Save the Smombies: App-Assisted Street Crossing

Kai Holländer
LMU MUnich
kai.hollaender@ifi.lmu.de

Andy Krüger
LMU Munich
krueger.andy@campus.lmu.de

Andreas Butz
LMU Munich
andreas.butz@ifi.lmu.de

Figure 1: Evaluated design concepts for smartphone-assisted street crossing. From left to right: Bars, Traffic Light and Map.

ABSTRACT
Using a smartphone while walking in urban traffic is dangerous. Pedestrians might become distracted and have to split their attention between traffic, walking and using the mobile device. The increasing level of automation in vehicles introduces novel challenges and opportunities for pedestrian-vehicle interaction, e.g., external displays attached to automated vehicles. However, these approaches are hardly scalable and fail to provide clear information in a multi-user environment. We investigate whether a smartphone app could provide individual guidance to enhance pedestrian safety in future traffic. To this end, we tested three app concepts in a user study (N=24) and found that on-screen guidance increases the frequency of successful crossing decisions significantly. In addition, all participants indicated that they would use the proposed system, preferably with an unobtrusive colored bar indicating the safest crossing decision in real-time. Integrating smartphones into the interaction between vehicles and pedestrians could increase situational awareness while crossing roads, solve the scalability problem and thus foster pedestrian safety in future traffic scenarios.

CCS CONCEPTS
• Human-centered computing → Interactive systems and tools.

KEYWORDS
Pedestrian Safety; Mobile Application; Smartphones; Automated Vehicles

ACM Reference Format:

1 INTRODUCTION
Pedestrians are not protected against collisions with motorized vehicles, thus they are among the most vulnerable road users in traffic [71]. In 2019, the National Roads and Motorists’ Association (NRMA) observed that 36% of pedestrians walked and crossed roads while actively using a smartphone or headphones [59]. The American Academy of Orthopaedic Surgeons (AAOS) found that 28% of more than 2000 respondents used their smartphone for reading tasks, texting, playing or taking selfies while walking [52].
Using smartphones while walking divides attention and increases reaction times [27, 53] as well as risky behavior [30, 42, 49, 66]. In prior investigations, pedestrians using their phone failed to notice a unicycling clown [36], someone in need of help [53] or money attached to a tree [35] due to inattentive blindness. Furthermore, Chen and Pai [10] observed that 22% of pedestrians playing Pokémon Go, an augmented reality game meant to be played while moving around outdoors [38], failed to observe traffic prior to stepping on the road. Recently, a collision of a motorcyclist and a boy who was focused on the game ended deadly [60]. Smartphone usage also causes people to walk slower, stagger or change direction more often [10, 36]. The nickname for people with these symptoms of divided attention is “Smartphone Zombies”, or “Smombies” for short, and the consequences of engaging in such behaviors can be injuries or even death [48].

In the future, we might also see automated vehicles displaying explicit messages or cues, for example via projections [8, 12, 51], displays [11, 22, 32] or light bands [5, 6, 26]. Such external human machine interfaces (eHMIs) present information about a vehicle’s intentions towards pedestrians in order to enable safe crossing decisions [57]. However, external vehicle interfaces have an inherent scalability problem. Future traffic interactions are very likely to include multiple vehicles at different stages of automation and multiple pedestrians. Thus, an eHMI could be seen by many pedestrians waiting on the road side while the signaled information addresses one person exclusively.

Such a mixed traffic scenario combined with different eHMI concepts is a risky ground for unwanted accidents, especially if inattentive pedestrians face increasingly complex traffic situations with various levels of vehicle automation. Therefore, we explore how to safely overcome the scalability problem and protect pedestrians by targeting them individually via their smartphone, the very source of their potential distraction. Our proposed solution aims to present mobile traffic guidance in a transparent manner directly on the screen of a personal device. We therefore named our prototype Smombie Defender.

In a lab study (N=24), we analyzed the crossing behavior of participants with video recordings. Participants continuously updated their willingness to cross while browsing images on a smartphone. Our results show that guidance by SmombieDe lead to a significantly higher success rate and more successful runs. A run is considered successful if the input given by the participant would result in a safe crossing decision in reality. Furthermore, SmombieDe reduces mental workload and all participants stated that they would use such an application. We conclude that integrating pedestrian guidance on smartphones could benefit overall traffic safety, reduce the negative effects of smartphone usage while walking and increase the acceptance of automated driving.

