# Unexpected Driving Behaviour of Autonomous Cars – Explainable AI for Pedestrians

Luca Rolshoven, *University of Fribourg*

***Abstract*—With the emergence of autonomous vehicles (AVs), a lot of questions about the explainability of the vehicle's actions arised. One of these questions is how such a vehicle should communicate its intentions to the pedestrians in a manner that they can decide whether to cross the street or not and feel safe in encounters with AVs. Recently, there has been quite a lot of research going on in this field and several interfaces have been proposed. However, a lot of these interfaces only use the visual modality. Those studies in which a combination of various modalities have been used often find that too much cues could lead to more ambiguity in the decisions of pedestrians rather than helping them in their reasoning. This paper tackles this problem by proposing a dynamic multimodal interface (DYNAMO), which uses vision as the main modality. Under certain conditions, the interface activates an auditive component, a haptic component or both components at the same time. This should decrease the possibility of an information overload and increase the communicational efficiency, particularly for pedestrians that suffer from visual impairment. At last, some evaluation methods are proposed and discussed.**

***Index Terms*—Autonomous Vehicles, Pedestrians, Intent Communication, Explainable AI, Interface, Multimodal**

## I. INTRODUCTION

WITH the emergence of autonomous driving, the question arises how a car should communicate its intentions to the driver, especially if the driving behaviour seems unexpected. While it is certainly important for the driver to know what the vehicle is planning to do next, it seems to be even more critical for pedestrians to understand the intentions of a self-driving vehicle, since a misunderstanding may endanger their health. In 2016, 5320 pedestrians died in road accidents in the EU, which corresponds to 21% of all road fatalities [1]. While the decisions of autonomous vehicles (AVs) may be based on more sensor data and therefore be safer, the decisions of pedestrians when crossing a street should also be considered. We do not want to increase the number of road fatalities involving pedestrians, hence a comprehensible intent communication interface is needed to replace the communication that was previously done by the driver itself.

Many studies in the field of behavioral psychology have been conducted about the social aspects of driving. This paper aims to give a better understanding of how we might shift these social aspects that serve as an indicator of the driver's intention towards other mechanisms that enable the communication between AVs and pedestrians. In Section II, I will briefly talk about the current state of the art and about ongoing research in the interaction between intelligent transportation systems and pedestrians. Then I will propose a novel interface in Section III that can be used to communicate the intentions of the vehicle using up to three different modalities. In Section IV, we will look at some suggestions of how to evaluate the efficiency and robustness of the proposed interface and in Section V I will discuss the connection between the parts of the interface and the research on pedestrian behavior and I will briefly show that the interface satisfies the functional requirements that will be introduced in Section III.

## II. PREVIOUS WORK

### A. General Pedestrian Behavior

There are a lot of studies that examine the behavior of pedestrians with respect to traditional non-autonomous vehicles. These studies uncover a wide range of factors that influence the conduct of pedestrians. One of those factors is the number of pedestrians that are present at the same time. In an older study of 1991, Harrell [2] shows that pedestrians pay less attention to the traffic when there is a larger group of pedestrians present at the other side of the street. Social norms seem to be another important factor, which is why Wilde [3] studied the effect of formal and informal rules and used the terms *psychological right of way* and *natural right of way* to denote either the social or the lawful right of way at a non-signalized intersection. Russell et al. [4] found that the social status of a jaywalking role model and the number of such role models increased the number of pedestrians that followed their example. These findings were consistent across male and female role models of different ethnicities. Women seem to be more cautious when crossing a road than men [5] because their perceived risk is higher [6]. Furthermore, Tom and Granié [7] found that women had a higher law compliance than men when it comes to temporal crossing rules, i.e. traffic lights. They also showed that men and women differ in their gaze patterns when crossing the road: while women mainly focus on other pedestrians, men tend to focus on vehicles instead. Another important factor is the age of the pedestrians. Older pedestrians tend to be more cautious [8] and they need a bigger gap before they feel safe enough to cross the road [9].

The speed of the pedestrians is important as well. Oudejans et al. [10] argue that pedestrians have a better estimation of whether crossing is possible or not when they are already walking. They attribute this, inter alia, to the optical flow information that is generated while walking, which might lead to a better distance estimation. However, the speed at which pedestrians walk is influenced by a lot of factors such as perceived or actual space [11], [12], [13], intention of crossing [14], group size [11], [12], [13], age [11], [12], [13], [15], gender [11], [12] and time of day [11].

Attention is without doubt another important factor for safely crossing a road. Geruschat et al. [15] divide the action of crossing a road into three stages: (1) walking to the curb, (2) waiting at the curb and watch out for a gap that allows safe crossing, (3) crossing the street. They found that their test subjects fixated mostly crossing elements such as bollards, curbs and crosswalk lines during the first stage. In the second stage, most of the attention was on the cars, which was to be expected. In the third and last stage, the subjects fixated the crossing elements again most of the time. Especially in stage two and three, attention is crucial for safely crossing the street. Hyman et al. [16] conducted an experiment in which 75% of the test subjects that were using a cell phone while walking through Red Square, a large central plaza of Western Washington University, did not notice that there was a unicycling clown present. The use of cell phones and other electronic devices may therefore lead to severe inattentional blindness and should be considered when building a communication channel between an AV and pedestrians.

It seems that the walking direction of the pedestrians is also an important factor to consider. Schmidt and Färber [17] found that pedestrians had a tendency to choose more dangerous gaps when a car approached them from behind and not from the front. The ability to estimate the speed and the distance of a vehicle both depend on it's actual speed. Sun et al. [18] found that pedestrians were underestimating the velocity of an approaching vehicle if it was driving with a speed of more than 40 km/h (sunny condition) or 45 km/h (rainy condition) and that they underestimated the stopping distance of the vehicle if it travelled faster than 65 km/h.

Of course, traffic signals also have an impact on the crossing behavior of pedestrians. An experiment by Tom and Granié [7] shows that pedestrians pay more attention to vehicles at unsignalized crossroads. Their results show that 86% of pedestrians look at vehicles if they are standing at an unsignalized crossroad in contrast to 69.5% of the pedestrians that were standing at a signalized crossroad. The weather also has an influence on how cautious pedestrians are when crossing a road: lower temperatures lead to a decrease in the level of cautiousness [2]. Moreover, pedestrians seem to be more cautious and conservative in their crossing decisions under bad weather conditions such as rain [18].

