluttered

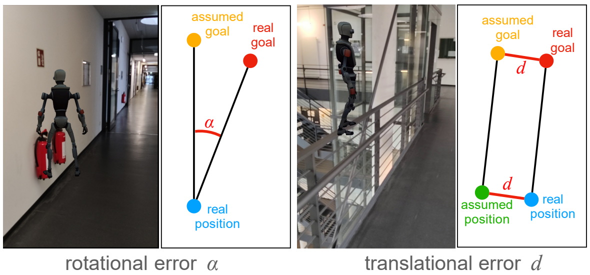


Figure 1: Types of error: **Rotational errors** (left) result in misleading direction information as the tracking system's assumed goal and the real goal differ. **Translational errors** (right) result in irritating positioning of the augmented content due to disparity between the assumed position and the real position as well as assumed goal and real goal of the tracking device.

Tracking System	Accuracy
Ultra-wideband Positioning [YNA+17]	1 cm - 0.3 m
Wi-Fi Based Positioning [YS15]	1 - 10 m
Magnetic Positioning [SBS+15]	1 - 8 m
Global Positioning System [LGD+15]	1 - 10 m
Bluetooth [RLJ+15]	0.5 - 10 m
Vision-Based Positioning [KJ08]	resolution dependent
Pedestrian Dead Reckoning [KH15]	distance dependent

Table 1: Accuracy of user tracking systems

environments introduce multipath, non-line-of-sight, and shadowing artifacts that affect either the arrival time, angle, or strength of a signal reaching the sensor [RC17, YNA+17]. Even the human body affects tracking accuracy [APM+16].

In Table 1 we provide an overview of several tracking techniques commonly found in pedestrian navigation systems and list typical accuracy ranges.

AR applications require a high precision tracking in order to display the virtual content correctly. This makes vision-based positioning systems currently the only alternative. These systems mostly rely on feature point tracking with one or more cameras and/or depth information [KJ08]. Due to their computational demands vision-based techniques are often coupled with faster techniques [BEP15]. While lighting conditions, surface properties (reflecting/refracting), and occlusion may negatively impact tracking precision [DRMS07], accuracy of vision-based positioning systems is theoretically only limited by image resolution.

Other exotic positioning systems utilize sound, light beacons, FM radio, or RADAR, to localize a user within a building but are rarely used in practice [YNA+17].

3 AR NAVIGATION DISPLAY TECHNIQUES

Commonly all AR indoor navigation display techniques require the following input:

1. The path from the current position towards the goal, usually provided as a list of 3D coordinates

2. The geometry of the building, including walls, doors, stairs, or elevators.

3. The user's pose that represents their position and rotation within the building.

Using this information the user gets feedback where to head next in order to reach their goal.

Besides the technical accuracy of these tracking systems, the way navigation information is presented has a strong impact on how accuracy of the navigation system is perceived by the user [PDCK13]. Misaligned or constantly jittering visual elements are distracting at best or misleading at worst [MCJ02]. Adapting how these instructions are delivered strongly increases performance of users when navigating through unknown environments [RC17]. But instead of interface design, the focus in indoor navigation applications is mostly put on their localization techniques.

The user's expectation of the system influences how they perceive it. If a technique is usually perceived as very accurate, based on past experiences, an inaccurate tracking will be overly distracting [MBMH01]. Whereas a less common technique which makes an imperfect tracking state less obvious could more easily prepare the user for the actual experience when using the system [MMC00].

Extending on Pankratz *et al.* [PDCK13] we distinguish three categories of AR navigation display techniques:

1. *Discrete information* which shows navigation hints as one or a series of next steps;

2. *Guiding information* which shows only the direction towards the next waypoint;

3. *Context information* which shows also the area around the user in an exocentric view.

Within these we found several representative techniques, see Table 2. Examples for the display techniques emphasized within the table are shown in Figure 2.

Discrete Information	**Lines on the Ground** Waypoint Marker **Bending Words**
Guiding Information	Guiding Arrow Shining Light **Digital Avatar** **Haptic Feedback**
Context Information	Paper Map **World in Miniature**

Table 2: Display technique examples. We evaluate the representatives most resilient against tracking errors (marked in bold) in section 5, and present Bending Words in (section 6).

Discrete Information

Discrete Information display techniques provide information about the next steps along the path at any time. A common example is *Lines on the Ground* that lead towards the goal. Instead of providing a continuous path, *waypoint markers* display only the next corner. These techniques are prone to tracking errors as they do not provide much context information to allow the user to compensate for the error. A typical example are non-aligned waypoints due to rotational error. A line on the ground makes it sometimes hard to guess where the systems wants the user to go if the line is off it's intended direction (Figure 2, top left).