From a technical standpoint, Galileo (GNSS) [13] navigation and 5G cellular networks [54] will provide novel opportunities for a smart and connected infrastructure. We envision a future including automated vehicles, car-to-x communication and ubiquitous digital mobile devices. A result could be that city infrastructure changes fundamentally. For example, traffic signals might extend from their current fixed positions onto mobile devices. Our proposed concept could inspire future applications for such a scenario and integrate seamlessly into smart city concepts.

**Contribution Statement:** Given the two observations above (distracted pedestrians, ineffective eHMIs), we propose to address both problems at the same time by integrating a smartphone app into traffic communication. We developed three concepts for a smartphone app called SmombieDe and verified their validity in a user study. We also contribute five precise design recommendations concerning user interfaces for mobile app-assisted street crossing.

## 2 RELATED WORK

Our prototype builds on understanding pedestrians’ crossing decisions, as well as aspects of digitally supported street crossing approaches. The novelty of our work lies in the design of the user interface.

### 2.1 Pedestrian Crossing Decisions

Crossing decisions of pedestrians are mainly based on implicit vehicle signals [17, 47, 55, 61], such as motion [46, 56], braking [4] or the gap size between vehicles [67, 74]. Further influencing factors are the time of day and age [1, 43] as well as physical and mental states [19]. Unaccompanied children have a 177% higher risk of an accident in comparison to adults [73]. Drivers’ braking decisions are often based on pedestrians’ movement patterns [58]. This shows that mutual awareness between drivers and pedestrians is essential for safe crossings. We thus propose a warning system for pedestrians which could be implemented for drivers as well.

### 2.2 Digitally Supported Street Crossing

A major challenge for pedestrian guidance on mobile devices is reliable positioning data. Hwang and Jeong [34] introduce SANA, which calculates the probability of a safety-critical situation based on GPS data. SANA shows a collision warning for both, drivers and pedestrians on mobile devices. However, Jain et al. [40] state that GPS does not provide sufficient precision to support pedestrian safety. According to the authors, GPS includes large errors and delays, leading to many false-positive warnings. Their proposed solution is to combine smartphones, inertial sensor data, pedestrians’ movement patterns and GPS data to create a more accurate position estimation. A concept presented by OKI adds dedicated short range communication (DSRC) wireless modules to vehicles and mobile devices. If two entities equipped with DSRC sensors get close to each other, a warning is issued.

Other approaches do not rely on positioning data: Wang et al. [70] present WalkSafe, which uses the back facing camera to detect vehicles with a machine-learning-based image recognition algorithm. WalkSafe is supposed to protect pedestrians talking on the phone while participating in traffic and indicates warnings via vibrations and sound. An application called Visiblezone includes a proprietary communication protocol and works independently of a cellular network connection [75]. The application needs to be installed on the phone as well as in the car and targets drivers rather than pedestrians. Available concepts aiming at pedestrians specifically include see-through displays utilizing the smartphone camera [44] or expanding infrastructure with LED-panels [50]. Pavement lights illuminate the ground in front of pedestrian crossings, enabling people to recognize the currently displayed traffic light color without looking up from their phones. Schartmüller et al. [63]
We evaluated four user interface design concepts as guidance methods:

- **Bars**: Directional warnings on a mobile device and highlight that (over)trust is a crucial aspect of pedestrian guidance.
- **Traffic Light**: The Traffic Light concept is inspired by pedestrian traffic lights. If it is safe to cross, this concept shows a green sign on the lower part of a rectangular gray background (see Figure 1 middle). Otherwise, it shows a red light on the upper part of the rectangle. Users can adjust the size and position of the visualization at any time, in order to omit occluding a fixed area on the screen.
- **Map**: The Map guidance is based on Google Maps [45] and features a mini-map. It can be dragged around the screen by the user to avoid occupying a specific screen area. The map dynamically shows the user’s position (pedestrian icon) together with the location of nearby vehicles (colored circles, see Figure 1 right) in real time. Each circle moving towards the users’ location is displayed in red or green, depending on the right-of-way decision. If a vehicle is going to yield the circle is green, otherwise it is red.
- **Notify**: The Notify concept uses standard notifications which pop up from the top of the screen and include a text and an icon (a red or green traffic light symbol), accompanied by a corresponding ‘WALK’ or ‘STOP’ text. This was inspired by explicit vehicles’ signals investigated in the work of Fridman et al. [22] (see Figure 3).

