Tell It Your Way: Technology-Mediated Human-Human Multimodal Communication

Helena Cardoso(✉), Nuno Almeida, and Samuel Silva

IEETA – Institute of Electronics and Informatics Engineering of Aveiro,
DETI – Department of Electronics, Telecomunications and Informatics,
University of Aveiro, Aveiro, Portugal
{helenamncardoso,nunoalmeida,sss}@ua.pt

Abstract. Communication plays a pivotal role in our daily lives. With the advances of technology we are now able to use it to communicate with others at a distance. However, while in direct human-human communication we are able to adjust how we pass a message based on our context and the perceived context of the receiver. When we do it at a distance, using a messaging tool, one of the most popular choices, nowadays, this becomes harder. In fact, most messaging tools, such as Whatsapp or Messenger, provide some degree of flexibility regarding the way a message is sent (e.g., text, audio, image), but the receiver is limited to receiving it in the format of sender choice. In this regard, providing more flexibility in such technology-mediated communication scenarios might foster increased adaptability of these tools to multiple user abilities and contexts, and provide important alternatives for those with some disability (e.g., aphasia, blindness). The work presented here adopts a user-centered approach to design and develop a first proof-of-concept for a multimodal messaging system than enables message modality conversion regardless of the format used by the sender.

Keywords: Multimodal communicator · Technology-mediated communication · Multimodal messaging

1 Introduction

Communication plays a crucial role in the individual's social, personal and professional development, but as it is so common, sometimes we do not give it due importance. Communication is a dynamic interactive process that involves sharing information, ideas, thoughts, feelings, and values and, in our daily lives, we use a wide variety of ways to communicate with others. We say communication is successful when the person transmits the information clearly and accurately and when the receiver interprets and understands the information correctly. However,

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2022
Published by Springer Nature Switzerland AG 2022. All Rights Reserved
X. Gao et al. (Eds.): MobiHealth 2021, LNICST 440, pp. 343–357, 2022.
https://doi.org/10.1007/978-3-031-06368-8_23

communication can face several obstacles to the efficient and reliable transmission of information, in many everyday contexts (e.g., a noisy environment) and, particularly, for those with speech disabilities.

The communication between humans is often multimodal since we interact with someone taking advantage of different ways that we can use to transmit our information to the recipient (e.g., speech, gestures) and, sometimes, we even use several of them at the same time. The way we communicate is often adapted to the nature of the message we want to transmit, but also to our current context, e.g., where we are and what we are doing. Additionally, we also adapt our way of communicating to what we know about the context and abilities of the receiver, e.g., if we know the receiver cannot hear us, we try, e.g., to resort to some visual form of communication (e.g., gesture, written).

Technology is increasingly present in our daily life. Through this, significant achievements were made that revolutionized the way of living in community, and thanks to its evolution, we can spend almost all day connected to everything and everyone. While video is an ever increasing possibility, most communication resorts to messaging due to its flexibility, low bandwidth requirements, and possibility of asynchronous message exchange, i.e., the interlocutors do not need to be present simultaneously. The best and most used messaging apps[1], for Android and iOS, are WhatsApp, Facebook Messenger, WeChat, and Telegram, among others. These apps provide sophisticated services, offering all kinds of interactions between users, from texting to voice and video calling to sharing images and audio. However, when technologies mediate communication, sometimes sending the message in a particular modality (e.g., audio, text, or pictograms) may not be appropriate (or understandable) for the receiver. Often, due to contextual conditions (being in a noisy place and not being able to hear the audio or driving and not being able to view the text message) or even health problems (e.g., hearing or visual impairment), this can lead to some communication difficulties. Therefore, having a messaging system that could adapt both to the sender and receiver context, preferences and abilities, enabling asymmetric use of message modalities, i.e., not limiting the message to the modality it was sent on, would improve the range of suitability of these tools.

And while technology-mediated communication supporting a versatile articulation of different forms of sending a message can be useful for all, it can provide an important support to those who have their abilities to communicate hindered due to a persistent or temporary condition. Most of the systems found on literature are limited in the number of supported modalities, typically supporting two or three conversions between modalities. Representative examples of messaging system capable of converting from speech-to-text and text-to-speech are the Stimme [2], BridgeApp [13] and ASRAR [9]. The first can also convert text to tactile feedback, the second, text to sign language and the third from text to gestures. AbleChat [5] enable the conversion from text to pictograms. These applications already provide valuable support. Nevertheless, these tools often

[1] https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/.

target a small number of conversion possibilities (e.g., from pictograms to text), are tailored for a specific group of users, and can sometimes work as barriers for integration, since they are tools that are different from what everyone else around uses (or is willing to use).

