Text to Speech through Bluetooth for People with Special Needs Navigation

Eirini Barri  
Department of Computer Engineering and Informatics  
University of Patras  
Patras, Greece  
e-mail: ebarri@ceid.upatras.gr

Apostolos Gkamas  
University Ecclesiastical Academy of Vella  
Ioannina, Greece  
e-mail: gkamas@aeavellas.gr

Evangelos Michos  
Department of Computer Engineering and Informatics  
University of Patras  
Patras, Greece  
e-mail: emichos@ceid.upatras.gr

Christos Bouras  
Department of Computer Engineering and Informatics  
University of Patras  
Patras, Greece  
e-mail: bouras@cti.gr

Christina Koulouri  
Department of Computer Engineering and Informatics  
University of Patras  
Patras, Greece  
e-mail: christinakoul1995@hotmail.gr

Spyridon Aniceto Katsampiris Salgado  
Department of Computer Engineering and Informatics  
University of Patras  
Patras, Greece  
e-mail: ksalgado@ceid.upatras.gr

Abstract—As far as outdoor navigation is considered, the Global Positioning System (GPS) technology is still one of the most (if not the most) commonly used approaches. Even though it is still considered an ideal solution for navigating in outdoor areas, challenges and problems arise when GPS is considered for navigation inside buildings due to the fact that GPS signals cannot penetrate walls/ceilings (e.g., shopping malls, hospitals, etc.) and because signals can be absorbed by the building walls. This paper’s contribution is navigation system that assists people with special needs with an audio guidance system that incorporates input from a voice recognition system. The central part of the system is a device that is able to identify the position and orientation of the person that carries it and provides the ability to navigate and route by voice commands. The suggested voice synthesis system is used, so as to guide the user through obstacles in indoor locations. The information of the precise location and orientation of the device is made available to the whole system, through the building’s network infrastructure, so that the user’s mobile phone, being connected to the same network and also to the user’s headset through BLE, is able to send audio commands. For the voice commands, Google Cloud Text-To-Speech (TTS) will be used, assuming that an online connection is active on the user’s device.

Keywords—GuideMe; GPS; BLE; TTS; indoor; navigation; trilateration; pathfinding; audio; voice.

I. INTRODUCTION

Undoubtedly, there is an increasing demand for efficient indoor navigation systems, demand that mainly derives from smart cities, robots and visually impaired people, only to name a few. As far as outdoor navigation and pathfinding are concerned, the Global Positioning System (GPS) is still considered among the most commonly used technologies. Yet, this is only efficiently applicable in outdoor locations, because when indoor navigation comes into play, issues do arise. Of course, indoor navigation is very important to us and has many applications for humans and robots. Two of the most common issues that arise are a) when facing physical obstacles inside buildings that cannot be labeled as obstacles by the GPS and b) the fact that signals cannot be absorbed by walls inside buildings. Multiple floors, rooms and obstacles inside each and every indoor area we can think of pose a major problem. Additionally, the inability to use the GPS technology inside buildings makes indoor navigation more complicated, for reasons already explained above [1].

Many recent studies have been and are still conducted in order to make indoor navigation more effective and efficient. The direct need for new applications and technologies that can efficiently tackle such issues can luckily be covered by other available indoor navigation technologies that do exist nowadays, such as Wireless-Fidelity (Wi-Fi), Bluetooth and sensors.

Some of the most promising technologies for indoor positioning are Bluetooth Low-Energy (BLE) beacons [2] and Ultra-Wide Band (UWB) beacons [3]. More specifically, such beacons have a low deployment cost and are suitable for a wide range of mobile devices. BLE and UWB beacons have undoubtedly great potential because:

- After installation of such beacons, the only equipment needed is a smartphone that incorporates BLE support, whereas foot-mounted positioning needs special inertial sensors.
- The BLE approach supports a large variety of mobile devices (both Android and IOS devices).
• The installation cost of UWB and BLE beacons is relatively low. After deployment, beacons continue to operate for a long duration using their batteries, due to the very low energy consumption rates.