### 3 RESEARCH QUESTION & HYPOTHESES

We aim to create a smartphone application which displays important guidance information while pedestrians may use any other application simultaneously. The main research question behind our work is whether a smartphone app can increase pedestrians’ awareness of oncoming traffic and therefore prevent dangerous crossing behavior. This question implies the challenge of an appropriate user interface design. Thus, we set up a prototypical application in a user-centered approach. The app combines visual stimuli and tactile feedback, and runs simultaneously with any other function on a smartphone. We investigated the following hypotheses in regards to the SmomDe prototype:

- **H1**: Using SmomDe increases the success rate in crossing decisions over a baseline without the application.
- **H2**: Using SmomDe results in a lower mental workload than the baseline condition.

We evaluated four user interface design concepts as guidance methods: **Bars, Traffic Light, Map and Notify**. To this end, we analyzed corresponding user behavior, subjective ratings, acceptance and mental workload.

### 4 PROTOTYPE DESIGN

Our initial idea was to transfer explicit vehicle signals and traffic lights to mobile devices, since pedestrians might benefit from being actively addressed. [9]. On-screen guidance should furthermore be scalable and work in urban traffic scenarios, given that a majority of pedestrian-related accidents happen in urban areas [73] and pedestrians tend to be less patient in bigger cities [14]. We created several design sketches and evaluated them in a focus group with four human-computer interaction experts (all male, 25 to 32 years, mean: 29.25 years, SD: 2.98). The main outcome was that the application should feature: (1) recognizable warnings even if the mobile phone remains in the pocket or the display is turned off, (2) tactile feedback and sound, (3) an on-screen overlay independent of running applications and (4) colors inspired by traffic lights (red and green). Furthermore, each pedestrian and vehicle would be assigned to a safety-area depending on the corresponding speed and movement direction (see Figure 2). If safety-areas overlap, involved entities receive a warning. The focus group’s discussions resulted in four visual design concepts, which are described below. All concepts use sound and tactile feedback with built-in native smartphone hardware.

- **Bars**: The Bars guidance method displays a green or red colored vertical border on the side of a smartphones’ screen (see Figure 1 left). If a vehicle is approaching from the right side, a colored bar is visible on the right side of the screen and vice versa. It is also possible for both screen borders to be active simultaneously. The bars are colored red, when the user has to stop. Otherwise they are green, indicating that no danger is nearby or all approaching entities will yield.

### Figure 2: Concept of safety areas for pedestrians and vehicles, depending on speed and direction of each entity. The hue of an area intensifies according to its potential danger.

![Figure 2: Concept of safety areas for pedestrians and vehicles, depending on speed and direction of each entity. The hue of an area intensifies according to its potential danger.](image)

- **Traffic Light**: The Traffic Light concept is inspired by pedestrian traffic lights. If it is safe to cross, this concept shows a green sign on the lower part of a rectangular gray background (see Figure 1 middle). Otherwise, it shows a red light on the upper part of the rectangle. Users can adjust the size and position of the visualization at any time, in order to omit occluding a fixed area on the screen.

- **Map**: The Map guidance is based on Google Maps [45] and features a mini-map. It can be dragged around the screen by the user to avoid occupying a specific screen area. The map dynamically shows the user’s position (pedestrian icon) together with the location of nearby vehicles (colored circles, see Figure 1 right) in real time. Each circle moving towards the users’ location is displayed in red or green, depending on the right-of-way decision. If a vehicle is going to yield the circle is green, otherwise it is red.

- **Notify**: The Notify concept uses standard notifications which pop up from the top of the screen and include a text and an icon (a red or green traffic light symbol), accompanied by a corresponding ‘WALK’ or ‘STOP’ text. This was inspired by explicit vehicles’ signals investigated in the work of Fridman et al. [22] (see Figure 3).