Guiding Information

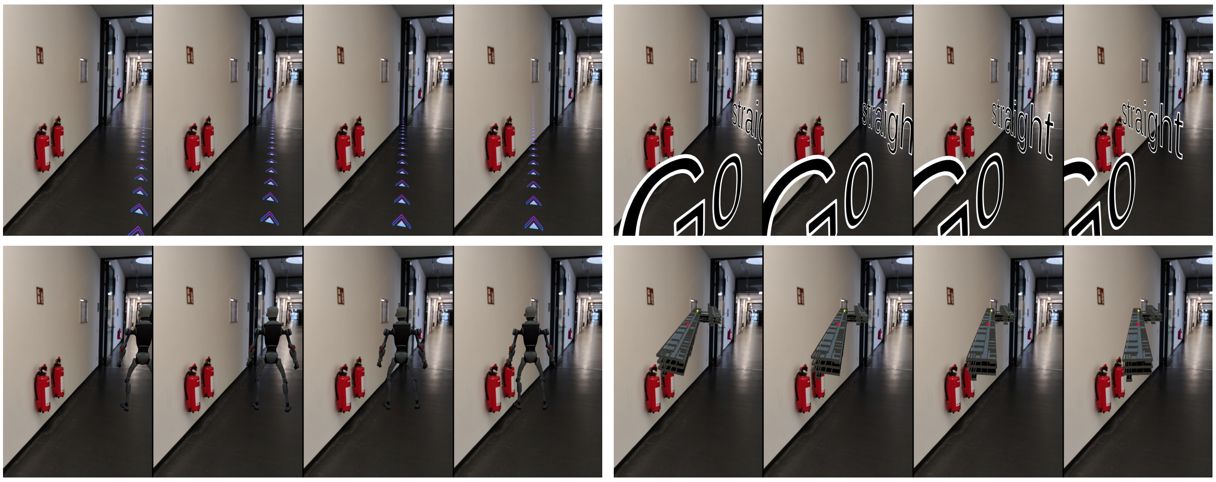
Guiding Information display techniques provide a series of guiding step-by-step information that are always limited to the next waypoint. This reduces cognitive load of the user and thereby positively affects their performance [WLPO94]. One contributing factor is the egocentric point of view that these display techniques provide [SCP95].

An example of such a technique is a *Guiding Arrow* that is positioned at the user and points towards the next waypoint [LMM16], or a *Shining Light* cone [MEN15]. This has the advantage of being perceived as less accurate than an arrow, which can dictate the users' expectations beforehand. The last approach in this category is the *Digital Avatar* that guides the user towards their goal by walking ahead of them [PDCK13].

A somewhat different approach to Guiding Information display techniques is the design of interfaces for visually impaired people [ZYZH19]. In this approach information is mapped to auditory or haptic cues (*Haptic Feedback*) to guide the user towards the next waypoint using different rates of tone frequencies or vibrations [ZYZH19].

Context Information

Context Information display techniques provide the user with additional information about their surroundings. This includes the traditional paper map. Users are shown the context in which they are positioned instead of direct navigation hints. By providing landmarks or other identifiable information about their surroundings, users are able to orient themselves using these techniques even without displaying their approximate location [BP13]. The goal of providing contextual information is to understand the connections between places within an environment and eventually help the user form their own route memory. This comes at the cost of an increased cognitive load while using the navigation system [RC17].



0° 5° 10° 15° 0° 5° 10° 15°

Figure 2: Simulated rotational error, increased in 5° steps from left to right for each display technique. Top left: Lines on the ground, top right: Bending Words (section 6), bottom left: Digital Avatar, bottom right: World in Miniature. As a non-visual technique, Haptic Feedback is omitted here.

A classic example is *World in Miniature (WIM)*. It essentially is a bird-eye view onto a small model of the building. Self-localizing is made easier by aligning the orientation of the miniature with the orientation of the real building and showing a marker for the position of the user within that model [HFH04, SCP95].

4 OBJECTIVE EVALUATION CRITERIA

In the following we describe how to evaluate navigation techniques based on objective criteria. We gathered these from multiple previous works to provide a comprehensive approach to rank display techniques within the context of creating error-robust indoor augmented reality navigation. They are not intended to replace systematic user studies but may be useful for preliminary investigations beforehand or to complement small-sized user studies in times where elaborate user studies are difficult to conduct, e.g. during a pandemic. The criteria are not metrics but instead provide guidelines to compare two or more display techniques against each other argumentatively. A score-based comparison was considered, as it can show variations between methods better, but due to the lack of calculable measures in some criteria it was decided against. Assigning each criterion a weight can be done and should be based on the individual requirements and target group.

We partition the aggregated list of criteria into two main categories: visibility (how to present information, subsection 4.1) and interaction (how to promote interaction, subsection 4.2). Each criterion within these categories was chosen based on the relevance to not only AR interfaces in general, but also to diminish the problem of uncertainty in tracking. Both technical aspects and human factors, including familiarity [AZLK12], are covered.