We have already seen that older pedestrians need a bigger gap to feel safe enough to cross the street. This *gap acceptance* is different for each individual. The *critical gap* [19], where pedestrians will cross the street, can be measured in terms of the Time To Collision (TTC) [20], which denotes the time it takes for the vehicle to reach the pedestrian. Usually, these critical gaps have a TTC between 3s and 7s for most pedestrians [21], [22], [17]. Interestingly, Schmidt and Färber [17] found that, at least under time pressure, pedestrians are more likely to rely on the estimated distance rather than the TTC. Since the TTC is a function of the distance between the vehicle and the pedestrian and the speed of the vehicle, this implies that within the same TTC values, pedestrians prefer to cross the street when the vehicle is farther away even though it approaches them at a higher velocity.

Risto et al. [23] show that pedestrians may infer the intentions of a car driver by observing the *movement gestures*, i.e. the way the vehicle moves. They found that pedestrians often appeared to be uncomfortable when certain vehicle behaviors were missing. A common movement gesture that drivers use to communicate that they won't let pedestrians pass is to keep the same speed or to even accelerate [24]. Rasouli et al. [22] observed that in more than 90% of the clips in their dataset, pedestrians showed some sort of non-verbal communication, with the change of head orientation being the most prominent one. More than 80% of these head movements occurred prior to crossing. An interesting study by Guéguen et al. [25] found that more car drivers stopped when confederates stared directly at the drivers than when they looked above the driver's heads. Although not all studies support the importance of eye contact in the negotiation of traffic situations between drivers and pedestrians [26], there seems to be more support that eye contact indeed is an important factor [27].

There are other factors that also play an important role in crossing situations such as culture and past experience. However, these attributes are difficult to measure directly and are therefore somewhat irrelevant for an AV, since it won't be able to base it's decisions and communication patterns upon these particular factors of variation. For a more detailed overview of those and other factors mentioned here, please refer to Rasouli et al. [27].

### *B. Interaction Between Pedestrians and Autonomous Vehicles*

While some authors believe that there is no need for a communication channel between pedestrians and autonomous vehicles, mainly because the AV has to stop either way if a pedestrians crosses the street [28], there are more arguments that do support the importance of communication [29], [30], [31]. Another important point is made by Prakken [32] who states that an AV should also be able to distinguish between pedestrians and police officers that give directions.

To investigate the interactions between pedestrians and AVs, the Wizard of Oz (WOZ) technique [33] is often applied to make the participants believe that they are interacting with a vehicle that is actually in an autonomous driving mode. Lagström and Lundgren [29] used this technique in a small study to observe the effect of the driver's apparent attention on the crossing decisions of pedestrians. All of the pedestrians crossed the street if there was direct eye contact between the driver and the pedestrians. However, less than 40% of the pedestrians decided to cross the street when there was no driver present or if the driver was reading a newspaper. These findings illustrate that the pedestrians do not entirely feel safe when an AV approaches them and that efforts should be made to improve the communication of the intentions of AV. Yang [34] reports similar findings as Langström and Lundgren and further points out, that pedestrians felt uncomfortable and worried when the AV had black windows.

It has been shown that *intent displays* can help to resolve potentially dangerous deadlock situations [35]. Böckle et al. [36] found in a virtual reality (VR) study that their intent communication interface, consisting of four LED columns and an audio speaker, increased the perceived level of safety and comfort of the pedestrians. However, in an experiment conducted by Clamann et al. [37], only 12% of the pedestrians reported that the intent display on the vehicle influenced their decision to cross. They found no significant effect of the display on participants decision times. In a VR study by Chang et al. [38], pedestrians were confronted with normal cars (type 1) and cars that had moving eyes installed at the front (type 2). Whenever a car of type 2 stopped, the eyes were looking at the pedestrian. The results show that two thirds of all participants made faster crossing decision when confronted with a car of type 2 rather than a normal car. After giving the pedestrians a hint that they should look at the eyes of the cars, but without explaining their functionality, the number rose to 86.6%. While the eyes in front of the car led to an increase in perceived safety in this study, Mahadevan et al. [39] found that their animated face which was used in one of their interfaces was not received very well. The authors found an overall positive feedback for the use of human-like cues in interactions between pedestrians and AVs, yet they expect this anthromorphism to become a thing of the past as pedestrians interact more with AVs and gain familiarity.

Zimmermann and Wettach [40] showed that a very short stopping distance or abprupt acceleration of an autonomous vehicle are perceived as erratic behavior and very aggressive, thus influencing the crossing decisions of pedestrians. However, the interpretation of certain movement patterns may be altered depending on various factors. Beggiato et al. [41] showed that pedestrian's perception and expectations of an automated braking action were dependant on the time of day, the vehicle speed and the age of the pedestrian.

### C. Interfaces for Vehicle-Pedestrian Communication

There are many different ways for an AV to communicate it's intention. Since it is not always clear whether a vehicle is being used in manual driving mode or if it is driving autonomously, it is a good idea to convey this information, which, for example, can be done using LED lighting patterns or different colors of LED lights [29], [42]. In his M.A. thesis presentation, Graziano [43] presents an interface called AutonoMI that communicates to the pedestrians that they have been detected. It does this by activating a light signal on a stripe. This light signal follows the movement of the pedestrians to let them know that they are still recognized. Other interfaces include information about the current speed of the vehicle [37] or they are able to project messages onto the ground such as a zebra crossing to show that the pedestrians can safely cross the street [44]. These interfaces are all mostly visual.

Besides Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) technologies, which aim to communicate information about the AV to other vehicles or the infrastructure, efforts have also been made to establish Vehicle-to-Pedestrian (V2P) technologies, that send this information directly to the pedestrians [45], [46]. In 2014, Ken McLeod, Policy Director at the League of American Bicyclists, conducted a survey with 357 participants which included questions about the acceptance of V2P technology [47]. The results show that 38% of the respondents, which consisted of bicyclists and pedestrians, would not make use of V2P technology. Some of the reasons were privacy concerns, perceiving V2P technologies as a burden and the belief that the responsibility might shift away from cars and their drivers. While these results indicate that V2P technology should maybe not serve as the main communication channel between pedestrians and AVs, it would arguably still be a good idea to include it as a secondary channel, since 62% of the respondents in the survey said that they would consider using such technology if it increases their level of safety.