One notable example of a condition hindering communication is aphasia, which often occurs after a stroke or head injury, and can affect the person's ability to speak, read, write or remember the names of objects. For these users, having multiple alternatives to send and receive messages can foster a greater adaptivity of the communication tool to their specific (often idiosyncratic) needs, which depend on the type of aphasia they have.

In this context, our research goal it to explore the interchangeable use of multiple messaging modalities in communication mediated by technology towards increased adaptability to different contexts, preferences and user abilities, adopting a vision of a communication tool designed for all. In this regard, relying on a user-centered approach, and aiming at a potential future application scenario assisting aphasics, in the scope of project APH-ALARM[2], the work presented here, establishing a ground zero for this research, is a first proof-of-concept for a communication tool supporting the exchange of messages using different modalities (e.g., text, images, audio) and introducing the ability of converting among them based on user choice or according to the user's declared abilities.

The remainder of this document is organized as follows: Sect. 2 describes the methods adopted to perform a first characterization of potential users and identify their expectations and receptivity to a multimodal communicator; Sect. 3 presents the overall steps and outcomes regarding the design and evaluation of a high-fidelity mockup of the system's interface to validate the paradigm and overall features; Sect. 4 presents the overall architecture and features for the first iteration of the multimodal communicator system; finally, Sect. 5 presents conclusions and a set of ideas for future developments.

2 Users, Scenarios and Requirements

The first stage of the work was to get further understanding of the potential target audience and of the characteristics the platform should have to be perceived as useful. Some personas and scenarios of use of the system were created, this allowed to define a list of requirements that the system must meet in the long-term. Following this way a user centered design methodology [4].

Before starting the new platform's development, a survey was performed to obtain some information about messaging use, and preferences in that context. The survey was answered by 69 persons, the main goals were to know their preferences regarding the communication systems, namely: (1) what system they usually use; (2) what modalities they know; (3) whether they are in favor of certain features; (4) if they would eventually join the new multimodal communication platform; and (5) if it solves many problems in a technology-mediated

[2] https://aph-alarm-project.com/.

communication. The analyses of the questionnaire allowed to conclude that many users frequently use communication platforms. The most used modality is the text modality, mostly because it is the one users feel most comfortable, the only one that they know how to use, or, in some cases, the only modality available in the used platform. Some people answered they would alternate the modality depending on what they want to communicate and the context around them. After explaining what was the goal of conversions between modalities, it was asked which ones they would find interesting to exist in this type of system. The most voted were the text-to-audio and audio-to-text, the least voted was pictograms-to-audio.

2.1 Personas

Personas are fictional representations of a person, since they are based on the behaviors of real people. They are created based on research to represent the different user types that might use the product. Personas provide helpful information to determine what a product should do and how it should behave [4]. With the results from the survey, 5 Personas were created trying to cover most of the questions addressed, such as context, disabilities and user preferences. Given their extent, only one Persona is presented here as an illustrative example:

Image adapted from Wikimedia

Madalena Vagos is 30 years old, lives in Aveiro alone and was born into a large and very communicative family. Sometimes she wants to share some message with her family, but given the limitation of certain family members it would consume some time. Her uncle John lost his sight in an accident, her mother suffer from headaches, her father is always on the road, and some younger members of the family, who have access to electronics, still cannot read. Personalize a message to each member takes much time, as she works several hours as a nutritionist, Madalena wanted this function to be more optimized and not only for communication with her family members but also to communicate with all clients.

Motivation: Madalena would like to be able to send the same message to everyone, but is such a way that they could receive it in the most suited formats.

2.2 Scenarios

After defining the Personas, several scenarios were designed. These consist of small scenes illustrating how the users' motivations can be served by a new system depicting the contexts and ways of use that users will face [4]. Although several scenarios have been proposed to guide the development of the application, given their extent, only one is presented, here. As can be observed, the scenario that follows depicts Madalena (the Persona presented above) using a system addressing her motivation for a particular context.