4.1 Visibility

Visibility describes elements that influence the ability of a user to perceive the navigation information and process it. This includes the visibility of elements, how often and how much of the information is located within the screen, and how effective it is in guiding the user.

Deviation Range: The deviation range refers to the range in which a display technique is still perceived as conveying the right information even though tracking errors exist [MKD+14]. The more precise the alignment between the desired position within the world and the position as shown on the display, the lower the chance of user failures such as taking a wrong turn or following a wrong path [MSS11]. It was shown that target detection performance decreases notably from precise ($< 7.5°$ rotational error) to partially degraded ($< 22.5°$) to poor ($< 45°$) given the increase in error [YWMB01]. Möller et al. also showed that users "perceived [wrongly estimated orientation] more negatively than a wrongly estimated location" [MKH+12]. Thus, tolerance against angular tracking error is an important evaluation factor, which we call the Deviation Range criteria. The display techniques can be evaluated and compared using the above mentioned thresholds in a qualitative manner or by comparing the overlap between the displayed content and the next goal on the display.

Path Information Visibility: To help prevent navigation errors, instructions have to be comprehensively distributed across the path and must be visible at appropriate points in time [RC17, SK00]. An overview of the upcoming tasks and especially the ability to easily see the next target location can also increase the performance of users [MSS11]. To summarize these effects, we devised this criterion where one needs to compare the amount of displayed path information for each technique.

4.2 Interaction

The way a display technique promotes interaction with the user, influences the user experience, but also impacts the tracking system itself. In the following we will explain the different criteria related to user interaction and their relevancy.

Device Orientation: The device orientation is mainly relevant for non-head-mounted vision-based tracking devices such as smartphones. The natural way for a user to hold these devices is at a 45° angle towards the ground [MKD+14]. This limits the number of feature points visible to the camera which negatively impacts the performance of a vision-based tracking system. A display technique should therefore enforce an upright position of the tracking device in a subtle and non-disturbing manner, which increases robustness.

Instant Feedback: Reaction time to changes in the tracking information of a display technique is an important factor for overall performance and user experience [RC17, SG04]. A technique, which updates the displayed information e.g. only at the next waypoint or at too large of a translational or rotational error can lead to deteriorated performance as the user might overshoot the target. Displaying lengthy animations are problematic for the same reasons.

Environmental Awareness: The awareness of the surrounding environment is an important factor when using digital navigation aids in general [BP13,MOP+09]. It decreases the requirement to use a navigation aid over time as the user becomes familiar with their surrounding and it also improves safety to avoid dangerous situations and obstacles. During times where the positioning system accumulates too much error to display reliable information, users can continue their navigation in the right direction using their acquired spatial knowledge [KAZ04]. A display technique can increase environmental awareness by using landmarks as part of the localization method or by making the environment stand out more. Techniques that cover too much of the screen or require the user to constantly look at it decrease environmental awareness. Environmental awareness using different display techniques is not measurable but needs to be compared argumentatively.

Multimodality Count: Multimodality Count is a measure counting the number of natural sensory receptors being utilized to convey information [GLB05]. In other words, how many senses does the display technique address? The main modalities relevant to current AR applications and the most researched within that context are: visual, auditory, and haptic modalities [Liv05, KSS20]. It is important to choose the right combination and number of modalities for the task at hand: Using more than one creates a more natural interaction with the system that grants more flexibility in a mobile situation such as during navigation [Gri09]. It can also increase the application's accessibility by allowing a user to freely choose their preferred modality [KSS20]. And it can reduce navigation errors of users, especially in reduced visibility conditions or if one modality conveys ambiguous information, therefore increasing effectiveness of the user-computer system [CFBM13].

Familiarity: A familiar display technique can help users build trust in the navigation information [AZLK12]. Existing navigation applications have introduced a set of display techniques that are widely accepted and understood, such as arrows or lines. Using these known forms can help users understand the system's intention when being guided. New, unfamiliar, technologies can sometimes lead to inadequate user experience for users with no previous knowledge of it and increase the cognitive load [ASB18]. This criterion is not directly measurable and depends on previous user experiences, though we expect differences to be mostly cultural. Therefore, a ranking using the familiarity criterion must be based on argumentative comparisons.

5 CRITERIA APPLICATION

To apply the presented evaluation criteria, we chose at least one representative from each navigation visualization category. From the Discrete Information category we chose lines on the ground as they generally provide more information than waypoint markers. From the Guiding Information category we chose the Haptic Feedback [ZYZH19], as the only non-visual navigation technique; and the Digital Avatar [PDCK13], as Guiding Arrow and Shining Light are conceptually only specialized instantiations of the Digital Avatar. As Paper Map is a non-digital navigation technique we opted for World-in-Miniature [SCP95] from the Context Information category.

