Univ Access Inf Soc (2015) 14:527–536
DOI 10.1007/s10209-014-0350-7

LONG PAPER

Right mix of speech and non-speech: hybrid auditory feedback in mobility assistance of the visually impaired

Ibrar Hussain · Ling Chen · Hamid Turab Mirza · Gencai Chen · Saeed-Ul Hassan

Published online: 26 March 2014

Abstract Despite the growing awareness about mobility issues surrounding auditory interfaces used by visually impaired people, designers still face challenges while creating sound for auditory interfaces. This paper presents a new approach of hybrid auditory feedback, which converts frequently used speech instructions to non-speech (i.e., spearcons), based on users' travelled frequency and sound repetition. Using a within-subject design, twelve participants (i.e., blind people) carried out a task, using a mobility assistant application in an indoor environment. As surfaced from the study results, the hybrid auditory feedback approach is more effective than non-speech and it is pleasant compared with repetitive speech-only. In addition, it can substantially improve user experience. Finally, these findings may help researchers and practitioners use hybrid auditory feedback, rather than using speech- or non-speech-only, when designing or creating accessibility/assistive products and systems.

I. Hussain · L. Chen (✉) · G. Chen
College of Computer Science, Zhejiang University,
Hangzhou 310027, Zhejiang, People's Republic of China
e-mail: lingchen@zju.edu.cn

I. Hussain
e-mail: ibrar@zju.edu.cn

G. Chen
e-mail: chengc@zju.edu.cn

H. T. Mirza
Department of Computer Science, COMSATS Institute of Information Technology, Lahore 54700, Punjab, Pakistan
e-mail: drturab@ciitlahore.edu.pk

S.-U. Hassan
Information Technology University, Lahore, Punjab, Pakistan
e-mail: saeedulhassan@gmail.com

Keywords Mobility · Auditory feedback · Speech · Non-speech · Hybrid · Visually impaired

1 Introduction

Mobility is a major characteristic of our times. Some tasks that we used to accomplish in the office or at home are now carried out on the go. Though, mobility requires a high degree of visual attention; however, previous research [1–4] shows that in the absence of vision or in the eye-busy environment, focus can be shifted from vision to hearing for obtaining information about the surrounding environment of the user. Blindness is a larger segment of the world's human population: It is reported [5] that worldwide, an estimated 40–45 million people are legally blind, 135 million have low vision, 314 million have some kind of visual impairment. It is also anticipated that the number will be double by the year 2020.

Recent developments in the field of mobility assistance for people with visual disability have led to a renewed interest in auditory interfaces. These interfaces are either speech based or non-speech based. Speech-based interfaces are based on human-speech, which can be *recorded, processed, played, or synthesized by a computer*. On the contrary, non-speech interfaces are based on *music, environmental sounds, or different artificial sound effects*. In mobility assistance, speech is used to convey information about the immediate environment of the user with minimal ambiguity; however, repetitive speech-only instructions become annoying and irritating for the visually impaired users and imposes higher cognitive load [6, 7].

Major non-speech audio cues (e.g., auditory icons, earcons and spearcons) are short and pleasant compared with repetitive human-speech. However, they are inadequate in presenting critical and detailed information. On the other hand, auditory icons are brief sounds and can be seen as an alternative to the visual icon in the computer interface. They are defined by Gaver [8] as "everyday sounds mapped to computer events by analogy with everyday sound producing events." Earcons are defined as a non-speech sound technique used in computer interfaces to convey or represent information to the user regarding entities, processes or interactions [9]. Brewster et al. [10] explained that they are "abstract, synthetic sounds that can be used in structured combinations to create sound messages to represent parts of an interface." Simply, they are based on music or other artificial sound effects, such as ringtones. Finally, spearcons can be created, "when spoken phrase is speeded up to the limit that it cannot be identified/recognized as speech." If speech sound is compressed up to 40–50 %, while the pitch is not squashed at all, it can result in the birth of new spearcons [11]. Spearcons are acoustically unique, as Palladino and Walker [12] elaborated that a user can hardly get confused with the use of successive spearcons. Since, each phrase when compressed results in a unique sound. Therefore, they can be easily differentiated among each other. In addition, earlier studies [11, 13–15] have reported that spearcons are very convenient in learning and may relatively require a short training period compared with auditory icons and earcons. Keeping these relative strengths of spearcons over earcons in view, the authors have chosen spearcons as a non-speech audio mode for this study.

The main contribution of this paper is to analyze the new approach of *gradual conversion of sound feedback from speech to non-speech (i.e., spearcons) based on sound repetitiveness and route travelled frequency of the user which may result in improved mobility experience.* By combining the good aspects and eliminating the inadequacies of speech and non-speech sound, this paper proposes the idea of 'hybrid auditory feedback' (combination of speech and spearcons) in mobility assistant applications for people with visual disabilities. Hence, the diminished or lack of the visual channel as a source of information for users with visual disabilities is compensated with the information obtained from sense of hearing. In this way, the auditory feedback that a user gets should convey meaningful information along with minimal irritation caused in the user's ear. Therefore, it is suggested that providing users with an auditory feedback in hybrid sound may result in a better user experience as compared with repetitive speech-only and non-speech audio feedback.