This paper provides the design and development of a navigation system that assists people with special needs using an audio guidance system that incorporates input from a voice recognition system. At its core, the system consists of a device that provides the ability to navigate and route by voice commands, based on the device's location and orientation capabilities. The instructions are based on the device's location and orientation capabilities and this device shall be connected to the server via the user's mobile phone (Android). The suggested voice synthesis system is used to guide the user through obstacles in indoor locations. Wireless connection between the user's mobile phone and the mobile device are made available through low energy consumption BLE and UWB protocols. The overall system (when completed) will consist of the following components:

• Equipment permanently installed in selected areas.
• A cloud server that will synchronize and coordinate the various parts, store information about the facilities and users, and will be responsible for the accounting and invoicing parts.
• Portable devices.
• Software that will run on smartphones.

As far as the voice commands are concerned, Google Cloud Text-To-Speech (TTS) will be used that supports different programming languages through its Application Programming Interface (API) (C#, GO, JAVA, NODE.JS, HYPERTEXT PREPROCESSOR (PHP), PYTHON, RUBY) [4]. The API uses online resources (thus, an active network connection is mandatory) and is provided as a service, or free, or comes with a small cost. The commands for the TTS conversion are provided through Speech Synthesis Markup Language (SSML) language [5]. The GuideMe device will give commands through UWB beacons to the Android application of GuideMe and the application – using the Google Cloud TTS - shall provide the audio commands.

The rest of this paper is organized as follows. Section II describes the motivation behind our work. Section III provides a literature review of other current works on this subject. Section IV addresses the system’s architecture whereas Section V goes into finer details in regard to the proposed algorithm for TTS through Bluetooth navigation in indoor spaces. Finally, Section VI summarizes our main findings and conclusions and suggests probable future work. The acknowledgement and conclusions close the article.

II. MOTIVATION

Blindness is the condition of lacking visual perception due to physiological or neurological factors. People with blindness encounter many problems in everyday life. Blind people always depend on others. They can not move easily from one place to another without help from others.

According to World Health Association [31] the following are the key facts regarding blindness and vision impairment

• Globally, at least 2.2 billion people have a vision impairment or blindness, of whom at least 1 billion have a vision impairment that could have been prevented or has yet to be addressed.
• This 1 billion people includes those with moderate or severe distance vision impairment or blindness due to unaddressed refractive error, as well as near vision impairment caused by unaddressed presbyopia.
• Globally, the leading causes of vision impairment are uncorrected refractive errors and cataracts.
• The majority of people with vision impairment are over the age of 50 years.

The GuideMe project [30] will develop a platform that provides guidance and security for out-of-home travel. The solution is built around a discreet portable device capable of indoors localization with accuracy of 10 cm using UWB technology. The device can also determine the orientation of the user, receive voice commands, and transmit voice instructions. It is also capable of detecting crashes and sending updates. The feature is based on the collaboration of a mobile device, a service running on the user's mobile phone and a server.

The motivation of the paper is to improve two areas in the life of the blind people and people with special needs in general: convenience and security. Specifically, with the use of the proposed system, users will feel more comfortable visiting public places such as airports, shopping malls, stations, etc., as they will be guided by the system to reach their destination. At the same time, in case of emergencies involving both the user (accidents) and the building (fire, earthquake etc.), the system will inform the users of the exact location of the users, whilst also guiding them to the nearest exit. The ultimate goal is to increase the presence of the population with mobility or other problems in buildings by 20%.

III. RELATED WORK

In this section, we present previous research work on indoor navigation systems targeted for people with special needs and provide a summarized overview of the research conducted in this field. Our literature review is categorized in research work on indoor positioning and indoor navigation.

Following the previous studies, in this section, we will present similar projects to GuideMe. Indoors and San Francisco International Airport worked together to create an app for visually impaired passengers. The Entrepreneurship-in-Residence (EIR) project is an Edwin M. Lee collaboration with the White House and other San Francisco business partners. At the beginning of 2014, they chose to help the San Francisco Airport (SFO) create a tool to assist blind and visually impaired travelers [11].