### 4.1 Pilot Study

We conducted a think-aloud pilot user-study (N=8, 2 female, 6 male, 24 to 42 years, mean age: 32.25 years, SD: 5.88) to collect early user feedback as the first iteration of a user-centered design process [20, 31]. The order of visual concepts was counterbalanced throughout the experiment. Participants used the image application [Imgur] for several minutes together with the SmomDe prototype while the experimenter changed the status (safe to cross / not safe to cross) five times for each of the guidance concepts. [Imgur] is a popular image-based application with more than 10,000,000 installations (as of January 2020) and can be used while walking. Each notification
method was tested with different modalities: multiple vibrations, vibrations and sound or visual only. The pilot study included a semi-structured interview and a questionnaire to evaluate users’ perceptions of the different methods. The questionnaire included a ranking of the guidance methods with a five point Likert scale [41] (see Figure 4). The most relevant findings from the pilot study are:

- The Bars concept was ranked first by four out of eight participants. Participants valued that it indicates the direction of potential dangers and its unobtrusiveness.
- The Traffic Light method was ranked first by two participants and described as the most intuitive and easy to understand concept.
- Six out of eight participants put the Notify concept in the last place. It was perceived as annoying and seven people reported that they are very likely to ignore it due to an overload of incoming notifications caused by other apps (e.g., messengers, games or calendars).
- All participants mentioned that sound is not needed as an additive cue, since tactile feedback would be sufficient even if the smartphone is not actively used.
- The implemented vibration pattern was perceived as “too aggressive” by two participants (P2 and P7).

As a result, we decided to not further investigate the Notify method in the main study and omitted auditory stimuli for the final prototype. The vibration pattern was shortened to a single 200 ms vibration at full force, generated only when a change from safe to unsafe situations occurred.

5 EVALUATION & USER STUDY

We measured the following dependent variables: mental workload to assess how much attention each guidance method demands via the raw NASA-TLX (short version) questionnaire [29]; the willingness to cross measured with a slider in real time (Crossbox); and a subjective rating ranging from 1 (very good) to 5 (not sufficient) where each of the ratings was unique. Therefore, it was not possible to rank two guidance methods with the same score. In addition, participants were encouraged to express their personal preferences, perceived acceptance and possible improvements regarding the different guidance methods in a concluding interview.

5.1 Crossbox

In order to measure participants’ willingness to cross in real time, we built the Crossbox, based on a research tool introduced by Walker et al. [69]. The Crossbox is a wooden box with an Arduino Mega 2560 [3] connected to a 10 kΩ sliding potentiometer (see Figure 5). A Bluetooth module (HC-05) connects it to the smartphone. We recorded the slider position via Processing [23] together with a timestamp for mapping measured values to recorded video frames. The slider values were converted to a corresponding percentage. For evaluating user behavior and the success of a run, we calculated A-slider positions for each trial by comparing participants’ input to a fixed ideal crossing behavior set by the experimenter. We decided to use the Crossbox over a binary yes/no button to get a more fine-grained estimation of participants’ situation assessment.

5.2 Videos

To ensure the safety of our participants, we could not allow them to cross a real road while using a smartphone. Instead, we recorded six videos on a local road with a speed limit of 30km/h. The videos were filmed using a GoPro Hero 7 Black [24] with activated linear filming mode, to counter the built-in wide-angle distortion. The camera was mounted on a tripod at a height of 156cm, which corresponds to the average human eye height [37]. The position of the camera mimicked the field of view of a pedestrian to create the effect of an ego-perspective.

In each video, all vehicles (driven in reality by colleagues) approach from the left side at a distance of 300m with a speed of 30km/h and either yield the way for pedestrians or continue driving. Yielding vehicles start to brake gently at a distance of 20m until reaching a full stop 5m away from the pedestrian, resulting in a
breaking distance of 15m. Therefore, the applied deceleration rate is in line with previously observed naturalistic driving behavior [15]. We recorded three scenarios including multiple traffic entities, each one with and without yielding vehicle(s). The videos have a run time of 39 to 46 seconds and contain: (1) a single car approaching, (2) two vehicles approaching, and (3) a pedestrian already waiting for an oncoming vehicle (see Figure 6).