This paper is organized in the following way: Sect. 2 covers in detail the related work that has been carried out regarding auditory feedback for mobility assistance and pedestrian navigation in an eyes-free environment. Section 3 describes the proposed approach of hybrid auditory feedback. Section 4 presents a user study and experimentation carried out with the involvement of twelve legally blind participants. Section 5 reports results from the experiment. Section 6 discusses the research questions raised in the user study. Finally, Sect. 7 concludes the paper and proposes new avenues for future research.

2 Related work

The research presented in this paper is motivated by previous attempts that have been made in using sound as an auditory feedback for mobility assistant systems in an eyes-free environment. AudioGPS is a spatial audio navigation interface for global positioning system (GPS), and it is designed and developed by Holland et al. [16]. The proposed interface provides the user with non-speech audio (i.e., spatial audio) to convey distance and direction information while carrying out the location tasks. A similar approach is adopted by Strachan et al. [7] in GPSTunes. Using the system, distance from the target destination is conveyed to the user by change in the volume of a song, for example, when the user is getting close to the destination the volume of a song increases. Furthermore, the directional information in the GPSTunes is conveyed by panning the sound 'source' around the user's head.

Ghiani et al. [17] proposed a location—aware museum guide for the people with visual impairment. The guide has a vocal user interface (VUI), which exploits an embedded text-to-speech (TTS) engine provided by Loquendo,[1] which synthesizes speech for describing artworks/sections and for giving direction tips on the fly ("… rotate left…", "… carry on in this direction",…"please, stop!"*).* TravelMan, a mobile multimodal pedestrian and public transport route guidance application is being introduced by Kainulainen et al. [18]. The application uses non-speech sound (i.e., auditory icons), which complements speech in conveying route information to the user. In experiments with the application, basic awareness information regarding four modes of public transportation, for example sound of walking, trains, tram and buses, are conveyed to the user in the form of auditory icons. Furthermore, the temporal distance information is conveyed, instead of spatial distance. This is achieved with the use of 'pause' in a speech output, which varies considerably and depends upon the complexity of the route and addresses of the destination. Similarly, another system called Ontrack, designed and developed by Jones et al. [19] also assists users in pedestrian navigation. The system conveys directional information to the user through non-speech sound (i.e., music) which helps the user reach to the target destination. The system works in such a way that if

[1] http://www.loquendo.com.

a user is walking in the direction of the target destination the music plays in both ears via headphones. However, if the direction of the destination is on the right side of a user the music will start playing in the right ear and vice versa, if the target destination is on the left side of the user. In case of a wrong direction, the volume of the music played would gradually decrease.

In a recent study conducted for the interior sound design of speed trains project (ISHT), Hasen and Bresin [20] have designed a non-speech sound as auditory feedback to represent distance between two train stations. The sound used for the purpose is an iconographic representation called soundscapes. In their proposed design of audio feedback, when the train is approaching to the station they mix different sounds (e.g., traffic, train station and crowds) and when the train gets further away from the station they use sounds of birds and forest which resembles with the sound of the natural environment. Vazquez-Alvarez et al. [21] designed an urban sound garden which supports overlapping audio landmarks. Symbolic earcons were used, i.e., animal sounds which represented physical landmarks. They carried out an experiment in which a recording of an animal sound, i.e., goose, cricket, nightingale, frog and owl represented five different landmarks in the environment. Results from the experiment showed that it provides the user with great sense of immersion and pleasure and can be used in the vicinity of smartphones to represent nearby surroundings of the users. The earcons used by Vazquez-Alvarez et al. [21] were designed based on the previous study of Stahl [22] and McGookin [23].

More recently, many mobility assistant systems have been developed for people with visual impairment, representing both indoor and outdoor environments. These assistant systems [7, 16] either rely totally on non-speech (e.g., auditory icons, earcons and spearcons) sounds or dependent on speech (mainly, synthetic) for information transmission about the nearby environment of the user, such as, the personal guidance system (PGS) [24, 25], mobility of blind and elderly people interacting with computers (MoBIC) [26] and the Drishti System [27]. Moreover, though the TravelMan system [18] and the system for wearable audio navigation (SWAN) [28] combine speech and non-speech sound in the interface, none of the earlier research intelligently combined speech and non-speech in the auditory feedback. Furthermore, a new approach to gradually change the auditory feedback from speech to non-speech (i.e., spearcons) based on parameters, such as user travelled frequency and sound repetitiveness in the mobility assistant system is suggested. In the authors' opinion, when users get familiar with the environment in mobility assistant applications, speech sound becomes repetitive and annoying in the user's ears, and thus, non-speech sound is the best alternative.

3 Approach

In an early study, Hussain et al. [29] proposed a new form of audio feedback, i.e., hybrid auditory feedback. This audio feedback is based on measures, such as sound repetitions and frequency of the same route travelled by the user. On the basis of these parameters, the approach presented in this paper achieves the right mix between speech-only and non-speech (i.e., spearcons-only) auditory feedback to the visually impaired. The two parameters are described as follows:

1. *Sound repetitiveness* is a measure on the basis of which it is argued that speech-only instructions can be converted to non-speech (i.e., spearcons). In this paper, sound repetitiveness is defined as repetition of speech instruction to the user while they are indoors and on the move. Though, this measure is dependent on the distance traveled (e.g., 200, 250, 270 m, etc.). However, for the user study, a hallway that is 290 m long having multiple doors and turns was selected.
2. *Route travelled frequency*. It is argued that in addition to sound repetitiveness, there is another measure that may help in conversion of speech instruction to spearcons. This measure is the route travelled frequency which can be defined as the number of times a user opted for the same route to reach a selected destination while navigating indoors.