Recommendation ITU-T F.921 [12] determines the way a navigation system that is audio-based can be developed to guarantee that it is dedicated and responsive to the needs of
people with visual disabilities. It takes on a technologically neutral approach and sets the operating attributes of the system. The goal is to provide visual network system designers with the audio data they require in the early stages of development to prevent any problems that keep vision impaired users from making full and independent use of the built environment. This system explains how to adapt the user experience to audio-based network navigation systems and to secure the interoperability of these technologies. In addition, it acknowledges that, to meet the needs of people that are visually impaired, networked audio navigation systems can also benefit people with other age-related disabilities, as well as the general public.

The purpose of [13] is to implement a module-based application developed in the context of preliminary projects for the mobile mass market, through an appropriate user interface that responds to the needs of the visually impaired. The blind user should be able to use public transport independently in a secure manner and navigate complex public transport terminals. As a result, the system combines real-time communication to and from public transport vehicles with precise positioning and guidance while it also provides additional navigation assistance.

Same as [13], INK 2016, Indoor Navigation and Communication in ÖPNV for blind and visually impaired people [14], combines real-time communication to and from public transport vehicles with precise positioning and guidance and has additional video call navigation assistance where the person can communicate with a professional operator.

Project Ways4all [15] is a new personalized indoor navigation system that can increase public transport accessibility for all passengers and especially the visually impaired, who will be able to access public transport and the necessary up-to-date traffic information in a very simplified way.

Finally, project "Using an Integrated Technique for Developing Indoor Navigation Systems to Allow the Blind and Visually Impaired People to Reach Precise Objects" [16], uses a set of different technologies (WiFi, Bluetooth, and Radio-frequency Identification (RFID)) to help the user reach a micro element in the navigated environment. As a proof of concept, Yarmouk University will test the system framework on their library to help the blind user find a specific book. Therefore, it constitutes an intelligent interface for precise indoor navigation for blind and visually impaired people using a smart phone.

A. Text-To-Speech (TTS)

In this section, we present previous research work on TTS techniques targeted for people with special needs and provide a summarized overview of the research conducted in this field. Our literature review is categorized in research work on TTS and Bluetooth TTS for blind people.

There are several studies concerning TTS. Speech synthesis is the artificial production of human speech. Attempts to control the quality of voice of synthesized speech, several prototypes and fully operating systems have been built based on different synthesis techniques. The authors of [17] review recent advances in research and development of speech synthesis, so as to provide a technological perspective. Their approach is based on the Hidden Markov Model (HMM) and aims to summarize and compare the characteristics of various speech synthesis techniques used by presenting their advantages and disadvantages.

The purpose of [18] is to try to explain the aim of a TTS synthesis system and how it works. In more detail, the authors try to give a short and inclusive review of developing a system that is equal to human performance. The rule-based techniques (formant and articulatory synthesis) are described by the authors and, finally, they propose an HMM synthesis combined with a Harmonic plus Noise Model (HNM), so as to get a TTS synthesis system that requires lower development time and cost.

As mentioned before, the commands for the TTS conversion are provided through SSML language [5]. SSML is a component of a bigger set of markup specifications for voice browsers developed through the open processes of the W3C. It is scheduled to provide a rich, XML-based markup language in order to assist the generation of synthetic speech in Web and other applications. A TTS system (a synthesis processor) that supports SSML will be responsible for providing a document as spoken output. It will also be responsible for using the details contained in the markup to provide the document, as intended by the author. According to [18], a significant job of the markup language is to provide authors of synthesizable content a standard way to control some characteristics of speech such as pronunciation, volume, pitch, rate, etc. across different synthesis-capable platforms.

Special reference needs to be made on the API for TTS services. The GuideMe device will give commands through UWB beacons to the Android application of GuideMe and the application – using the Google Cloud TTS -shall provide the audio commands. Particularly, [20] refers to a Cloud TTS conversion powered by machine learning. Google Cloud TTS transmits text into human-like speech. This is accomplished in more than 180 voices across 30 variants and languages, or more. It applies groundbreaking research in speech synthesis (WaveNet) and Google's powerful neural networks to deliver high-fidelity audio. With this easy-to-use API, anyone can make live interactions with users that constitute customer service, device interaction, and other applications.