5.3 Apparatus
To validate SmomDe, we equipped our lab with a white canvas (2.3m × 2m) and mounted a Canon SX Mark II (1400px × 1050px resolution) projector on top. Participants were shown the aforementioned videos and provided with a Huawei P20 Pro (Android Version 9, EMUI Version 9.1.0) including SmomDe. The computer connected to the Crossbox was a VR-ready, Windows 10 machine (Nvidia GTX 1980Ti GPU, IntelCore i7-6700k processor, and 16 GB RAM).

5.4 Study Design & Participants
The concepts were tested with 24 participants in a within-subjects, 4 (guidance methods + baseline) × 6 (videos) repeated measures experiment (N = 24, 9 female, 15 male, between ages 19 and 36, mean: 25.87 years, SD: 4.06). Trials were counterbalanced with a Latin square. This resulted in a data set containing 576 study cases. Each guidance method occurred 144 times, the approaching vehicle(s) drove by without stopping in 192 study cases whereas they yielded in 384 cases.

5.5 Study Procedure
Upon arrival at the lab, participants were greeted and introduced to the study. They filled out a consent form in accordance with local ERB regulations, were given an explanation of the Crossbox and went through two practice trials (yielding or driving vehicles) to become comfortable with the setting. Participants stood at a marked position with the Crossbox on a table next to them, about 1.5m away from the screen and held the smartphone looking at memes (mainly pictures, gifs or videos shared over the internet) on Imgur [62]. They were oriented towards the canvas showing videos of recorded traffic scenarios (see Figure 7) and indicated their willingness to cross in real time with the Crossbox. If the slider was at the front end, this meant that participants would absolutely cross (100% sure); the back end meant that participants were completely certain to stop. During the whole video, the vehicles’ intent was shown by a visual guidance method in SmomDe except for a baseline in which participants browsed images without additional guidance. Status changes from safe to unsafe conditions in SmomDe were accompanied by a 200ms vibration. After each study run, participants filled out a raw NASA-TLX questionnaire. After all runs, we conducted a semi-structured interview with the corresponding participant regarding the visual design concepts and possible improvements. Participants spent around 50 minutes in the lab and received a compensation of ten Euros in cash.

6 RESULTS
The Bars concept was rated highest (mean ranking: 1.71), followed by the Traffic Light visualization with a mean ranking of 2.67 and the Map concept with a mean ranking of 3.29. To investigate whether the means differ significantly, we calculated pairwise one-sample t-tests: Bars versus Traffic Light (t = −2.65, p = 0.01); Bars vs. Map (t = −4.65, p < 0.001); Traffic Light vs. Map (t = −1.54, p = 0.13), see Figure 8. Hence, Bars achieved a significantly (p < α 5%) higher user rating than the other concepts. For this work, indicated levels of
significance in figures are marked with asterisks: (*) $0.05 > p > 0.01$; (**) $0.01 > p > 0.001$; (***) $0.001 > p > 0.0001$; (****) $p < 0.0001$.

In the semi-structured interview, all 24 participants stated that they would accept the general concept of SmomDe. In the context given in the study, 45.8% answered they could imagine to fully rely on the guidance cues of such an application. Conversely, 29.2% answered they would accept the system, but only as additional feedback and still mainly rely on observations. The remaining 25% also mentioned to accept SmomDe in their daily life, if the following conditions were met:

- A sufficient familiarisation phase.
- An approval by traffic safety authorities or positive results in long-term testing.
- If there is a lack of eye contact while crossing (e.g., due to driverless vehicles).
- If guidance is only visible when “I am actually willing to cross the street” (P23).

Additionally, participants suggested the app should:

- Include the app status, e.g., reception strength of vehicle signals similar to the indication of smartphones’ cellular network connection quality.
- Make the Map visualisation resizable.
- Add directional arrows pointing towards oncoming vehicles to the Traffic Light concept.
- Combine the concepts Bars and Traffic Light.
- Show an additional timer which displays a countdown for red signals.

Participant 23 mentioned a possible perceived loss of control, since such an application could be interpreted as patronizing. The participant did not like the possibility of an application imposing certain actions. However, he still stated to accept SmomDe and did not have a suggestion on how to overcome this potential issue.