In the following, we briefly describe the implementations of the techniques. Lines on the ground are implemented as stripes of arrows, that lead from the user along the path (Figure 2, top left). The walking direction is indicated by the arrow directions. The entire path is visible as occlusions from the environment are not taken into account. World in Miniature (WIM) shows the surrounding area as a small model within the view (Figure 2, bottom right). The model, including the user's tracked position depicted as a red dot, is fixed in front of the user, while its orientation is constantly aligned with the tracked orientation of the real building for an improved user performance [WLPO94]. Besides model and user position, the only additional information is the very next step of the navigation path, displayed as a yellow dot. The Digital Avatar (Figure 2, bottom left), is a humanoid robot entity that guides the user towards the next waypoint along the path. It walks in front of them and waits when the user is not moving. Haptic Feedback links angular difference between current viewing direction and next waypoint to vibration frequency of the tracking device. If the tracking device is pointed towards the correct direction (the next waypoint is inside the green region), it vibrates at the highest rate. If the next waypoint is in the outer thirds of the screen it vibrates with a medium frequency. If it is to the left or right of the screen the slowest vibration is applied to make the user aware that the system is still running.

5.1 Results

We obtain the **deviation range** ranking by measuring the range of rotational error that a display technique can be exposed to until it no longer overlaps with the position of the next waypoint. We verified the results qualitatively by simulating the rotational error in steps of 2.5° (Figure 2 showing every 2nd step).

Lines on the ground shows perceptible mismatch already at a small rotational error, the Digital Avatar and Haptic Feedback display accurate navigation information within a 5° rotational error. The value for Haptic Feedback has been obtained by measuring the range in which the vibration rates change. Besides the quantitative value of the deviation range we also include unique characteristics of the display techniques. E.g., even though the deviation range of WIM is only 5° it can still be used to identify the next steps by comparing the landmarks within the model with landmarks of the surroundings.

Concerning the **path information visibility** Haptic Feedback provides the fewest information as it only roughly points towards the next waypoint. Digital Avatar gives hints on where the waypoint is as a user approaches it. Because of the lack of occlusion, Lines on the ground can show more than just the path to the next waypoint, although information further down the path becomes more and more disassociated to the environment. WIM performs best, as it displays the complete environment as a map and can potentially show the complete path.

Most techniques that rely on displaying spatially registered information, such as lines on the ground and Digital Avatar, enforce an upright **device orientation** of 90° which is beneficial for feature tracking. With Lines on the ground the user has a slight tendency to look down and follow the arrows. Haptic Feedback is designed to aid visually impaired people who don't use visual cues and therefore tend to hold the device at a more natural angle of 45°. The camera then tracks mostly the floor which has fewer features. Linkage of the displayed model in WIM with the orientation of the tracking device [WLPO94] makes a rotation of smaller degrees more likely to improve visibility of path information, which has an inverse effect on the device orientation criterion. WIM also never gives incentives to point the camera in the direction of travel, making it rank worst.

Instant feedback All techniques except Digital Avatar update the navigation information or a deviation from the planned path in real-time. The Digital Avatar is restricted to human speed to not break the immersion. Note that this is strongly implementation dependant and should not be generalized.

Environmental awareness describes the trade-off between providing information from the navigation technique and environmental information to avoid obstacles and to become familiar with the environment over time. The WIM model shows the best support by providing context information in form of a model of the surroundings. Lines on the ground provides some but lesser context information in the form of directional changes at junctions and corners. Haptic Feedback was ranked third due to its minimum amount of information. We consider the digital avatar to be the least awareness-friendly technique, as it not only provides very little path information and occludes a large area of the screen, but also inherently enforces the user to focus on the avatar instead of the actual path.

The **multimodality count** refers to the number of natural senses addressed. All techniques solely rely on visual information except haptic feedback, which vibrates the device and even generates acoustic feedback thereby. Again, this is strongly implementation dependent.

The **familiarity** of each technique is based on how often elements of it are found within other everyday navigation situations. Digital Avatar implements the common situation where a user has to follow someone through an unknown environment. Lines on the ground in a similar form is broadly used for car and pedestrian navigation systems. Transferring this to AR preserves this familiarity. WIM resembles the well-known paper maps. Haptic Feedback is the least common approach as it makes the assumption that users can correlate an increase in vibration speed to positive directional feedback.

6 BENDING WORDS

Bending Words is our proposed discrete information display technique based on optimizing the criteria from section 4. In its core, this technique takes advantage of a person's ability to naturally follow turn-by-turn instructions. [Kim10] Yet, it overcomes the problem that turn-by-turn instructions are usually unable to show the exact location of the next turn [PB10]. During navigation Bending Words shows a three-dimensional text containing turn-by-turn instructions and adapts to the path in front of the user (Figure 2, top right, and Figure 3) both in terms of position and displayed text. The displayed text consists of two parts: The action keyword which is either *Go* or *Destination reached* to indicate if the goal is reached; and the direction keyword which is one of *right/left/straight/up/down* to indicate the next action at the next waypoint.