In the following, a real-life scenario that highlights the significance of the presented approach in auditory feedback is provided. For instance, Mr. X is visually impaired and an employee in a company with his office on the third floor. Every day when he enters the office building, he follows the same path. Thus, over a period of time speech instructions become very repeating/irritating. Therefore, the approach proposed is to convert all the repetitive speech instructions to non-speech (i.e., spearcons), which makes the experience less-irritating and more enjoyable. Same goes from his desk to washroom, his desk to lunch room, etc. However, once every month, he needs to go to first floor to submit his monthly bills. Obviously, this is not his very frequent/regular path; therefore, auditory feedback on a less frequent path may remain as speech-only.

In the following, the definitions of the terms used in the paper are presented:

Definition 1 *Speech Auditory Feedback (SF)* A mode of audio feedback in which all information regarding the immediate environment of the visually impaired person is provided in speech.

Definition 2 *Non-Speech Auditory Feedback (NF)* A mode of audio feedback in which all information regarding the immediate environment of the visually impaired is provided in spearcons-only.

Part 1	Part 2
Room Number	04
Room Number	05
Room Number	06

Fig. 1 Breakdown of the repetitive part (i.e., part 1) and dynamic part (i.e., part 2) in the speech instruction

Definition 3 *Hybrid Auditory Feedback (HF)* The audio in this mode is provided to the user in such a way that repetitive speech-only instructions are converted to non-speech (i.e., spearcons) and less repetitive speech instruction do not change and remain as speech.

For the implementation of hybrid auditory feedback, consider an example based on the current study environment.

Figure 1 shows that 'Part 1' (i.e., Room Number) remains constant, and there is no change in the representation. However, 'Part 2' (i.e., 04, 05 and 06) is dynamic and keeps on changing. The spearcons of the constant speech instructions are produced by speeding up part of the speech to the limit where it is not recognizable as speech and the pitch remains unchanged. The language of the spearcons used in the study is English.

For this study, an already developed mobility assistance application called BlueView [30] for smartphones was modified. The system is developed to assist visually impaired people in improving their perception of points of interest (POIs) in the nearby surrounding. BlueView has two components: Viewer device and Beacon points (BPs). Viewer device is a smart phone which is carried by users. Beacon point is a Bluetooth tag. BP's are small in size and can be mounted on any object of interest for the users. The name of a beacon point, e.g., the number of a room, is banded with its MAC address, which is unique to every Bluetooth device. The viewer device can uniquely identify BPs by their MAC addresses. For the evaluation of the hybrid auditory feedback approach, multiple Bluetooth tags were placed on the path of the user from his/her starting position to the target destination. As the participants' path is familiar, all the route instruction was employed in the viewer device and as the participant reaches the tag, instructions are played to the user in the different audio modes (e.g., speech-only, hybrid or non-speech) based on the proposed measures (i.e., users' traveled frequency of the path and sound repetitiveness).

Furthermore, to record participants' interest with the type of auditory feedback, following a few trials when the hybrid audio is achieved, participants were provided with an interface on the viewers' device on which they can select between speech-only and non-speech (i.e., spearcons-only). The selection can be made by tapping the screen once for speech-only which will make all instructions played along the path in speech-only mode. To select the non-speech listening, the participants have to tap the screen twice in quick succession and all the instructions along the route will be played in non-speech mode (i.e., spearcons-only). If a participant does not tap the screen, the audio listening mode remains as hybrid (i.e., speech plus spearcons). It is to note that participants can only make the selection before the start of each trial. Once the trial is started, they cannot shift between the speech-only and non-speech and the tapping feature in the application is disabled.

4 User study

The purpose of the current study is to evaluate speech-only, non-speech (i.e., spearcons-only) and hybrid (i.e., speech plus spearcons) as auditory feedback in mobility assistance systems for the visually impaired, in the context of indoor environment. It is assumed that hybrid sound provides adequate information compared with non-speech, and at the same time, it is more enjoyable and less-irritating compared with repetitive speech-only.

4.1 Research questions

This user study aims to answer five main research questions:

1. How many times does listening to the same speech instruction become repeated/annoying for the visually impaired?
2. Does the hybrid auditory feedback convey meaningful information compared with non-speech sound?
3. What factors may decide that the auditory feedback based on speech may be converted to non-speech (i.e., spearcons)?
4. What type of auditory feedback (i.e., hybrid, speech-only) successfully guides the visually impaired in relatively less time to the target destination in a familiar indoor environment?
5. Is a user experience with a hybrid auditory feedback more pleasant as compared with a repetitive speech-only or non-speech?