## III. PROPOSED INTERFACE

### A. Modalities and Cues

In Section II-A we saw that pedestrian behavior is influenced by a lot of factors such as pedestrian group size, social norms, vehicle speed, age, gender, etc. While it might not be possible for an AV to consider all of these factors, it is important that it considers as much of them as possible when communication its intention to the pedestrians. Furthermore, the interface should also make use of various modalities since some of the pedestrians may exhibit limitations regarding one particular modality. In addition, if we only used one modality, the communication could be disrupted by environmental factors such as bad weather conditions. The Automated Vehicle Interaction Principle (AVIP) proposed by Habibovic et al. [48] states that our interface should satisfy the following functional requirements: 1) pedestrians should be able to tell whether the vehicle is in autonomous or manual driving mode 2) the vehicle must communicate it's future state and intentions 3) there must be a replacement for eye contact 4) the interface should not make suggestions about crossing decisions 5) the interaction with the interface should be calm.

Especially the second last point is important, since the AV may fail to consider other vehicles and road users and therefore suggest a pedestrian to cross the street when the situation is actually dangerous. But first, let's talk about which modalities that can be used in the interface. Vision might be the most obvious one that has been used a lot in the communication interfaces discussed previously. It may be useful to include audition as well, especially for those pedestrians that have impaired vision. Furthermore, if V2P technology is included as an alternative communication channel, we can also make use of taction, e.g. by broadcasting a vibrational pattern to smartphones in a certain range of the AV to warn or inform pedestrians about the intentions of the vehicle. These different modalities and the cues that could be generated for each of them should however be chosen carefully. Mahadevan et al. [39] argue that there exists a possibility of information overload [49] when using too many cues. One subject in their study stated that she or he had the feeling of having to wait for all of the cues before crossing, which takes a lot of time and could be confusing to many people. A possible solution for this could be to only use auditory and haptic cues when there is a need for them. For example, the AV could decide to use auditory cues when the weather is bad and there is no good sight. It could then play a sound whose frequency increases and decreases with the speed of the vehicle, in order to indicate whether it will stop or keep it's speed. In the case of pedestrians that have some sort of impairment, the additional cues could be activated upon registering pedestrians that have a special V2P application installed on their smartphones. This way, vision still is being used as the main communication modality, but the other modalities are activated dynamically whenever they are needed, which should prevent the occurrence of an information overload.

### *B. Primary Communication Channel*

The primary communication channel of the proposed interface is a visual one. For the sake of familiarity, it might be a good idea to use LED lights instead of actual intention displays. Habibovic et al. [48] mounted an LED stripe at the top of the windshield that was controlled via an Arduino microcontroller. Their interface, which can be seen in Figure 1, communicated four different states by expanding and shrinking the area of LED light that is activated. They later removed the third state that communicated that the car was waiting because several of their participants stated that it did not contribute to their experience, which makes somehow sense, since the waiting state is already encoded at the end of state 2. The authors were able to show that the perceived safety of their participants when they were approached by the autonomous car with the LED interface increased to the same level as when they were approached by a conventional car. The dynamic multimodal interface (DYNAMO) proposed in this paper will make use of the interface designed by Habibovic et al. While one can argue that the led stripe interface already satisfies the third AVIP principle, there are other approaches which are more suitable for the replacement of eye contact. One of these approaches is used in the AutonoMI interface, which was already mentioned in Section II-C. One problem with this approach is that it can only acknowledge the recognition of one pedestrian at a time. The proposed DYNAMO interface includes another LED stripe at the bottom of the windshield with a higher density of LED lights. This second LED stripe works similarly as the the AutonoMI interface [43], only that it detects groups of pedestrians as well as single pedestrians rather than only one pedestrian at a time. Whenever a group is detected by the AV, a circle of LED lights is activated on the LED stripe, based on the position of the group relative to the vehicle. A larger group results in a bigger circle. These circles then follow the movement of the pedestrian groups or the single pedestrians respectively. Since it could cause confusion to always show these circles, the lower LED stripe is only being used whenever the speed of the AV is below a certain threshold. A visualization of the interface can be seen in Figure 2.

### *C. Secondary Communication Channels*

As already discussed earlier, the DYNAMO interface makes also use of other modalities such as audition and tactition. When the AV detects bad weather conditions, it enables the auditive interface, which plays a sound whenever a pedestrian is noticed. The sound should not be disturbing and it is only activated if the vehicle travels below a certain speed threshold. By linking the frequency of the sound with the speed of the AV, the pedestrians are able to tell whether the vehicle is going to stop or if it does not slow down and is therefore not going to yield. If the hardware, the software and other factors such as speed and distance to the pedestrians allow it, the AV may also activate the auditive interface when it notices the presence of an elderly pedestrian. Quinn et al. [50] from the Nissan Technical Center Europe write about an audible pedestrian alert system called eVADER, which is an European Commission, part funded project. The system uses an acoustic beamforming loudspeaker array, which is mounted at the front of a vehicle to direct sound towards any detected vulnerable road user (VRU). In this paper, I propose to first use a simple, non-directed loudspeaker to test the basic concept of the auditive interface. If the results show that the concept is indeed effective, the auditive interface could be improved by drawing inspiration from the eVADER system and other existing interfaces. Furthermore, the proposed interface also includes V2P. However, the V2P communication channel is only used to communicate with pedestrians who exhibit some kind of impairment and may not be able to perceive the visual or auditive communication cues of the vehicle. Those pedestrians would have to install an application on their smartphones before the V2P communication could succeed. While we are primarily focussing on smartphones in this paper, the V2P application could also be installed on other wearable devices such as smartwatches. Of course, this would only make sense if there was a general agreement of car manufacturers that this kind of communication with impaired pedestrians via a smartphone application makes sense. The same also holds for the visual part of the interface: if the pedestrians are not familiar with the light patterns, they won't know how to interpret the signals. The DYNAMO interface should therefore be seen as a prototype and for the sake of simplicity, we will assume that the pedestrians were instructed or that they were already familiar with the interface before using it. When the AV detects a smartphone that has the application installed, it sends a vibrational pattern to the smartphone that encodes the intention of the vehicle. For example, a continuous vibration could be triggered, meaning that the car is not going to stop. On the other hand, short and long lasting vibrations could be triggered cyclically to inform the pedestrian that the AV is going to yield. Another approach would be to just use the vibration as a notification that the vehicle is establishing contact. Then, the information about the intentions of the car could be displayed on the smartphone or they could be spoken to the pedestrian in case of limited eyesight. Such a use of haptic and auditive feedback for pedestrians with impaired sight is nothing new and has already been proposed before, for example in the context of obstacle detection [51]. The advantage of this V2P communication with impaired pedestrians is that they could change the communication cues in the settings of their application such that the information is presented in the best way possible.