Madalena sends Christmas wishes to the family group—Another Christmas has arrived and Madalena wants to send a Merry Christmas message to her family. She created a group with all family members in the app, wrote: "Merry Christmas to all and good entries" and send it to the family. Her uncle, who has visual problems, listens to an audio message. The message is automatically converted from the text because he has set the disability in his profile. Her younger cousin, who has autism spectrum disorder and prefers to view messages in image format rather than text, has the option image as the default modality to view the messages, so he always sees messages in image format. Her mother, due to headaches sometimes cannot read text messages or listen the messages, when is the case, she chooses to convert the message to image. Finally, while his father is driving the message is converted to audio and read aloud.

Madalena no longer needs to send a message individually in the most appropriate format for each family member since the system allows this to happen transparently.

2.3 Requirements

Requirements describe the necessary capabilities of the product. Based on the scenarios devised for the different Personas, we extracted requirements from the actions and features depicted in them [4]. The requirements were divided into two sets: functional and interaction and a summary of those deemed most relevant is presented in Table 1. These served as grounds for the design of the first prototypes.

Table 1. Overall requirements for the multimodal communication system as extracted from the scenarios identified for the different Personas.

Functional Requirements

- Allow the user to select the modality which they want to send/receive the message.
- Support sending message in text, audio, image and gesture format.
- Convert from one modality to another according to the users' preferences.
- Convert from text to audio, text to image, audio to text, image to text, gesture to text format.
- Showing the message to the user in text, audio and image format.
- Adapt the modality depending on the context.
- Adapt the modality depending on the users' disabilities.
- Store users' preferences.

Interaction Requirements

Modality	Technology	Interaction
Text	Touch Screen	The application must allow users to navigate using touch inputs or mouse,
	Keyboard	depends on the device
	Mouse	The application should allow users to write his message;
Speech	Microphone	The application should allow users to record his voice;
	Speakers	The application should allow user to hear the message;
Image	Camera	The application should allow to take a picture;
	Gallery	The application should allow to choose a picture from gallery;
Gesture	Camera	The application should allow to record video, to recognize the gestures;

3 High Fidelity Mockup Design and Evaluation

Following the adopted iterative user-centered design methodology, our aim was to perform short prototyping sprints followed by evaluation to inform further

developments and refinements [3]. After setting the requirements for the system, those considered with a higher priority were selected for the initial prototype. At this early stage, we opted for building high-fidelity mockups to perform first validations of the design, flow, and features of the system. There are many design tools for making prototypes (e.g. Adobe XD, Proto.io). For our work, we chose the graphic editor Figma. One of the significant advantages of this editor is that it has no limitation on sharing the prototype, and someone else may be working on the same prototype. The developed mockups are high-fidelity, meaning that they are already more advanced than common paper prototypes (low-fidelity), are more aesthetically pleasing, and already support interactions (systems' flow). Thus, users have a better perception of the application, both aesthetically and functionally. This section presents the developed high-fidelity mockups, their evaluation (e.g., Heuristic Evaluation, Usability Tests) and the result of the evaluation.

3.1 First Mockup

One of the results from the initial survey taken by 69 people was that 61 persons used Messenger as a means of communication. Since many people are already acquainted with Messenger, it would be easy to handle in an application with a similar flow and interface. Therefore, the developed mockups were roughly based on the flow and element disposition adopted by Messenger.

On first use, the user has to login into the application (Fig. 1(A)). The next page shown is the Chats Page, which has a list of all the user chats (Fig. 1(B)). At the bottom of each page has a menu (Bottom Navigation), with three options: chats, friends and user profile. This menu will be present on all pages for quick access to the main pages. The friends' page (illustrated in Fig. 1(C)), displays a list of all the user's friends. The friends' profile (Fig. 1(D)) contains the usual user name, gender, avatar, birthday, but also includes information regarding the disabilities (hearing, speech or vision problems). In this page it is possible to add/remove the person to/from your friends' list and if it is already a friend, we can start a conversation. To add a new friend, the user needs to press the new button on the Friends Page and it is possible to see a list of all users registered on the application. Furthermore, on the user's profile page it is possible to see data similar to what was in the friends' profile. Adding information on disabilities to the user profile is beneficial because the system will convert the received messages to the most suitable format for their situation. For example, if the user has hearing problems, all the audio messages received will be converted to either image or text, without the user having to change each message to the desired modality.

In a private chat it is possible to send image, audio, and text messages, just like for any other chat applications that exist, these days. If the user wants to convert a message to other modality, he can click on the desired modality to convert. In this prototype, the considered conversions include text to audio or image, audio to text or image, and image to text or audio. Figure 2 illustrates examples of conversions.