4.2 Participants

Twelve blind participants, ten male and two female, were recruited for the study from a local blind organization. Their age ranged from 20 to 65 with a mean of 36 (SD = 11.4) years old. All participants were informed they would be completing surveys and documenting items to which they

Table 1 Description of participants involved in the study to evaluate auditory feedback

Users	Gender	Age	Description of vision impairment
P1	Male	20	Blind
P2	Female	28	Blind
P3	Male	38	Retinitis pigmentosa
P4	Female	25	Blind
P5	Male	32	Blind
P6	Male	45	Blind
P7	Male	65	Blind
P8	Male	42	Severe global loss of acuity
P9	Male	36	Blind
P10	Male	35	Blind
P11	Male	24	Blind
P12	Male	34	Blind

fully consented. Blind participants in the context of the study are defined as people who are legally blind (visual acuity 20/400 or less). Participants were tested and found having no significant hearing loss (as assessed by an audiogram) and they also did not suffer from any memory loss. Among all of the participants, ten were totally blind (presumed to be in complete vision loss) and two had low vision, e.g., one participant (P9) had severe global loss of acuity, and one (P3) was suffered from retinitis pigmentosa. Finally, eight of the blind participants were congenitally blind (i.e., no visual experience); two participants had become blind by the age of 10 and 15 years old (i.e., have some visual experience) and two participants are with low vision. All participants reported having O&M (Orientation and Mobility) training and self-ratings on various mobility tasks (1 = Well below average and 5 = Well above average), provided mean ratings between 2.5 and 3.7, with a group mean of 3.4. For research focusing on users with impairments, it is generally acceptable to have 8–10 users to take part in the study [31]. Table 1 shows the details of each participant, including gender, age and description of the visual impairment.

4.3 Smart phone usage

Only one participant (P7) stated that he did not use smartphone, because he feels uncomfortable with the complexity and accessibility functionalities of a smartphone. He rather prefers to use simple mobile phone to make/receive phone calls. The other 11 of 12 usually carry smart phones, as well as a white cane while on the move.

4.4 Apparatus

Participants executed the experimental tasks using smartphones, such as, Samsung Galaxy SII and Sony Ericsson. The smartphone ran Android and an application (i.e., customized version of the indoor mobility assistant application, BlueView [30]). All participants' actions were logged, and the experiment was video tapped to observe the participants' behaviors.

4.5 Procedure

Before commencement of the experiment, all participants were thoroughly briefed about the purpose of the study. Once the participants understood the purpose, the researchers elaborated on the idea of introducing hybrid auditory feedback in mobility assistant applications based on sound repetition and users' travelled frequency of the route. Furthermore, three training sessions (based on the participant's availability) were conducted on multiple days. The purpose of the training sessions was to develop a better understanding of the participants regarding non-speech audio mode (i.e., spearcons-only) used in the experiment. During the training, initially speech-only examples, such as, turn right, turn left, walk straight, etc., were played, followed by a spearcons version of the above examples, played using a laptop with external amplified speakers. The training sessions with the participants were conducted in the laboratory conference room. Each training session lasted for one hour. Following the training sessions, the researchers explained and exemplified how to interact with the mobility assistant application. Initially, most participants were reluctant to interact with the mobility assistant application. However, they seamlessly coped with the application and easily understood the purpose of the study. Participants carried out the task using within-subject design, and the order was counterbalanced.

4.6 Task

The experiment was carried out in a real environment. A suitable hallway approximately 290 m in length on the ground floor of a building having several doors and turns was selected. Figure 2 shows a sketch of the ground floor of the building where the user study was conducted.

The tasks were carried out on three different days. Because of the participants' limited availability and single indoor environment, the participants were divided into three groups. Each group consisted of four individuals. Group I consisted of P1, P2, P5 and P6, whereas, group II comprised of P3, P4, P7 and P8, and group III involved P9, P10, P11 and P12. Each day participants repeated these tasks ten times in limited time duration. The task was to walk in a hallway with the help of the mobility assistant application deployed on the smartphones and listen to the audio feedback instructions while they walk until the destination was reached. Based on the path followed,

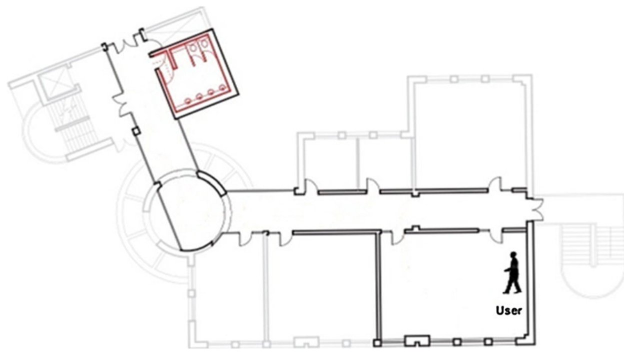

Fig. 2 Sketch of a building used in the experiment, where user in the figure shows starting position of the participant and *red square* shows the target destination

Table 2 Sample of auditory feedback instructions used in the task

Category	Instructions	Suffix distance	Suffix object
Directions	Go ahead		
	Take turn to your (left | right)	1 m	Door
	Go ahead to the corner and then turn right	4 m	
	Stop		
Door	Now on your (left |right) Room No (1st |2nd |3rd |4th |5th |6th |7th |8th |9th)		
	You reached your destination room no (10th)		

Direction instructions can be combined with distance and object suffix. Instruction parts shown in parenthesis are alternate to more commands, e.g., "turn (right | left)" stands for two instructions "turn right" and "turn left"

participants listened to the auditory feedback instructions. Sample instructions are shown in Table 2. These instructions were clustered into two main categories: "Directions" is a category which contains instructions related to route guidance, e.g., "go ahead", "turn left", "turn right", etc. These instructions can be combined with two suffixes (i.e., distance and object) to indicate the required distance to walk or an object (i.e., door, corner) to reach. In case of distance, the combined instruction reads as 'go ahead for four meters'. In case of object, the combined instruction reads as, 'go ahead to the corner'. These two suffixes can be combined, e.g., 'go ahead for four meters until the corner'. The category *"Door"* contains instructions related with specific rooms. The instructions in this category are considered meaningful. This category was included because user may be interested to know his/her exact location (e.g., a specific door or room).