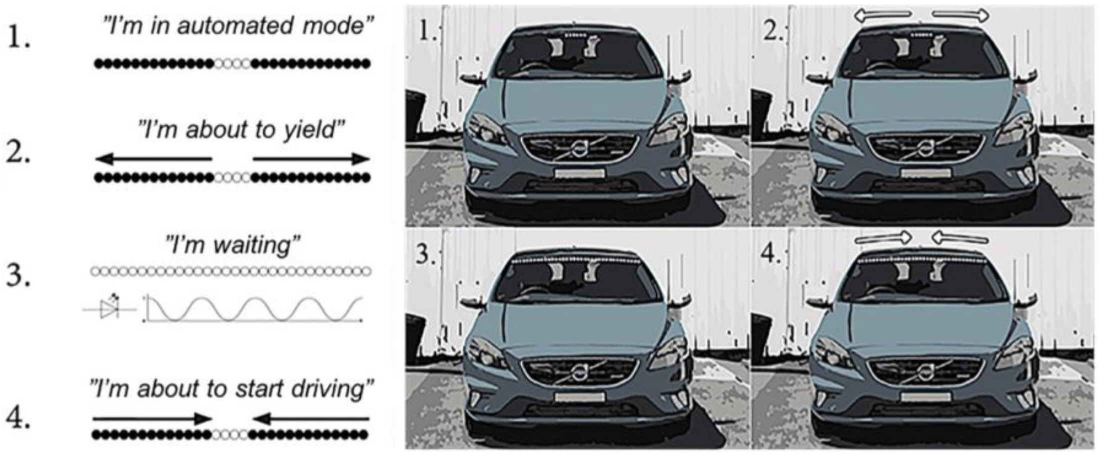

Fig. 1. Visualization of the four states in the interface proposed by Habibovic et al. [48].

### D. Limitations and Justification

Altough the idea of combining several modalities into one communication interface is not new [39], [35], [52], there do not seem to be a lot of studies exploring this idea further. Mahadevan et al. [39] were also using auditory and haptic cues in their study. They found that the auditory cues were liked by 6 out of 10 participants and that phone vibration was not very well received as it is a very subtle communication channel and the participants were not always confident whether the phone had vibrated. Therefore, it is important to have a vibrational pattern that is clearly distinguishable from other kind of notifications. This is achieved by giving the user the total control over how notifications are sent through her or his V2P application. The pedestrian may choose an individual vibrational pattern that she or he can easily detect. Furthermore, the visually impaired pedestrians should either wear their smartphone close to their body rather than in a bag to perceive the vibrations or they should use another device such as a smartwatch.

In a study by Hudson et al. [53], the participants preferred a verbal message saying *safe to cross* over music in a communication interface. They argue that the participants may have perceived the music as a negative and alarming signal to indicate that they should not cross the road. This finding may suggest that it would be better to use spoken messages than the proposed speed indicator sound. Nonetheless, it should also be stated that the DYNAMO interface does not simply use music but a specific sound that encodes information about the velocity of the AV. This approach is novel and has not been tested before to the best of my knowledge. Furthermore, the spoken message used in [53] did not comply with the fourth AVIP principle. By telling the pedestrians that it is safe to cross, the AV may endanger them if it is not able to see other potential threats in the scene. One positive aspect of the auditive modality in the DYNAMO interface is that it should be understood quite intuitively. Most pedestrians have a lot of implicit knowledge about the sound of cars and maybe even ambulances passing by. They are therefore familiar with the Doppler effect that causes them to hear an increased sound frequency when the car is approaching them and a decreasing sound frequency when the car has already passed. Therefore, if the DYNAMO interface emits a sound with a raising frequency, the pedestrians may implicitly perceive this as a cue to not cross the road. This makes a lot of sense, since the car would be accelerating in this case. On the other hand, if the frequency declines, this is similar to the case where the car has already passed the pedestrian and the perceived safety should be higher. Again, this makes sense, since the AV communicates that it is decreasing it's speed to yield to the pedestrian.

## IV. PROPOSED EVALUATION METHOD

### A. Threefold Study Design

With the ongoing COVID-19 pandemic, it is difficult to evaluate the DYNAMO interface right now. Therefore, I will propose a method to evaluate the interface such that it could be evaluated in the future. There are mostly two types of study that are used to evaluate intent communication interfaces: WOZ and VR studies. In general, WOZ studies may produce results that are more accurate, since the participants are experiencing the apparently autonomous vehicle rather than a merely virtual simulation of it. However, it can be expensive to build working prototypes of certain interfaces. In a VR study, the prototype could simply be programmed and thus the cost of a working prototype may decrease drastically. Moreover, recent studies were able to show the effectivenes of virtual environments in pedestrian safety research [54]. On the other hand, we cannot know for sure if the results of the VR study are accurate in a real world scenario. Chang et al. [38] argue that pedestrians may always feel safe in a virtual environment because their lives are never in danger. Furthermore, they state two weaknesses of VR studies: low-fidelity in VR and VR (motion) sickness. The authors state that the latter can be reduced by increasing the size of the interface elements in the VR environment and by slightly reducing the speed of the participants in the scene. Moreover, they suggest to arrange short breaks if the VR experiment takes longer than 30 minutes. Coming back to the proposed interface, there are parts that can be tested better in a virtual environment. For example, the AV might activate the auditive modality upon