Another API for TTS service is [21]. This is part of the Speech service of Microsoft and builds apps and services that speak naturally. This API creates lifelike voices with the Neural Text to Speech capability built on breakthrough research in speech synthesis technology. It offers a wide range of voices and languages. One specific characteristic of this API is that it provides its users with models for customization in order to create a unique voice for everyone’s solution and brand.

Similar to [21], another part of the Speech service of Microsoft is [22]. Respectively, it allows to convert text into synthesized speech and get a list of supported voices for a region using a set of REST APIs. Each available endpoint is
associated with a region. Also, a subscription key for the endpoint / region someone plans to use is required. The differentiation from [21] is that TTS REST API supports neural and standard TTS voices, each of which supports a specific language and dialect, identified by locale.

Watson is IBM’s suite of enterprise-ready AI services, applications, and tooling. Watson TTS [23] converts written text into natural-sounding audio in a variety of languages and voices. Watson TTS develops interactive toys for children, automates call center interactions, communicates directions hands-free, and beyond. It delivers a seamless voice interaction that caters to the audience with control over every word. Like [21], it offers this pronunciation across many languages and voices, which is very important for the users.

It is important to mention that every API for TTS service described above includes documentation, in order to facilitate usage and implementation by the users.

B. Bluetooth TTS for blind people

In [24], the PERCEPT system is introduced. PERCEPT is an indoor navigation system for the people who have visual issues and cannot see. PERCEPT is intended to make improvements to the quality of life and health of the visually impaired community and try to give the ability for independent living. Using PERCEPT, blind users are supposed to access public health facilities such as clinics, hospitals, and wellness centers, without the help of other people, but only using their own means. PERCEPT system tests were held with the participation of 24 blind and visually impaired users in a multilevel building. The results show that PERCEPT system is effective in providing the right navigation instructions to these users. The key aspects and advantages of this system are that it is inexpensive and that its design follows orientation and mobility principles, according to the authors.

The creators of the PERCEPT system have also created the INSIGHT system [25], which is an indoor location tracking and navigation system for the blind people using RFID (Radio Frequency Identification) and Bluetooth connectivity technologies. The workflow is as follows. The PDA based user device interacts with the INSIGHT server and provides the user navigation instructions through voice commands. They have implemented accurate navigation as well they have integrated a PANIC button in case of emergency. Moreover, the system is able to understand if the user is in wrong place and heads towards false direction and helps them to re-route in order to move in the correct direction.

In [26], a system with a portable TTS converter is designed in order to assist the blind listen to an audio of a text that has been scanned. The system, according to the authors, consists of a page scanner that can be carried with one hand, an Android phone that is able to scan the image and send it over Bluetooth, and an app that helps with the extraction of the text from the scanned image and to convert the extracted text to speech. Moreover, another positive impact of this system is that it comes with a page scanner which scans the entire page containing the text. So, blind users do not need to take photos and focus on the area of text that is needed, and then crop it in order to remove the background pictures etc, something that happens in the case of other systems that exist.

Furthermore, [27] tries to design and create a solution for visually impaired travelers and specifically, for the use case of train transportation using only a smartphone and no other hardware. Particularly, using the BLE and the integrated compass of the smartphone, the system is able to provide turn by turn voice commands inside the Tokyo station.

In [28], the authors designed a system for visually impaired people in order to help them navigate through work zones in a safer manner. According to statistics taken from the Federal Highway Administration (FHWA), every year about 17% of all work zone fatal accidents happen to pedestrians [28]. People who have problems seeing often must deal with physical and information difficulties that limit their accessibility and mobility. After a survey conducted, some elements of the results were implemented in a smartphone application that incorporates both GPS and Bluetooth technologies to calculate the user’s location. When the user goes to a work zone, the smartphone is supposed to vibrate, so as to alert users, and the application will then announce an appropriate audible message to users. Blind users, if they want, they could do a single tap on the smartphone to repeat the audio messages.

IV. System Architecture

In this project, the main components are a small wearable that helps in the user’s positioning through Ultra-Wide Band (UWB) technology. This technology provides very accurate positioning, up to 10 cm divergence. The system, apart from the ability to locate the user, has the ability to provide guidance via voice commands.