There were three listening conditions for the auditory feedback instructions; speech-only (SF), hybrid (i.e., speech plus spearcons) (HF), and non-speech (i.e., spearcons-only) (NF). At the end of each task, all the participants were provided with a questionnaire that had five measures of the impression of the task for each type of audio listening. The numbers in parentheses refer to the scores of a 7-level assessment: Feeling Good (7)–Uncomfortable (1), Enjoyed (7)–Painful (1), Love (7)–Hatred (1), Fun (7)–Boring (1), Calm (7)–Annoyed (1). In addition to these questions, data regarding their *choice* of audio listening and the *time* to complete each task, using specific type of audio mode (i.e., speech-only, non-speech and hybrid) was also recorded.

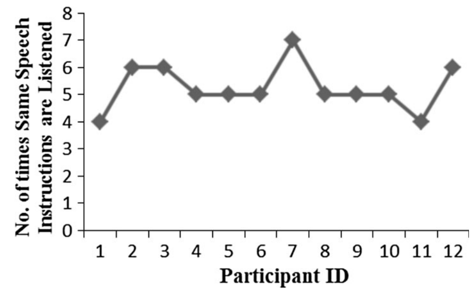

Fig. 3 Blind participants listened to similar human-speech instructions

4.6.1 *Threshold identification for repetitive human-speech*

In the laboratory environment, a small experiment was conducted with the participants involved in the user study. In the experiment, the participants were requested to listen to the same speech instructions continuously for the identification of thresholds of repetitive speech. Then, they were asked "how many times listening to the same speech instruction becomes repeated/annoying." Though the responses to the above question varied, the mode was five times. Similar to listening to the same speech instruction for five times, such human-speech may be regarded as repeated/annoying speech instruction and participants would like to change that speech instruction to non-speech sound.

The graph in Fig. 3 demonstrates that participants have listened to the identical speech instruction several times. Overall, six participants (i.e., P4, P5, P6, P8, P9, P10) agreed that listening five times to same speech instructions makes it repetitive for them. Whereas, three participants (i.e., P2, P3, P12) responded to it six times, two participants (i.e., P1, P11) stated it four times, and the only participant (i.e., P7) specified that listening to the identical human-speech seven times makes it a repetitive speech instruction for him.

5 Results

The mean task completion time while listening to the speech-only as an auditory feedback for the participants,

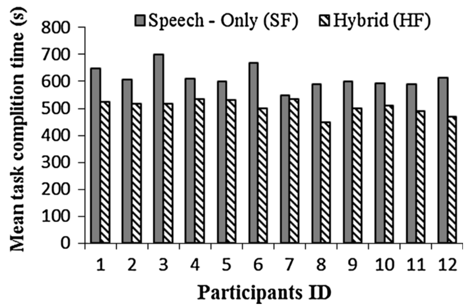

Fig. 4 Mean task compilation time for the participants using speech and hybrid auditory feedback

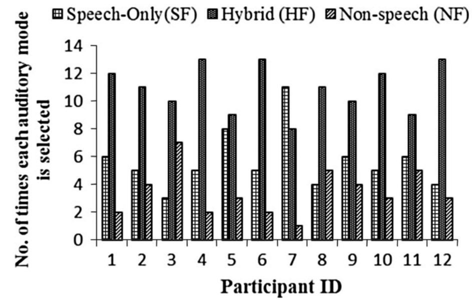

Fig. 5 Audio mode selection in the auditory feedback over a period of time following same path

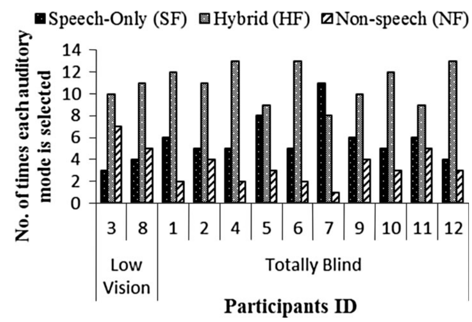

Fig. 6 Preference of audio mode selection in the auditory feedback over period of time by participants with low vision and totally blind

i.e., P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12 were 628, 609, 609, 610,619, 620, 639, 624, 620, 632, 641 and 641 s, respectively. Similarly, the mean task completion time for the hybrid auditory feedback for each participant, i.e., P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12 were 524, 517, 517, 534, 540, 524, 536, 498, 539, 528, 531 and 535 s, respectively. Compared with speech-only, it is evident from the graph in Fig. 4 that using a hybrid auditory feedback, participants needed less time to successfully reach the target destination. Further statistical tests, such as one-tailed paired *t* test, analyses of variance (ANOVA) and chi-square test were applied to the results.

Comparisons between the two auditory feedback modes, i.e., HF (speech plus spearcons) and NF (spearcons-only) regarding the conveyance of meaningful information to the participants were made using one-tailed paired t test with a 5 % statistical significance level. The results from paired t test revealed that the difference between HF and NF was statistically significant (tcal $= 6.9432$, $t = 1.895$, $df = 7$, $p < 0.05$).