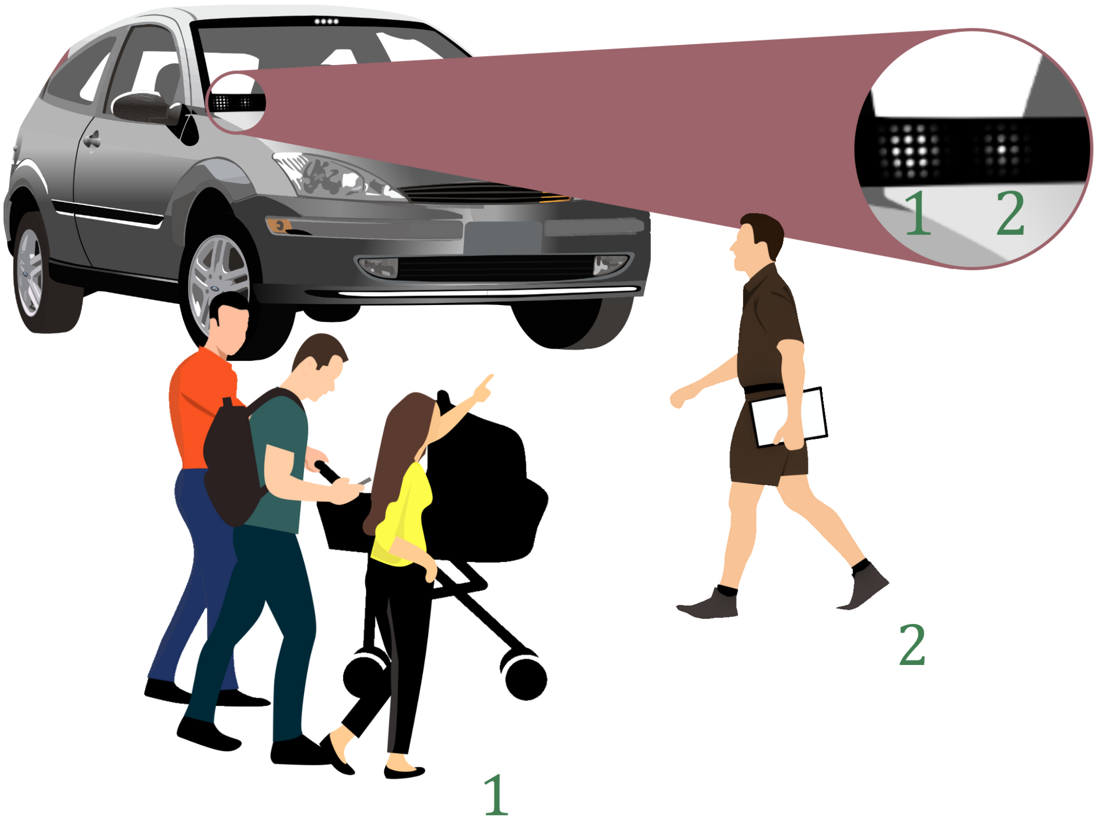

Fig. 2. The LED lights indicate that there are two groups of pedestrians present: one lager group on the left and one small group consisting of only one pedestrian at the right.

registering bad weather conditions and limited visibility of the street. It would be difficult to evaluate this condition in a WOZ study. A virtual environment could just simulate these weather conditions and, what's more, it could simulate the experience of a person with impaired eyesight. This way, we are able to collect more data for the *impaired eyesight* condition than in a WOZ study. But would these results really say something about encounters with AVs in the real world? To make sure that the results are somehow transferable to the real world scenario, we can divide the evaluation of the interface into three larger experiments, one that uses the WOZ principle and two that take place in a virtual environment.

The first experiment only include those parts of the interface that can be tested easily. It is important to have participants with different cultural backgrounds, since this may influence the assessment of the interface and the perceived safety of the participants. Moreover, we should also make sure that the participants cover a wide range of age. The second experiment includes the same parts of the interface as the WOZ experiment, but this time it is conducted in a virtual environment. If the results of the VR study do not differ significantly from the ones obtained in the WOZ study, we hope that the results of the third VR study, which then incorporates the entire interface, do also generalize well. Note that the two first experiments should be carried out with different groups of participants, since their behavior and their answers might be biased from the first experiment that they participated in. Nonetheless, the two groups should more or less include the same amount of demographic variability.

## B. *Structure of the Experiment*

Because we want to make sure that the results are really transferable, we have to test all of the modalities once in the WOZ study and once in the VR study. For the sake of simplicity, we will only include the upper part of the visual LED interface, which is the one proposed by Habibovic et al. [48]. If the transferability of the results is given, we can then also include the lower part of the visual interface into the larger VR study. Habibovic et al. evaluated their interface in three different stages, collecting data about the interpretation of the interface before it was explained to the participants and about the confidence in their interpretations of the signals after the interface was explained to the participants. Moreover, pedestrians had to state whether they would have crossed the street based on the received signals before the car had actually stopped. The experiment leaders also asked the participants to complete a Self-Assessment Manikin (SAM) questionnaire [55] to evaluate the affective reactions of the participants in three different dimensions and to answer some questions in a semi-structured interview. While the authors tested their interface in two different scenarios, namely once at a zebra crossing and once in a parking garage, we will only test the interface in one scenario. The experiment will take place at a street without zebra crossing. Otherwise, the participants may think that the AV has to stop either way because of the traffic rules. However, a good interface would also increase the perceived safety under less obvious conditions where there are no clear crossing rules. It is further proposed to choose a street and a time where there's not too much traffic and ideally there are no other pedestrians present to minimize the number of uncontrolled variables and the effect of group dynamics. There are two different conditions for the AV behavior (yielding vs. no yielding) and three different interface conditions (the standard visual interface, visual + auditive and visual+ auditive + haptic). This results in a total of six different conditions that have to be tested in each of the three experiments.

### C. WOZ Experiment

*1) Implementation Details:* The LED interface can be implemented analogously to how Habibovic et al. did it with an Arduino microcontroller. For the auditive part of the interface, we can simply use a speaker and play one of two recordings: a sound with a constant frequency for the condition where the car keeps it's speed and will not yield to the pedestrians and a sound whose frequency decreases while the car is slowing down before yielding to the pedestrians. For the V2P part of the interface, a simple call to a provided smartphone could be used. Note that we would have to assign different notification patterns to different caller ids to have two distinct vibrations for the two yielding conditions. The participants could choose one of the available vibration patterns before the experiment starts, such that they do perceive the preferred haptic cue.

*2) Evaluating Auditive and Haptic Cues:* The easiest way to test the effectiveness of the auditive and the haptic interface is to reduce the availability of the visual cues. This can be done using glasses that either introduce some blurriness into the visual field of perception or glasses that decrease the amount of light that the participants can perceive. While it would still be better to include participants that actually have an impaired sight, it is easier to find test subjects without the impairment and to introduce it artificially. Nonetheless, it is highly recommended to include participants with impaired eyesight if the availability is given.