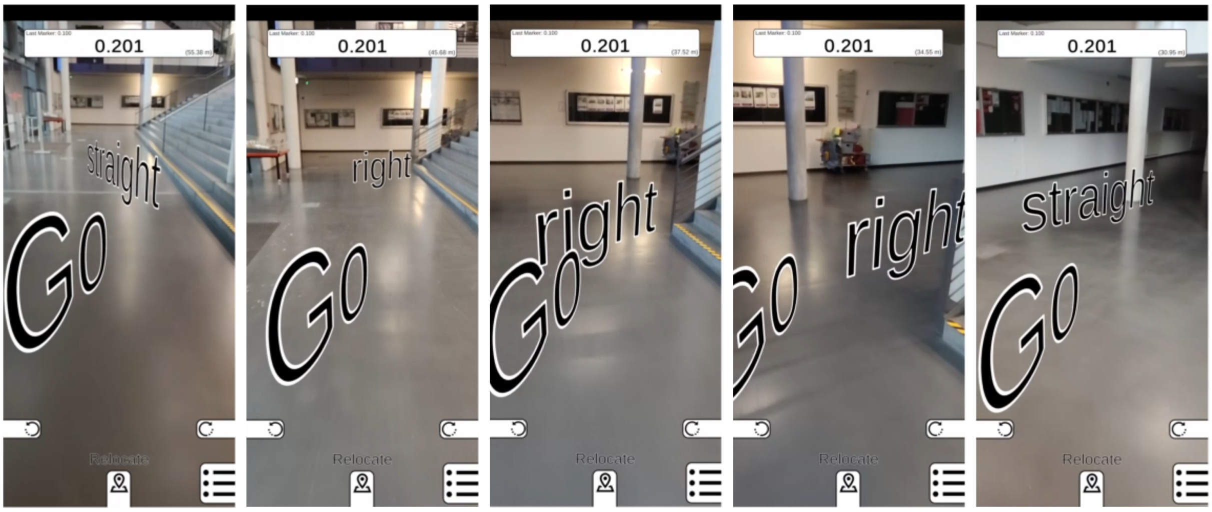


Figure 3: Bending Words: From left to right: The text aligns roughly with the path and the foreshortening as well as the content guide the user. If a new waypoint, e.g. corner, is approached the direction keyword changes and snaps to the waypoint's position. Once the waypoint is passed the technique returns to it's initial setup but always using the foreshortening effect to gently guide the user towards the next waypoint.

The displayed text is large enough to be easily readable but small enough to hide as little as possible of the environment. To improve readability and let the text stand from the envrionment, we opt for a black text with white contour.

The text aligns with the navigation path in front of the user. The action keyword *Go* is positioned at a small distance from the user and rotated around the y-axis to face the user at an angle. Due to the resulting foreshortening effect this indicates a walking direction for them. The absence of precise directions makes the user aware that this is only an approximate guidance. The direction keyword is placed at a larger distance and aligns with the path similarly as the action keyword. Its rotation is adjusted to keep a viewing angle that ensures readability. The angle adapts so that the foreshortening guides towards the next corner. When approaching a waypoint where the directional change would indicate an upcoming corner, the direction keyword snaps to the waypoint to emphasize this change before continuing along the path. An example is shown in Figure 3.

The advantage of this technique is that it shows users where to go and it tells them in text form, too. This supposedly increases the user's ability to understand the system's intention even if it suffers from low tracking accuracy. This way the user sees where they are headed to, based on the position of the words while the words themselves provide clear instructions in case of unclear situations.

7 DISCUSSION

In the following we discuss the proposed Bending Words display technique. Comparing it to the previous four display techniques, we obtain the ranking summarized in Table 3. A lower number means a better ranking in the respective category. The ranking of the baseline techniques has been discussed in section 5, here we only address the performance of Bending words in comparison to these techniques.

	Deviation Range	Path Information Visibility	Device Orientation	Instant Feedback	Environmental Awareness	Multimodality Count	Familiarity	**Average Rank**
Lines on the ground	5	2	3	1	3	2	1	**2.43**
WIM	2	1	5	1	1	2	3	**2.14**
Digital Avatar	4	4	1	5	5	2	1	**3.14**
Haptic Feedback	3	5	3	1	4	1	5	**3.14**
Bending Words	1	3	1	1	2	2	3	**1.86**

Table 3: Ranking of the different display techniques (left) within the selected criteria (top), where lower numbers are better.

The weighting of the criteria should be adapted based on the goals, given environment or target group. For example when developing a solution for people with mild cognitive impairments, the Familiarity criterion could be weighed more than Path of Information Visibility. The exact weights would need to be determined using a user study. For this work, we weight all criteria equally. In many criteria, the proposed Bending Words outranks the other display techniques (Table 3), followed by WIM.