Hence, HF conveys more meaningful information to the participants about the nearby environment than NF. While using ANOVA, the independent variable is feedback factor, i.e., NF, HF and SF (speech-only). In particular, for "Feeling good–Uncomfortable", a statistical test showed that SF and NF are more uncomfortable than HF ($F(2,14) = 9.32$, $p < 0.01$). For "Love–Hatred", a statistical test demonstrated that SF and NF are more dislikeable than HF ($F(2,14) = 8.79$, $p < 0.01$). For "Calm–Annoyed", a statistical test illustrated that participants were more annoyed with SF and NF compared with HF ($F(2,14) = 15.58$, $p < 0.01$). For "Enjoyed–Painful", a statistical test depicted that SF and NF were painful than HF ($F(2,14) = 13.65$, $p < 0.01$). There was no difference in HF compared with NF and SF for "Fun–Boring" ($F(2,14) = 3.21$, $p < 0.10$). The chi-square test showed a significant difference between HF, SF and NF ($p < 0.05$).

The result in Fig. 5 shows that in total, each participant listened to the auditory feedback instructions (i.e., speech-only, hybrid and non-speech) twenty times in the experiment. It is noted that participants started with listening to speech-only as an auditory feedback. However, over a period of time and with the users' familiarity with the environment, they were comfortable with the hybrid auditory feedback. During the experiment, participants also tried listening to non-speech (i.e., spearcons-only) audio. But in comparison, participants listened to non-speech audio fewer times than speech—only and hybrid. For instance, overall participants listened to speech-only mode 74 times, hybrid mode 120 times, and non-speech audio mode 46 for reaching to the target destination in the experiment. In Fig. 6, it is evident that participants with low vision listened to non-speech (i.e., spearcons-only) auditory feedback more times compared to speech-only audio mode. Nevertheless, all participants relied on hybrid auditory feedback more times compared with speech-only and non-speech for reaching to the target destination.

From the results and a detailed posterior interview held with all participants, it is observed that hybrid sound is more convenient as compared with non-speech and more enjoyable as compared with repetitive speech-only. A possible explanation for this might be that non-speech cannot possibly convey meaningful information of the immediate indoor environment to the visually impaired.

Secondly, speech-only instructions take longer time and with the repetition it becomes irritative as compared to hybrid. Whereas, hybrid auditory feedback takes less time and produces minimal irritation in user ears.

5.1 Participants' comments and preference

One participant commented, "Hybrid auditory feedback is easily learnable compared to non-speech and it can give me accurate information about the exact position of the target door in the same way as a human-speech does." Furthermore, few participants mentioned that information provided by hybrid auditory feedback is closely similar to human-speech. However, hybrid audio feedback is more pleasant and can guide blind people to a destination in relatively less time compared with speech-only. In addition, participants also mentioned that after they become familiar with the environment they create a mental map of the hallway and while relying totally on a non-speech auditory feedback they have the idea of the hallway, although extracting meaningful information about the nearby surroundings with only the use of non-speech audio feedback is quite difficult and confusing.

With respect to the questions asked related to the efficiency of the sound during information conveyance, the majority of participants expressed that hybrid and speech were efficient in terms of the information accuracy about the surrounding environment. They experienced some difficulties, however, due to the attention required in all types of sound i.e., speech, non-speech, and hybrid. The single most striking observation to emerge from the interview was "auditory feedback requires 100 % concentration, which makes it hard to talk to people on the way."

At the end of the study, participants were asked to rank the three auditory feedback modes according to their preferences. The least preferred the non-speech (i.e., spearcons-only) audio mode, which was not surprising because participants relied few times on non-speech for the task completion. Participants ultimately chose hybrid (i.e., speech plus spearcons) over speech-only, arguing that they would like to use it in the mobility assistance applications in the familiar environment.

6 Discussion

Following the analysis of all data, it is possible to answer research questions proposed at the beginning of this user study.

1. *How many times does listening to the same speech instruction become repeated/annoying for the visually impaired?* For this purpose, an experiment was carried out with the study participants to determine the number of times same speech instructions become repeated/ annoying. It was observed from the results and discussion with the participants that listening five times to the same speech instructions in limited time duration (e.g., listening to same speech instruction after every 6 s) makes it repetitive for the participants. Furthermore, it was also noted that identification of thresholds for repetitive speech instruction varies in visually impaired people and it may be depended on the numerous characteristics of the individual, e.g., age, cognitive ability, orientation and mobility training (O&M) and period of vision loss (e.g., blind from birth or blindness caused at a later stage in life).
2. *Does the hybrid auditory feedback convey meaningful information compared with non-speech sound?* The statistical test (i.e., one-tailed paired t test) shows that hybrid (speech plus spearcons) auditory feedback is statistically significant compared with non-speech (spearcons-only) auditory feedback in terms of meaningful information and conveyance, while navigating indoors. Furthermore, the interview results also suggested that hybrid auditory feedback conveys dynamic information more clearly than non-speech, while participants were on the move and passed through different rooms. Totally blind participants also mentioned that they were not comfortable in totally relying on non-speech for obtaining information regarding nearby surroundings.
3. *What factors may decide that the auditory feedback based on speech may be converted to non-speech (i.e., spearcons)?* A new type of audio feedback (i.e., hybrid sound) was introduced based on two measures: sound repetitiveness and users' frequency of the traveled route. The present study reveals that on the basis of these two parameters, participants are comfortable with the change. Future work will explore and identify new parameters, which may help in identifying improved ways in pleasant and less-irritative auditory feedback for the visually impaired while they are on the move.
4. *What type of auditory feedback (i.e., hybrid, speech-only) successfully guides the visually impaired in relatively less time to the target destination in a familiar indoor environment?* On the basis of the mean task completion rate, it is evident that hybrid auditory feedback guided users in relatively less time to reach the target destination in a familiar indoor environment compared with speech-only. In addition, participants also mentioned that though human-speech does not need any learning period, however, they still have to wait to finish the sentence for a complete understanding.