*3) Procedure:* Since the number of participants in one experimental group is usually quite limited, it is proposed to use all of them in every condition. The advantage of this approach is, that they can then give feedback about the perceived differences in the different conditions. There are three conditions, which will be presented to the participants in a randomized order. The first condition only includes the visual part of the interface. It is proposed to to randomly sample the yielding condition three to four times per participant and for each modality that is tested. This way, the participants cannot know beforehand if the car is going to yield to them or if it won't. After having sampled the yielding conditions, the car approaches the participant while only activating the LED interface. In a second condition, the car approaches them while the participants wear the glasses that introduce an artificial impairment. Besides the visual interface, the car now also activates the auditive interface, playing a sound depending on the randomly sampled yielding condition. In the third condition, the haptic (V2P) component of the interface is activated as well, together with the visual and auditive components.

*4) Collected Data:* After the experiment, participants are asked to fill out a form where they are asked different questions about which modalities they preferred, if their expectations were met and how safe they felt during the task. It might be a good idea to include a SAM questionnaire as Habibovic et al. did. Moreover, more data could be collected about the time between the signalization of the interface and the decision of the participant to start crossing the street (if they decide to do so). The walking speed could also be measured. These measurements can be used to infer the observed determination or uncertainty of the participants. Of course, we would also analyze the effectiveness of the interface by looking at the number of times that the pedestrians actually crossed the street when the interface communicated that the car was going to yield to them. Note that it might be necessary to tell the participants that they should decide what to do before the car has completely stopped, since completely still car poses no risk to the pedestrians and is therefore no good indicator for the effectiveness of the interface.

### D. First VR Experiment

*1) Implementation Details:* In this case, we would have to implement everything in the virtual environment. Instead of using a speaker, we can now simply attach a virtual sound source to the virtual car. The LED interface can also be added quite easily to the car. For the haptic modality, we might still use the same method as before by simply calling the smartphone at the right moment or we could program the simulation to send a SMS to the smartphone just before the car approaches the pedestrian.

*2) Evaluation Auditive Cues:* Instead of using glasses, we can simply introduce a certain amount of blurriness to the image on the VR headset screen. We may also just decrease the brightness of the scene or use a combination of both approaches. However, it is important to keep the level of visual impairment to one that is similar to the one caused by the glasses in the WOZ experiment. We might want to consider using approaches that have already been tested for simulating visual impairment in augmented [56] or virtual reality [57].

*3) Procedure:* The procedure is the same as in the WOZ experiment, only that we now use the virtual counterparts of the interface rather than the physical prototype and another group of participants.

*4) Collected Data:* To ensure that the results are comparable, we have to collect exactly the same data.

### E. Second VR Experiment

*1) General Remarks:* If the results of the first two experiments do not significantly differ, we can proceed to the VR experiment that includes the whole interface. The only part that is still missing is the lower part of the visual LED interface. We might want to add a random number of *virtual pedestrians* in each crossing task to evaluate this part of the interface, since the size of the displayed circles should change with the amount of pedestrians that are registered. However, this also raises the question whether the participants are influenced by the crossing decisions of the virtual pedestrians. A solution would be to only include these virtual pedestrians for a fraction of the participants. If possible, a third group of participants should be used for this experiment for the same reason as before.

*2) Visual Impairment:* So far, we have only tested the efficiency of the auditive and haptic interface for people with an (artificial) visual impairment. But the DYNAMO interface also activates the auditive part of the interface when the weather conditions are bad and the visibility of the scene is reduced. We can test this by just changing the weather conditions of the scene for some test subjects. It is proposed to use this approach as a replacement of the artificial impairment and not as an additional factor.

*3) Procedure:* It is proposed to use the same procedure as in the other two experiments. The only thing that changes is the visual interface and sometimes the virtual weather conditions as described above.

*4) Collected Data:* Once again, we propose to collected the same data as in the first two experiments.

*5) Controllability vs. Demographic Diversity:* The visual and the auditive modalities of the interface could also be tested by a larger group of participants remotely. There are large groups of VR enthusiasts on Facebook and probably also on other platforms. The *OcculusQuestCommunity* [1] is an example of such a group that contains more than 21 000 members. The Oculust Quest [2] is an All-In-One VR headset that allows roomscale tracking without using additional sensors or cables. This would be ideal for carrying out the proposed VR experiment. However, the haptic part of the interface could not be tested and we would lose a great amount of controllability over the experiment. Therefore, this last step could be seen as an addition to the controlled VR experiment. It is quite possible that a lot of participants could be found in such Facebook Groups when they are told that the experiment offers insight into the research of AVs and possible technologies of the future. After participating in the experiment, the could also be asked to answer some questions, just as the participants in the controlled VR experiment.

## V. DISCUSSION

### A. Coverage of Behavioral Factors

While the proposed interface is not able to consider all of the factors that influence the crossing behavior of pedestrians which were presented in Section II, it does consider them partially. More specifically, the interface caters to elderly pedestrians by additionally activating the auditive modality. Whenever the auditive cues are active, the speed estimation is also alleviated through a direct link between the sound frequency and the vehicle speed. We could argue that there would be some more situations in which it is useful to activate the auditive modality of the interface, such as when the AV registers that a pedestrian is looking at his smartphone or another electronical device without paying much attention to the street. An important factor which is not considered by the DYNAMO interface is gap acceptance. This factor plays an important role in the crossing decisions of pedestrians. However, it makes more sense to include the empirical information about demographic gap acceptance in the algorithm that makes the decisions about yielding or not yielding to pedestrians and not in the intent communication interface. If a particular pedestrian perceives the gap as being too small, she or he probably won't cross the street, regardless of what the interface communicates.

### B. AVIP Compliance

The five principles of the AVIP are all addressed within the DYNAMO interface. First, the pedestrians are able to tell if the car is in autonomous mode just by looking at the visual interface. If it is not activated, the car is in manual mode. The second principle is also satisfied as the car communicates its future intentions through the upper LED stripe and maybe even through the auditive and V2P component of the interface. The replacement for eye contact is accomplished by activating the lower part of the interface, which shows LED circles that follow the pedestrians when they cross the street, such that they feel seen. Moreover, the interface does not make any kind of suggestion about crossing the street. It simply communicates the intentions of the car and lets the pedestrians decide how to handle the information. The interface is designed in a way that the interaction with the pedestrians should be relatively calm. However, this is the only principle that has yet to be evaluated more profoundly in the questionnaire at the end of the experiments. If the results show that the interaction with the proposed interface is not calm enough, the participants should be asked to state some suggestions which would improve their perceived calmness when interacting with the interface of the car.