Figure 1. Overview of the proposed architecture.

In our purposed system, our smart device can communicate to anchors via UWB technology, in order to locate the user. The anchors are calculating and measuring the distance between the user and the anchor. The distance data (between the user and the anchors) is transferred to a local server so as to measure the exact position, running positioning algorithms. The local server, based on the positioning and navigation algorithms, will give commands (using the Wi-Fi network) to the Android application of GuideMe running on the user’s Android smart phone and the
application – using the Google Cloud TTS - shall provide the audio commands. The API uses online resources and is provided as a service. The commands for the TTS conversion are provided through SSML language. Furthermore, there is a remote server that has a map of the building. This remote server, having the details of the building, the position of the user and the destination of the user, can provide guidance to the user, giving him directions. The directions are given by the smartphone to the user through wireless headphones, using voice commands. The system will also support the usage of pre-recorded voice command for the case where no Internet access is available.

As far as the wearable device is concerned, the processor that is chosen is the module made by Esspresif (Esspresif ESP32 [33]). This family of processors are energy efficient, in order to expand the battery life. A Wi Fi module is integrated as well in the system. For the connectivity through UWB, we have chosen the module DWM1000 of Decawave [32].

Figure 2 presents the general architecture of the device. The device consists of the magnetometer and accelerometer sensors, the UWB module, the Main Computing Unit and a module for the battery management as well.

V. PROPOSED SYSTEM

In this section, we describe the proposed system for TTS through Bluetooth navigation in indoor spaces. We propose a GuideMe device that will use UWB beacons to let the system identify its position and orientation precisely and provide this information to the Android application of GuideMe and the application – using the Google Cloud TTS-shall provide the audio commands. The commands for the TTS conversion are provided through SSML language. SSML is a component of a bigger set of markup specifications for voice browsers developed through the open processes of the W3C. It is scheduled to provide a rich, XML-based markup language in order to assist the generation of synthetic speech in Web and other applications. A TTS system (a synthesis processor) that supports SSML will be responsible for providing a document as spoken output. It will also be responsible for using the details contained in the markup to provide the document, as intended by the author. According to [18], a significant job of the markup language is to provide authors of synthesizable content a standard way to control some characteristics of speech such as pronunciation, volume, pitch, rate, etc. across different synthesis-capable platforms.

Special reference needs to be made on the API for TTS services. TTS is considered ideal for any application that plays an audio of human speech to users [29]. TTS operates by converting SSML input into audio data and by using TTS. The response string can be converted to actual human speech that will be played back to the user of the application. As for the process, the procedure of translating text input into audio data is called synthesis and the output is labeled as synthetic speech. The speech synthesis begins by generating raw audio data as a base64-encoded string and decoding of this string into an audio file is required in order to play from the application. Additionally, TTS offers a large variety of custom voices, depending on the needs (voices differ by language, gender and accent). The output settings are also configurable, concerning speaking rate, pitch, volume and sample rate hertz.

An indicative mode of the operation, followed by the system we described, is shown in Figure 3.

VI. CONCLUSION AND FUTURE WORK

This work refers to the project GuideMe using Bluetooth technology for text-to-speech directions given to people with visual difficulties. Some previous research works on indoor navigation systems using text-to-speech technology were presented. This work is the basis of the next step of the
project that relates to developing the software for smartphone or a wearable device using an audio guidance system that incorporates input from a voice recognition system. The text-to-speech technology through Bluetooth is used to guide the user through obstacles in indoor locations. For future work, we may include an extension of this current research by also covering outdoor areas through the application.

ACKNOWLEDGMENT
This research has been co-financed by the European Union and Greek national funds through the Regional Operation Program “Western Greece 2014-2020,” under the Call “Regional research and innovation strategies for smart specialization (RIS3) in Communication and Information Technologies” (project code: 5038620 entitled “System for indoors orientation and guidance - GuideMe”).

REFERENCES
[19] https://www.w3.org/TR/speech-synthesis11/ 2020.04.06
[29] https://cloud.google.com/text-to-speech/docs/basics 2020.04.06
[31] https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment 2020.04.06