5. *Is a user experience with a hybrid auditory feedback more pleasant as compared with a repetitive speech-only or non-speech?* Overall user experience with hybrid auditory feedback was pleasant and less-irritative compared with repetitive human-speech and non-speech. Participants indicated hybrid auditory feedback preference over repetitive speech-only auditory feedback. Objectively, it was observed that once participants followed the same path many times and the path becomes familiar to them, they raised concerns with repeated human-speech and instead, and were more interested to listen to the hybrid sound.

Finally, it is interesting to note that all participants liked the idea of hybrid sound and were very satisfied with its usage in mobile device for mobility assistance in an indoor environment. They agreed that hybrid sound produces less irritation and is more enjoyable compared with repetitive speech-only and non-speech sound. Furthermore, they also rated overall performance of hybrid sound much higher compared with speech-only and non-speech sound. Participants mentioned; "…the experience with hybrid sound was enjoyable and very convenient…".

7 Conclusions and future work

Vision is probably the major communication channel. However, for people with visual impairment or in the absence of vision, hearing substitutes for the visual channel and becomes the major form of non-visual communication [1–4]. A number of mobility assistance systems have been developed for helping people with visual disabilities in both indoor and outdoor environments. Although many research works (e.g., [7, 16–18, 21, 26]) have provided visually impaired people with non-speech and speech sound as an auditory feedback in mobility assistance or pedestrian navigation, none of the earlier studies has focused on intelligently combining speech and non-speech in auditory feedback to the users in eyes-free environment.

This paper has introduced a new approach of using hybrid sound for feedback purposes in mobility assistant applications for the people with visual disabilities. Speech sound seems to be an obvious choice for auditory feedback in an unfamiliar environment. However, repeated use of speech instructions on the frequent route of the blind can produce irritation. On the other hand, solely relying on non-speech sound for the presentation of critical information to the user is not helpful and the user can miss important details. Thus, the presented approach suggests to combine both speech and non-speech (i.e., spearcons) based on the sound repetition and users' frequency of the travelled route, which resulted in a creation of hybrid sound. The performance of the hybrid auditory feedback has been evaluated with the involvement of twelve legally blind participants who were recruited for the study. Each participant carried twenty tasks on three different days in limited time duration. The results from the experiment, questionnaire, and posterior interviews held with the participants were highly encouraging. Overall, participants appreciated the idea of hybrid auditory feedback and considered it more efficient and less-irritative compared with repetitive speech instructions and convey more meaningful information compared with non-speech.

During the study, it was observed that identification of the threshold for repetitive speech instruction varied due to factors related to the individual, such as cognitive ability, O&M training and time of the vision loss, e.g., early blind (i.e., no visual experience) or late blind (i.e., have some visual experience). It can be argued that these are very important factors for the future research in the designing of an appropriate auditory feedback for people with visual disabilities. However, these limitations can be addressed while focusing on a user-centric design and giving visually impaired people more freedom for making their own choices. Research plans for the near future include research in auditory feedback for the visually impaired in mobility assistance in both indoor and outdoor environments. Work is already being carried out on Chinese-based spearcons for the Chinese blind population. Furthermore, since the hybrid auditory feedback is less-irritative and more enjoyable in user ears, it is envisioned that its utility may go beyond mobility assistant systems. Future work may include the study of designing hybrid auditory feedback in desktop environment for people with visual disabilities using their desktop activities.

Acknowledgments The authors would like to thank reviewers for their valuable comments/suggestions on the early versions of the paper. Our gratitude goes to the participants for their time and involvement in the study. This work was funded by the Natural Science Foundation of China (Nos. 60703040, 61332017), Science and Technology Department of Zhejiang Province (Nos. 2007C13019, 2011C13042).