Deviation Range: As the words in Bending Words only hint towards the next waypoint the text supports a clear decision making, resulting in strong robustness even with a rotational error of up to 10° ranking it the highest among the compared techniques. While the user can also correct any error when using WIM by comparing the map with their surroundings, this comes at a slightly stronger cognitive load though.

Strong translational and rotational errors can also have more drastic negative effects. The less dramatic effect would be that the system guides the user against a wall, in a more dramatic case, which we encountered during our own tests, Digital Avatar and Haptic Feedback would guide the user right down the stairs, even though the correct path continued to the right directly after the stairs. WIM can make it hard for the user to decide for the correct path, as the goal is displayed in the map but not necessarily the path. Even though the placement of Bending Words may also guide the user towards the stairs, the textual information always gives the user the possibility to correct the presented information. E.g. if Bending Words states “Go right” it is obvious that one should not go down the stairs as it would state “Go down” or even “Go down the stairs” instead.

Path Information Visibility: While Haptic Feedback and Digital Avatar only roughly indicate the current direction, Bending Words and Lines on the ground provide some contextual information (position of next waypoint, direction beyond that), and WIM even reveals the whole surrounding. We deem Bending Words to be slightly worse than Lines on the ground with this regard but better than Digital Avatar as it provides more information about the next waypoint earlier on.

Device Orientation: Similar to the Digital Avatar, Bending Words enforces the desired upright orientation of the tracking device. All other techniques implicitly enforce a worse orientation towards the ground.

Instant Feedback: Bending Words constantly reflects updates from the tracking system and quickly corrects known errors. Therefore, all techniques, except for Digital Avatar, perform equally well.

Environmental Awareness: An active process of localization supports the learning process of spatial knowledge required to independently navigate the building [Kui16]. Bending Words provides a neat way to give just enough guidance to stay on track, supporting memorization of the path. Therefore, we rank it slightly better than Lines on the ground. While we could give more precise context information, e.g. "Go right at elevator", this might clutter up screen space. Such extensions to our basic approach should be evaluated in the future.

Multimodality Count: As all techniques, except for Haptic Feedback, Bending Words mostly focus on the visual sense. It would be very simple to include additional senses, though, e.g. through audio feedback.

Familiarity: One can argue that we are used to textual instructions e.g. from assembly manuals or street signs which resemble Bending Words. Though, we are probably more used to follow other persons, as with the digital avatar. However, few of us are used to interpret vibrations as an information channel.

Limitations: Within our study we did not investigate how display techniques could be combined to improve the shortcomings of each other. For example, the Digital Avatar could be combined with an indicator that tells the user where the avatar is currently waiting for them, or the non-visual Haptic Feedback technique could be combined with a visual technique for a similar effect.

Bending Words is also limited by its reliance on turn-by-turn instructions, which can have adverse effects on spatial learning [KAS]. Its utilization of 3D space imposes an additional constraint on the number of words that can be displayed, further limiting the amount of instructive information that can be conveyed.

8 CONCLUSION

This article has presented a set of seven objective evaluation criteria for error-robustness in AR indoor navigation. Based on these, we have introduced a new display technique called Bending Words that can reduce the impact of tracking errors within an AR navigation application. We have evaluated and compared it to four other baseline techniques. Bending Words ranks best within the criteria, closely followed by WIM. Bending Words expands the spatially registered information provided by an AR display with precise instructions that can be easily interpreted.

In the future, new display techniques could be constructed for each visualization category, using the criteria presented in this work. It would be interesting to see how these criteria and other human factors interact with each other.

9 ACKNOWLEDGMENTS

Partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491805996 / GR 5932/1-1.

10 REFERENCES

[APM+16] Askarzadeh, F., Pahlavan, K., Makarov, S., Ye, Y., and Khan, U. Analyzing the effect of human body and metallic objects for indoor geolocation. In *2016 10th Int. Symposium on Medical Information and Communication Technology (ISMICT)*, pages 1–5, 2016.

[ASB18] Arifin, Y., Sastria, T. G., and Barlian, E. User experience metric for augmented reality application: A review. *Procedia Computer Science*, 135:648–656, 2018.

[ASWK15] Adler, S., Schmitt, S., Wolter, K., and Kyas, M. A survey of experimental evaluation in indoor localization research. In *2015 Int. Conf. on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–10, 2015.

[AZLK12] Arning, K., Ziefle, M., Li, M., and Kobbelt, L. Insights into user experiences and acceptance of mobile indoor navigation devices. In *Proc. of the 11th Int. Conf. on Mobile and Ubiquitous Multimedia*, pages 1–10. ACM, 2012.

[BCL15] Billinghurst, M., Clark, A., and Lee, G. A survey of augmented reality. *Found. Trends Hum.-Comput. Interact.*, 8(2-3):73–272, 2015.