### 6.1 Mental Workload

The baseline condition results in the highest mental workload with a mean of 5.0 (max.: 10.0). To answer our second hypothesis, we paired the observations of each concept with the baseline and calculated corresponding paired t-tests, see Figure 9. Each SmomDe concept reduces the subjective workload significantly compared to the baseline: Bars ($t = -6.49, p < 0.0001$); Traffic Light ($t = -5.48, p < 0.0001$); Map ($t = -4.54, p < .001$).

### 6.2 Success Rates

To evaluate slider positions measured with the Crossbox, we calculated optimal runs for each yielding and driving scenario: If the vehicle yields, the slider position should stay at 100% or at least within the top 85% for an optimal run, since there is no danger present and pedestrians are safe to walk at any time. If the vehicle does not yield, the Crossbox knob should start moving down at around 28000 ms to a value between 0 and 15%. Otherwise, the participant would provoke a risky situation or collision. We chose a 15% threshold of the slider knob because (other than a button) there is no clear boundary of walking or standing in the slider data. The top and bottom 15% are an area which allows back and forth movements, hesitant behavior or cautious approaches of participants. Thus, participants are able to behave as in the real world where pedestrians might take a few timid steps before taking a final decision. To evaluate participants’ inputs, we compared measured slider positions with the described optimal runs. Our results show that the Bars concept lead to the highest success rate (85.4%; 123 of 144 trials), whereas Traffic Light (78.5%; 113 of 144), Map (71.5%; 103 of 144), and the baseline (8.3%; 12 of 144) achieved less successful runs (see Figure 10). To further analyze how the visual concepts influence success rates, we performed a logistic regression. The logit model includes a normally distributed random effect for the visual guidance methods. Bars (estimate = 4.33) shows the highest stochastic chance for success, followed by Traffic Light (estimate = 3.84) and Map (estimate = 3.46). Each of the guidance methods performed significantly better than the baseline (see Table 1). A pairwise analysis between the guidance concepts did not reveal any statistical significance. The six different videos, participants’
Table 1: Variance components model of success rates calculated by logistic regression analysis with dependent variables. Confidence intervals (2.5% and 97.5%) of success rates in relation to the baseline: intercept (-1.38% and 1.97%), Bars (3.55% and 5.10%), Traffic Light (3.11% and 4.58%), Map (2.74% and 4.17%).

|                     | Estimate | Std. Error | z value | Pr(>|z|) |
|---------------------|----------|------------|---------|----------|
| (Intercept)         | 0.3      | 0.85       | 0.35    | 0.73     |
| Bars                | 4.33     | 0.4        | 10.94   | < 0.001  |
| Traffic Light       | 3.84     | 0.38       | 10.23   | < 0.001  |
| Map                 | 3.46     | 0.37       | 9.47    | < 0.001  |

Figure 11: Mean Δ-values of unsuccessful runs per concept. Stars show significance of two-sided Mann-Whitney-U tests.

To further assess the design concepts, we calculated Δ-slider positions, which are a metric for unsuccessful runs. The Δ-value expresses the deviation of the slider position to the optimum for each millisecond. Thus, this data set consists of timestamps with non-optimal slider positions accumulated of all methods, participants and videos. Since the conditions for a parametric t-test are not fulfilled (no normal distribution), we investigated central tendencies of the samples with Mann-Whitney U-tests. Two-sided Mann-Whitney U-tests revealed the correlations shown in Figure 11. The Bars method lead to significantly more successful runs (lowest Δ-values in %) compared to any other visual concept, followed by Traffic Light and Map. The baseline performed significantly worse than any SmomDe concept.

For a visual comparison of guidance methods, we created graphs including slider positions for all runs and each method. Each run is represented by a colored line. For ideal runs, participants’ inputs should stay at the top green area and omit the red section. In the most intense red area (33.000ms to 37.000ms) a collision is inevitable (see Section 4). The graphs in Figure 12 show that the Bars guidance method produces the lowest number of outliers and better clustered lines around an ideal run in comparison to the other visualizations. In addition, the baseline condition seems to provoke the highest frequency of irrational behavior, e.g., unnecessary waiting (see left part of Figure 12 (d)).