## VI. CONCLUSION

There are a lot of different factors that influence the behavior of pedestrians in traffic situations. These factors may become more and more important with the emergence of self-driving vehicles. In order for the pedestrians to be safe, an intent communication interface for AVs has to be developed. A lot of interfaces have been proposed, but most of them are either only visual or they found that too much cues from different modalities could confuse pedestrians and lead to an information overload. This paper proposes a novel and dynamical interface called DYNAMO, which tries to solve this problem by only activating the different modalities of the interface when they are actually needed. Moreover, the proposed interface aims to be as comprehensible to visually impaired pedestrians as it is to pedestrians with intact vision. Some suggestions on how the interface could be evaluated using a WOZ and two VR experiments is given at the end of the paper.

## REFERENCES

[1] European Commission, Directorate General for Transport, "Traffic safety basic facts on pedestrians," https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/statistics/dacota/bfs20xx_pedestrians.pdf, June 2018.

---

[1] https://www.facebook.com/groups/OculusQuestCommunity

[2] https://www.oculus.com/quest/

[2] W. A. Harrell, "Factors influencing pedestrian cautiousness in crossing streets," *The Journal of Social Psychology*, vol. 131, no. 3, pp. 367–372, 1991.

[3] G. S. Wilde, "Immediate and delayed social interaction in road user behaviour," *Applied Psychology*, vol. 29, no. 4, pp. 439–460, 1980.

[4] J. C. Russell, D. O. Wilson, and J. F. Jenkins, "Informational properties of jaywalking models as determinants of imitated jaywalking: An extension to model sex, race and number," *Sociometry*, pp. 270–273, 1976.

[5] D. Yagil, "Beliefs, motives and situational factors related to pedestrians' self-reported behavior at signal-controlled crossings," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 3, no. 1, pp. 1–13, 2000.

[6] C. Holland and R. Hill, "The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations," *Accident Analysis & Prevention*, vol. 39, no. 2, pp. 224–237, 2007.

[7] A. Tom and M.-A. Granié, "Gender differences in pedestrian rule compliance and visual search at signalized and unsignalized crossroads," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1794–1801, 2011.

[8] W. A. Harrell, "Precautionary street crossing by elderly pedestrians," *The International Journal of Aging and Human Development*, vol. 32, no. 1, pp. 65–80, 1991.

[9] D. Sun, S. Ukkusuri, R. F. Benekohal, and S. T. Waller, "Modeling of motorist-pedestrian interaction at uncontrolled mid-block crosswalks," in *Transportation Research Record, TRB Annual Meeting CD-ROM, Washington, DC*, 2003.

[10] R. R. Oudejans, C. F. Michaels, B. Van Dort, and E. J. Frissen, "To cross or not to cross: The effect of locomotion on street-crossing behavior," *Ecological psychology*, vol. 8, no. 3, pp. 259–267, 1996.

[11] A. Willis, N. Gjersoe, C. Havard, J. Kerridge, and R. Kukla, "Human movement behaviour in urban spaces: Implications for the design and modelling of effective pedestrian environments," *Environment and Planning B: Planning and Design*, vol. 31, no. 6, pp. 805–828, 2004.

[12] M. S. Tarawneh, "Evaluation of pedestrian speed in jordan with investigation of some contributing factors," *Journal of safety research*, vol. 32, no. 2, pp. 229–236, 2001.

[13] R. L. Knoblauch, M. T. Pietrucha, and M. Nitzburg, "Field studies of pedestrian walking speed and start-up time," *Transportation research record*, vol. 1538, no. 1, pp. 27–38, 1996.

[14] R. Tian, E. Y. Du, K. Yang, P. Jiang, F. Jiang, Y. Chen, R. Sherony, and H. Takahashi, "Pilot study on pedestrian step frequency in naturalistic driving environment," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2013, pp. 1215–1220.

[15] D. R. Geruschat, S. E. Hassan, and K. A. Turano, "Gaze behavior while crossing complex intersections," *Optometry and vision science*, vol. 80, no. 7, pp. 515–528, 2003.

[16] I. E. Hyman Jr, S. M. Boss, B. M. Wise, K. E. McKenzie, and J. M. Caggiano, "Did you see the unicycling clown? inattentional blindness while walking and talking on a cell phone," *Applied Cognitive Psychology*, vol. 24, no. 5, pp. 597–607, 2010.

[17] S. Schmidt and B. Faerber, "Pedestrians at the kerb–recognising the action intentions of humans," *Transportation research part F: traffic psychology and behaviour*, vol. 12, no. 4, pp. 300–310, 2009.

[18] R. Sun, X. Zhuang, C. Wu, G. Zhao, and K. Zhang, "The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment," *Transportation research part F: traffic psychology and behaviour*, vol. 30, pp. 97–106, 2015.

[19] S. Das, C. F. Manski, and M. D. Manuszak, "Walk or wait? an empirical analysis of street crossing decisions," *Journal of applied econometrics*, vol. 20, no. 4, pp. 529–548, 2005.

[20] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, "Agreeing to cross: How drivers and pedestrians communicate," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 264–269.

[21] C. M. DiPietro and L. E. King, "Pedestrian gap-acceptance," *Highway Research Record*, no. 308, 1970.

[22] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, "Understanding pedestrian behavior in complex traffic scenes," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 61–70, 2017.

[23] M. Risto, C. Emmenegger, E. Vinkhuyzen, M. Cefkin, and J. Hollan, "Human-vehicle interfaces: the power of vehicle movement gestures in human road user coordination," 2017.

[24] A. Varhelyi, "Drivers' speed behaviour at a zebra crossing: a case study," *Accident Analysis & Prevention*, vol. 30, no. 6, pp. 731–743, 1998.

[25] N. Guéguen, S. Meineri, and C. Eyssartier, "A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing," *Safety science*, vol. 75, pp. 87–89, 2015.

[26] D. Dey and J. Terken, "Pedestrian interaction with vehicles: roles of explicit and implicit communication," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2017, pp. 109–113.