[BEP15] Bettadapura, V., Essa, I., and Pantofaru, C. Egocentric Field-of-View Localization Using First-Person Point-of-View Devices. In *2015 IEEE Winter Conf. on Applications of Computer Vision*, pages 626–633, 2015.

[BP13] Brown, M. and Pinchin, J. Exploring Human Factors in Indoor Navigation. In *The European Navigation Conf.*, page 7, 2013.

[CFBM13] Calvo, A. A., Finomore, V. S., Burnett, G. M., and McNitt, T. C. Evaluation of a mobile application for multimodal land navigation. *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 57(1):1997–2001, 2013. Publisher: SAGE Publications Inc.

[DRMS07] Davison, A. J., Reid, I. D., Molton, N. D., and Stasse, O. MonoSLAM: Real-Time Single Camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007.

[FPS+20] Feigl, T., Porada, A., Steiner, S., Löffler, C., Mutschler, C., and Philippsen, M. Localization Limitations of ARCore, ARKit, and Hololens in Dynamic Large-scale Industry Environments. In *VISIGRAPP*, 2020.

[GFW21] Gomes, A., Fernandes, K., and Wang, D. Surface prediction for spatial augmented reality applications. *Virtual Reality*, 25(3):761–771, 2021.

[GLB05] Grasset, R., Looser, J., and Billinghurst, M. A step towards a multimodal AR interface: a new handheld device for 3D interaction. In *Fourth IEEE and ACM Int. Symposium on Mixed and Augmented Reality (ISMAR'05)*, pages 206–207. IEEE, 2005.

[Gri09] Grifoni, P. *Multimodal Human Computer Interaction and Pervasive Services:*. IGI Global, 2009.

[HFH04] Hallaway, D., Feiner, S., and Höllerer, T. Bridging the Gaps: Hybrid Tracking for Adaptive Mobile Augmented Reality. *Applied Artificial Intelligence*, 18(6):477–500, 2004.

[KAS] Krukar, J., Anacta, V. J., and Schwering, A. The effect of orientation instructions on the recall and reuse of route and survey elements in wayfinding descriptions. 68:101407.

[KAZ04] Krüger, A., Aslan, I., and Zimmer, H. The effects of mobile pedestrian navigation systems on the concurrent acquisition of route and survey knowledge. In Brewster, S. and Dunlop, M., editors, *Mobile Human-Computer Interaction - MobileHCI 2004*, Lecture Notes in Computer Science, pages 446–450. Springer, 2004.

[KH15] Kang, W. and Han, Y. SmartPDR: Smartphone-Based Pedestrian Dead Reckoning for Indoor Localization. *IEEE Sensors Journal*, 15(5):2906–2916, 2015.

[Kim10] Kim, H. J. Turn-by-turn navigation system and next direction guidance method using the same. In *US Patent 7844394B2*, 2010.

[KJ08] Kim, J. and Jun, H. Vision-based location positioning using augmented reality for indoor navigation. *IEEE Transactions on Consumer Electronics*, 54(3):954–962, 2008.

[KSS20] Kuriakose, B., Shrestha, R., and Sandnes, F. E. Multimodal navigation systems for users with visual impairments—a review and analysis. *Multimodal Technologies and Interaction*, 4(4):73, 2020.

[Kui16] Kuipers, B. The "Map in the Head" Metaphor. *Environment and Behavior*, 2016. Publisher: SAGE Publications.

[LGD+15] Li, X., Ge, M., Dai, X., Ren, X., Fritsche, M., Wickert, J., and Schuh, H. Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo. *Journal of Geodesy*, 89(6):607–635, 2015.

[Liv05] Livingston, M. Evaluating human factors in augmented reality systems. *IEEE Computer Graphics and Applications*, 25(6):6–9, 2005.

[LLY+15] Lymberopoulos, D., Liu, J., Yang, X., Choudhury, R. R., Handziski, V., and Sen, S. A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned. In *Proc. of the 14th Int. Conf. on Information Processing in Sensor Networks*, pages 178–189. Association for Computing Machinery, 2015.

[LMM16] Liu, K., Motta, G., and Ma, T. XYZ Indoor Navigation through Augmented Reality: A Research in Progress. In *2016 IEEE Int. Conf. on Services Computing (SCC)*, pages 299–306, 2016.

[MBMH01] MacIntyre, B., Bolter, J., Moreno, E., and Hannigan, B. Augmented reality as a new media experience. In *IEEE and ACM Int. Symposium on Augmented Reality*, pages 197–206, 2001.

[MCJ02] MacIntyre, B., Coelho, E., and Julier, S. Estimating and adapting to registration errors in augmented reality systems. In *Proceedings IEEE Virtual Reality*, pages 73–80, 2002.

[MEN15] Moura, D. and El-Nasr, M. S. Design techniques for planning navigational systems in 3-d video games. *Computers in Entertainment*, 12(2):2:1–2:25, 2015.