References

1. Fallah, N.: AudioNav: a mixed reality navigation system for individuals who are visually impaired. ACM SIGACCESS Access. Comput. **96**, 24–27 (2010)
2. Ran, L., Helal, S., Moore, S.: Drishti: an integrated indoor/outdoor blind navigation system and service. In: Proceedings of the 2nd IEEE Annual Conference on Pervasive Computing and Communications, pp. 23–30 (2004)
3. Loomis, J.M., Marston, J.R., Golledge, R.G., Klatzky, R.L.: Personal guidance system for people with visual impairment: a comparison of spatial displays for route guidance. J. Vis. Impair. Blindness **99**(4), 219 (2005)

4. Sánchez, J., Oyarzún, C.: Mobile audio assistance in bus transportation for the blind. Int. J. Disabil. Hum. Dev. **10**(4), 365–371 (2011)
5. World Health Organization: Factsheet visual impairment and blindness (2009) Accessible at: http://www.who.int/medacentre/factsheets/fs282/en/
6. Lutz, R.J.: Prototyping and evaluation of landcons: auditory objects that support wayfinding for blind travelers. ACM SIGACCESS Access. Comput. **86**, 8–11 (2006)
7. Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., O'Modhrain, S.: GpsTunes: controlling navigation via audio feedback. In: Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 275–278 (2005)
8. Gaver, W.W.: Auditory icons: using sound in computer interfaces. Hum.-Comput. Interact. **2**(2), 167–177 (1986)
9. Blattner, M.M., Sumikawa, D.A., Greenberg, R.M.: Earcons and icons: their structure and common design principles. Hum.-Comput. Interact. **4**(1), 11–44 (1989)
10. Brewster, S.A., Wright, P.C., Edwards, A.D.N.: Experimentally derived guidelines for the creation of earcons. In: Adjunct Proceedings of People and Computers (Huddersfield). British Computer Society, pp. 155–159 (1995)
11. Walker, B.N., Nance, A., Lindsay, J.: Spearcons: Speech-based earcons improve navigation performance in auditory menus. In: Proceedings of the International Conference on Auditory Display, pp. 63–68 (2006)
12. Palladino, D.K., Walker, B.N.: Learning rates for auditory menus enhanced with spearcons versus earcons. In: Proceedings of the 13th International Conference on Auditory Display, pp. 274–279 (2007)
13. Dingler, T., Lindsay, J., Walker, B.N.: Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In: Proceedings of the 14th International Conference on Auditory Display, pp. 1–6 (2008)
14. Walker, B.N., Lindsay, J.: Navigation performance with a virtual auditory display: effects of beacon sound, capture radius, and practice. J. Hum. Factors Ergonomics Soc. **48**(2), 265–278 (2006)
15. Tran, T.V., Letowski, T., Abouchacra, K.S.: Evaluation of acoustic beacon characteristics for navigation tasks. Ergonomics **43**(6), 807–827 (2000)
16. Holland, S., Morse, D.R., Gedenryd, H.: AudioGPS: spatial audio navigation with a minimal attention interface. Pers. Ubiquit. Comput. **6**(4), 253–259 (2002)
17. Ghiani, G., Leporini, B., Paternò, F.: Supporting orientation for blind people using museum guides. In: Extended Abstracts on Human Factors in Computing Systems, pp. 3417–3422 (2008)
18. Kainulainen, A., Turunen, M., Hakulinen, J., Melto, A.: Soundmarks in spoken route guidance. In: Proceedings of the 13th International Conference of Auditory Display, pp. 107–111 (2007)
19. Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D., Holmes, G.: ONTRACK: dynamically adapting music playback to support navigation. Pers. Ubiquit. Comput. **12**(7), 513–525 (2008)
20. Hansen, K.F., Bresin. R.: Use of soundscapes for providing information about distance left in train journeys. In: Proceedings of the 9th Sound and Music Computing Conference, pp. 79–84 (2012)
21. Vazquez-Alvarez, Y., Oakley, I., Brewster, S.A.: Urban sound gardens: supporting overlapping audio landmarks in exploratory environments. In: Proceedings of Multimodal Location Based Techniques for Extreme Navigation Workshop, Pervasive (2010)
22. Stahl, C.: The roaring navigator: a group guide for the zoo with shared auditory landmark display. In: Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 383–386 (2007)
23. McGookin, D.K.: Understanding and improving the identification of concurrently presented earcons. University of Glasgow (2004)
24. Loomis, J.M., Klatzky, R.L., Golledge, R.G.: Navigating without vision: basic and applied research. Optom. Vis. Sci. **78**, 282–289 (2001)
25. Loomis, J.M., Golledge, R.G., Klatzky, R.L.: GPS-based navigation systems for the blind. In: Fundamentals of Wearable Computers and Augmented Reality, pp. 429–446 (2001)
26. Strothotte, T., Fritz, S., Michel, R., Raab, A., Petrie, H., Johnson, V., Reichert, L., Schalt, A.: Development of dialogue systems for a mobility aid for blind people: initial design and usability testing. In: Proceedings of the 2nd Annual ACM Conference on Assistive Technologies, pp. 139–144 (1996)
27. Helal, A., Moore, S.E., Ramachandran, B.: Drishti: an integrated navigation system for visually impaired and disabled. In: Proceedings of the 5th International Symposium on Wearable Computers, pp. 149–156 (2001)
28. Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., Dellaert, F.: Swan: system for wearable audio navigation. In: Proceedings of the 11th International Symposium on Wearable Computers, pp. 91–98 (2007)
29. Hussain, I., Chen, L., Mirza, H. T., Majid, A., Chen, G.: Hybrid auditory feedback: A new method for mobility assistance of the visually impaired. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 255–256 (2012)
30. Chen, L., Hussain, I., Chen, R., Huang, W., Chen, G.: BlueView: a perception assistant system for the visually impaired. In: Adjunct Proceedings of the ACM Conference on Pervasive and Ubiquitous Computing, pp. 143–146 (2013)
31. Lazar, J., Feng, J.H., Hochheiser, H.: Research methods in human-computer interaction. Wiley, London (2010)