[27] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE transactions on intelligent transportation systems*, 2019.

[28] A. Millard-Ball, "Pedestrians, autonomous vehicles, and cities," *Journal of planning education and research*, vol. 38, no. 1, pp. 6–12, 2018.

[29] T. Lagström and V. Malmsten Lundgren, "Avip-autonomous vehicles' interaction with pedestrians-an investigation of pedestrian-driver communication and development of a vehicle external interface," Master's thesis, 2016.

[30] L. K. Vissers, S. van der Schagen, and M. INLG van & Hagenzieker, "Safe interaction between cyclists, pedestrians and automated vehicles: what do we know and what do we need to know?" 2017.

[31] H. Surden and M.-A. Williams, "Technological opacity, predictability, and self-driving cars," *Cardozo L. Rev.*, vol. 38, pp. 121–182, 2016.

[32] H. Prakken, "On the problem of making autonomous vehicles conform to traffic law," *Artificial Intelligence and Law*, vol. 25, no. 3, pp. 341–363, 2017.

[33] B. Hanington and B. Martin, *Universal Methods of Design Expanded and Revised: 125 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers, 2019, p. 254.

[34] S. Yang, "Driver behavior impact on pedestrians' crossing experience in the conditionally autonomous driving context," 2017.

[35] M. Matthews, G. Chowdhary, and E. Kieson, "Intent communication between autonomous vehicles and pedestrians," *arXiv preprint arXiv:1708.07123*, 2017.

[36] M.-P. Böckle, A. P. Brenden, M. Klingegård, A. Habibovic, and M. Bout, "Sav2p: Exploring the impact of an interface for shared automated vehicles on pedestrians' experience," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*, 2017, pp. 136–140.

[37] M. Clamann, M. Aubert, and M. L. Cummings, "Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles," Tech. Rep., 2017.

[38] C.-M. Chang, K. Toda, D. Sakamoto, and T. Igarashi, "Eyes on a car: an interface design for communication between an autonomous car and a pedestrian," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2017, pp. 65–73.

[39] K. Mahadevan, S. Somanath, and E. Sharlin, "Communicating awareness and intent in autonomous vehicle-pedestrian interaction," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.

[40] R. Zimmermann and R. Wettach, "First step into visceral interaction with autonomous vehicles," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2017, pp. 58–64.

[41] M. Beggiato, C. Witzlack, S. Springer, and J. Krems, "The right moment for braking as informal communication signal between automated vehicles and pedestrians in crossing situations," in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2017, pp. 1072–1081.

[42] Daimler Communications, "Mercedes-benz f 015 luxury in motion," https://media.daimler.com/marsMediaSite/ko/en/9920141, 2015, press Information.

[43] L. Graziano, "Autonomi autonomous mobility interface," https://vimeo.com/99160686, 2014, m.A. thesis presentation.

[44] (2015) The mercedes-benz f 015 luxury in motion. [Online]. Available: https://www.mercedes-benz.com/en/innovation/autonomous/research-vehicle-f-015-luxury-in-motion/

[45] J. J. Anaya, P. Merdrignac, O. Shagdar, F. Nashashibi, and J. E. Naranjo, "Vehicle to pedestrian communications for protection of vulnerable road users," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE, 2014, pp. 1037–1042.

[46] P. Sewalkar and J. Seitz, "Vehicle-to-pedestrian communication for vulnerable road users: Survey, design considerations, and challenges," *Sensors*, vol. 19, no. 2, p. 358, 2019.

[47] (2014) Survey: Bicyclists unsure of connected and automated car technology. [Online]. Available: https://bikeleague.org/content/survey-bicyclists-unsure-connected-and-automated-car-technology

[48] A. Habibovic, V. M. Lundgren, J. Andersson, M. Klingegård, T. Lagström, A. Sirkka, J. Fagerlönn, C. Edgren, R. Fredriksson,

S. Krupenia *et al.*, “Communicating intent of automated vehicles to pedestrians,” *Frontiers in psychology*, vol. 9, p. 1336, 2018.
[49] M. J. Eppler and J. Mengis, “The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines,” *The information society*, vol. 20, no. 5, pp. 325–344, 2004.
[50] D. Quinn, J. Mitchell, and P. Clark, “Development of a next-generation audible pedestrian alert system for evs having minimal impact on environmental noise levels–project evader,” in *Proc. of Inter-Noise*, 2014.
[51] A. M. KASSIM, T. Yasuno, H. Suzuki, M. S. M. ARAS, H. I. JAAFAR, F. A. JAFAR, and S. Subramonian, “Conceptual design and implementation of electronic spectacle based obstacle detection for visually impaired persons,” *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 10, no. 7, pp. JAMDSM0094–JAMDSM0094, 2016.
[52] E. Florentine, M. A. Ang, S. D. Pendleton, H. Andersen, and M. H. Ang Jr, “Pedestrian notification methods in autonomous vehicles for multi-class mobility-on-demand service,” in *Proceedings of the Fourth International Conference on Human Agent Interaction*, 2016, pp. 387–392.
[53] C. R. Hudson, S. Deb, D. W. Carruth, J. McGinley, and D. Frey, “Pedestrian perception of autonomous vehicles with external interacting features,” in *International Conference on Applied Human Factors and Ergonomics*. Springer, 2018, pp. 33–39.
[54] S. Deb, D. W. Carruth, R. Sween, L. Strawderman, and T. M. Garrison, “Efficacy of virtual reality in pedestrian safety research,” *Applied ergonomics*, vol. 65, pp. 449–460, 2017.
[55] M. M. Bradley and P. J. Lang, “Measuring emotion: the self-assessment manikin and the semantic differential,” *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
[56] H. C. Ates, A. Fiannaca, and E. Folmer, “Immersive simulation of visual impairments using a wearable see-through display,” in *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, 2015, pp. 225–228.
[57] J. Lewis, L. Shires, and D. Brown, “Development of a visual impairment simulator using the microsoft xna framework,” in *Proc. 9th Intl Conf. Disability, Virtual Reality & Associated Technologies, Laval, France*, 2012.

**Luca Rolshoven** was born in Bern, Switzerland in 1995. He received the Bachelor’s Degree in Computer Science from the University of Bern in 2020. Currently, he is pursuing the Swiss Joint Master of Science in Computer Science at the universities of Bern, Neuchâtel and Fribourg with a specialization in data science.