7 DISCUSSION

According to the presented results, we have to accept both stated hypotheses (see Section 3): The usage of SmomDe leads to a higher...
The aim of this work is to mitigate negative effects of mobile device usage in traffic. We want to specifically emphasize that our goal is to include visually impaired pedestrians. Based on the insights of this investigation we present five design recommendations for an implementation of mobile-device-assisted street crossing:

- **DR1:** Show unobtrusive and directed indications of oncoming traffic. User feedback suggests that mobile guidance should not cover a central area on the screen and clearly indicate the direction of the cause of an issued warning.

- **DR2:** Reduce presented information and include tactile feedback. Traffic guidance should be comprehensible within split seconds. Therefore, we suggest reducing presented information as much as possible and including multiple modalities to relieve cognitive load on the visual channel [18, 25, 72].

- **DR3:** Present warnings only. In order to avoid false positives, patronizing pedestrians and to mitigate overtrust, mobile app-assisted street crossing should give warnings only and not present instructions to move. This recommendation is derived from participants interview feedback.

- **DR4:** Do not use notifications for warnings. Our pilot study showed that notifications were associated with entertainment, messages or calendar applications and warnings were likely to be ignored if deployed as notifications.

- **DR5:** Consider the target audience and enable inclusion. Pedestrians can be of all ages as well as physical and mental abilities [7, 21, 68]. Therefore, it is especially important to consider inclusion when creating guidance applications for pedestrians. An example would be designing accessibility for color-blind people.

### 7.4 Pedestrians and Automated Vehicles

Warnings on a Smartphone could resolve the scalability problem of eHMIs. Therefore, we see this problem as an inspiration for the concept of SmomDe. Automated vehicles (AVs) equipped with eHMIs might increase traffic complexity and demand even more attention from pedestrians than nowadays, which counteracts smartphone usage while walking. Explicit information of an AV might be a source of confusion instead of meaningful decision support. To overcome this problem, we believe that targeting individuals via mobile devices could provide a promising solution. A centralized service at the backend of a street crossing application could even enable a two-way interaction, in which pedestrians become directly involved in traffic flow decisions. For example, a group of pedestrians outnumbering the people inside a vehicle could force it to yield. Such a system could prioritize pedestrians over vehicles, shifting traffic flow optimizations towards walking rather than driving. Especially in urban environments and in the context of climate change, creating safe and comfortable means for pedestrians to navigate through dense cities might gain in importance. For a real world implementation we envision a system using native mobile device hardware (e.g. camera, inertial sensors, location services) similar to the approach of Jain et al. [40]. A corresponding technical implementation could use a low latency 5G mobile internet connection combined with the Galileo positioning system, which features a higher accuracy than GPS (less than 1m compared to 3m [39]).

### 7.5 Limitations

The results of this investigation should be understood within the limitations of our study design. We did not observe behavior in real-life situations, but presented videos on a large screen. Participants indicated their walking behavior with the Crossbox in real time relative to the running video, but did not actually walk. This setup was chosen to protect participants against potential collisions and is inspired by prior work from the context of automated vehicle and pedestrian interaction [8, 16, 64].

---

**7.1 Visual Guidance Concepts**

Participants stated preferring the Bars concept, due to its unobtrusiveness and indication of direction of a potential danger. Additionally, the Bars method performed best in the statistical evaluation. The Traffic Light concept was described as familiar and easy to understand but lacks an indication of the direction from which vehicles are approaching. Participants explained that the Map visualization showed too much information and was confusing. This confusion might occur since a spatial mapping of the real world and the on-screen dynamic map is necessary to comprehend the signals. Still, all guidance concepts performed significantly better than the baseline condition. Implementing well-known colors from traffic signals for pedestrian guidance fosters comprehensibility [22, 32]. However, red and green signals alone are not suitable for color-blind pedestrians. Thus, additional indications, e.g., via the position or shape of the signals or adjustable colors should be considered to include visually impaired pedestrians.

**7.2 Design Recommendations**

Based on the insights of this investigation we present five design recommendations for pedestrians and automated vehicles.