[MKD+14] Möller, A., Kranz, M., Diewald, S., Roalter, L., Huitl, R., Stockinger, T., Koelle, M., and Lindemann, P. A. Experimental evaluation of user interfaces for visual indoor navigation. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 3607–3616. Association for Computing Machinery, 2014.

[MKH+12] Möller, A., Kranz, M., Huitl, R., Diewald, S., and Roalter, L. A mobile indoor navigation system interface adapted to vision-based localization. In *Proc. of the 11th Int. Conf. on Mobile and Ubiquitous Multimedia*, pages 1–10. Association for Computing Machinery, 2012.

[MMC00] MacIntyre, B. and Machado Coelho, E. Adapting to dynamic registration errors using level of error (LOE) filtering. In *IEEE and ACM Int. Symposium on Augmented Reality (ISAR 2000)*, pages 85–88, 2000.

[MOP+09] Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., Näsänen, J., and Juustila, A. Like bees around the hive: a comparative study of a mobile augmented reality map. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 1889–1898. Association for Computing Machinery, 2009.

[MRBT03] May, A. J., Ross, T., Bayer, S. H., and Tarkiainen, M. J. Pedestrian navigation aids: information requirements and design implications. *Personal and Ubiquitous Computing*, 7(6):331–338, 2003.

[MSS11] Mulloni, A., Seichter, H., and Schmalstieg, D. Handheld Augmented Reality Indoor Navigation with Activity-based Instructions. In *Proc. of the 13th Int. Conf. on Human Computer Interaction with Mobile Devices and Services*, pages 211–220. ACM, 2011.

[MSTSP+21] Mendoza-Silva, G. M., Torres-Sospedra, J., Potorti, F., Moreira, A., Knauth, S., Berkvens, R., and Huerta, J. Beyond euclidean distance for error measurement in pedestrian indoor location. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2021.

[PB10] Pielot, M. and Boll, S. "in fifty metres turn left": Why turn-by-turn instructions fail pedestrians. In *Proc. of Using Audio and Haptics for Delivering Spatial Information via Mobile Devices, Lisbon (Portugal)*, pages 1–3, 2010.

[PDCK13] Pankratz, F., Dippon, A., Coskun, T., and Klinker, G. User awareness of tracking uncertainties in AR navigation scenarios. In *2013 IEEE Int. Symposium on Mixed and Augmented Reality (ISMAR)*, pages 285–286, 2013.

[RC17] Rehman, U. and Cao, S. Augmented-Reality-Based Indoor Navigation: A Comparative Analysis of Handheld Devices Versus Google Glass. *IEEE Transactions on Human-Machine Systems*, 47(1):140–151, 2017.

[RLJ+15] Rida, M. E., Liu, F., Jadi, Y., Algawhari, A. A. A., and Askourih, A. Indoor Location Position Based on Bluetooth Signal Strength. In *2015 2nd Int. Conf. on Information Science and Control Engineering*, pages 769–773, 2015.

[SBS+15] Shu, Y., Bo, C., Shen, G., Zhao, C., Li, L., and Zhao, F. Magicol: Indoor Localization Using Pervasive Magnetic Field and Opportunistic WiFi Sensing. *IEEE Journal on Selected Areas in Communications*, 33(7):1443–1457, 2015.

[SCP95] Stoakley, R., Conway, M. J., and Pausch, R. Virtual reality on a WIM: interactive worlds in miniature. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 265–272. ACM Press/Addison-Wesley Publishing Co., 1995.

[SG04] Sutcliffe, A. and Gault, B. Heuristic evaluation of virtual reality applications. *Interacting with Computers*, 16(4):831–849, 2004.

[SK00] Sutcliffe, A. G. and Kaur, K. D. Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology*, 19(6):415–426, 2000.

[WLPO94] Wickens, C. D., Liang, C.-C., Prevett, T., and Olmos, O. Egocentric and Exocentric Displays for Terminal Area Navigation. *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 38(1), 1994.

[YNA+17] Yassin, A., Nasser, Y., Awad, M., Al-Dubai, A., Liu, R., Yuen, C., Raulefs, R., and Aboutanios, E. Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications. *IEEE Communications Surveys Tutorials*, 19(2):1327–1346, 2017.

[YS15] Yang, C. and Shao, H.-r. WiFi-based indoor positioning. *IEEE Communications Magazine*, 53(3):150–157, 2015.

[YWMB01] Yeh, M., Wickens, C. D., Merlo, M. J. L., and Brandenburg, D. L. Head-Up vs. Head-Down: Effects of Precision on Cue Effectiveness and Display Signaling. *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 45(27):1886–1890, 2001.

[ZYZH19] Zhang, X., Yao, X., Zhu, Y., and Hu, F. An ARCore Based User Centric Assistive Navigation System for Visually Impaired People. *Applied Sciences*, 9(5):989, 2019.