- **DR1:** Show unobtrusive and directed indications of oncoming traffic. User feedback suggests that mobile guidance should not cover a central area on the screen and clearly indicate the direction of the cause of an issued warning.
- **DR2:** Reduce presented information and include tactile feedback. Traffic guidance should be comprehensible within split seconds. Therefore, we suggest reducing presented information as much as possible and including multiple modalities to relieve cognitive load on the visual channel [18, 25, 72].
- **DR3:** Present warnings only. In order to avoid false positives, patronizing pedestrians and to mitigate overtrust, mobile app-assisted street crossing should give warnings only and not present instructions to move. This recommendation is derived from participants interview feedback.
- **DR4:** Do not use notifications for warnings. Our pilot study showed that notifications were associated with entertainment, messages or calendar applications and warnings were likely to be ignored if deployed as notifications.
- **DR5:** Consider the target audience and enable inclusion. Pedestrians can be of all ages as well as physical and mental abilities [7, 21, 68]. Therefore, it is especially important to consider inclusion when creating guidance applications for pedestrians. An example would be designing accessibility for color-blind people.
We did not investigate the mental effort of the image browsing task on Imgur [62], since the scope of this investigation focuses on the user interface design concepts and their effects. Thus, we do not know whether our results remain stable with another secondary task. We believe, that the task has a major influence on immersion, distraction and resulting inattentiveness and should therefore be considered in future work.

Tactile feedback via vibrations was perceived as a valuable addition to the visual guidance. Vibrations as a warning modality were not part of the baseline. This could be a reason why concepts featuring tactile feedback lead to significantly better outcomes. We did not investigate tactile feedback as an isolated warning channel. Auditory cues were rejected by the participants in our pilot study. This might be a side effect of the study setup. We tested a single beeping sound as a modality in a closed lab condition and not in a lively street. However, the Multiple Resources Theory [72] states that humans have a limited set of capabilities for cognitively processing information per modality. Since observing traffic and using a smartphone both include visual processing, we think that adding tactile feedback with sound could support pedestrians’ awareness and traffic safety. Redundancy in user interfaces can furthermore foster accuracy of use [25]. Auditory cues could provide a channel to redundantly indicate directions of potential dangers, e.g., via earcons or spearcons. In this context, Hakkinen and Sullivan [28] highlight that auditory warnings have to be carefully designed.

7.6 Future Work

Future work should consider how a universally understandable application could be designed, including multiple modalities, e.g., auditory displays for warnings [28, 65] and directed tactile feedback. The development of applications for assisted street crossing should include users of various age groups and with mental or physical limitations. Future developments of mobile guidance applications should also consider (over)trust in the design process [33, 63].

Additionally, concepts which target cyclists, (E-)scooter riders and motorcyclists could foster traffic safety. When developing corresponding applications, various devices could be investigated, e.g., smart watches, tablets or handheld computers. Approaches for traffic guidance could occupy a screen completely, forcing users to steer their visual attention to the road. Although we believe that this approach might fail to find acceptance for pedestrians, it could be worthwhile to investigate it further in the context of other vulnerable road users, such as cyclists or E-scooters.

We currently do not see how a real world study investigating mobile device usage in traffic could be in line with the Declaration of Helsinki [2]. Thus, we suggest implementing immersive virtual reality environments with eye tracking to further elaborate on the results of this first evaluation. In this context, the Crossbox provided valuable outcomes and could be used in future studies for investigating user behavior.

8 SUMMARY AND OUTLOOK

The presented mobile application SromDe follows a human-centered design, targets pedestrians specifically and aims to overcome the scalability problem of eHMs in the context of mixed traffic with automated vehicles. Pedestrians using their smartphone in traffic are in danger and current AV-to-pedestrian communication prototypes lack the functionality to address multiple pedestrians individually at the same time. SromDe aims to overcome those issues by presenting guidance on personal devices. We conclude that tactile feedback combined with a visual indication of Bars (which show the direction of oncoming traffic entities) provides a promising approach for the design of mobile application supported street crossing. Our results show that on-screen guidance increases the success rate of crossing decisions significantly and that users accept the concept. Additionally, we present five design recommendations which are suitable to improve the presented SromDe prototype.

We are convinced that the combination of traffic and mobile devices will gain importance. People might walk or (E-)bike more for ecological and economical reasons. In addition, novel mobile applications could tempt pedestrians to use them while participating in traffic. In combination with automated vehicles, digitally assisted street crossing could prevent accidents. However, the presented concept in this work could be implemented independently of automated vehicles and save the smombies if the required technological infrastructure is available.

REFERENCES