Heuristic Evaluation The first step to evaluate the prototype was to conduct a heuristic evaluation. At this point, the goal was to identify major usability problems in the interface. To this end, four evaluators performed a heuristic evaluation adopting Nielsen's heuristics [10]. The evaluators were two females and two males, students of Computer and Telematics Engineering, aged 22–25 years old and with experience in performing heuristic evaluations. Each evaluator marked usability issues based on the adopted heuristics and according to a severity scale ranging from 0 to 4, in which zero consisted of a low-impact usability problem and four consisted of a high-impact usability problem. Taking into account the identity usability problems that resulted from this evaluation, the most concerning ones, with severity 4, were related to the lack of error messages and provided feedback in cases of adding or removing friends and signing out. In terms of what is related to conversions, they stated that the button to go back to the original modality was not very intuitive, and the result of the image to text conversion was not very coherent either. They also identified the lack of a landing page explaining what the application is about and introducing the user to its features, but this issue was scored as having low impact.

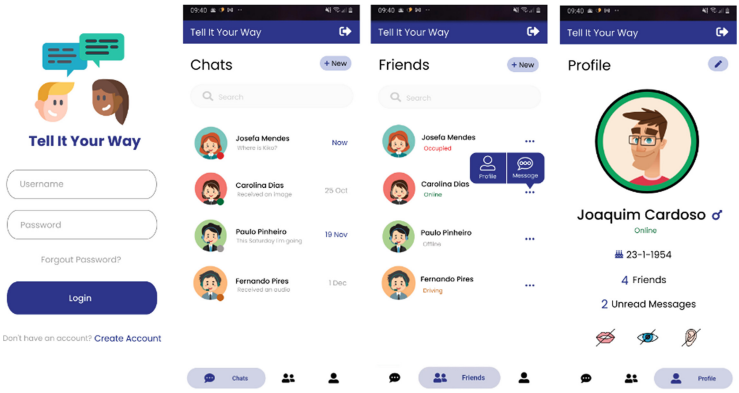

Fig. 1. Screen from the first prototype of the main pages. (A) Login page; (B) Chats page; (C) Friends page; (D) Profile page;

3.2 Refined High-Fidelity Mockup

In the second mockup several improvements were made based on the heuristic evaluation. The problem of not understanding the result of the conversion from image to text resulted from the fact that in the conversion many objects are identified, leading to the text sometimes becoming incomprehensible. To make the message a little bit more understandable, at this stage, without going into more sophisticated methods to generate a sentence, the message starts with "The image contains:".

The user's avatar in the conversation was downsized, leaving more space for the messages. It was added the "..." button in front of the user's data, through

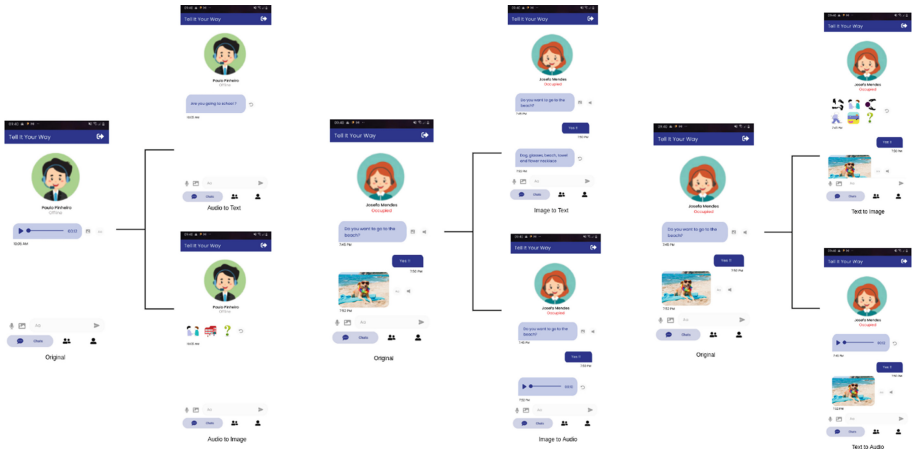

Fig. 2. Screen from the first prototype doing conversions from different modalities. (A) (1) Audio-to-Text and (2) Audio-to-Image; (B) (1) Image-to-text and (2) Image-to-audio; (C) (1) Text-to-image and (2) Text-to-audio (Color figure online)

which more information about the user can be accessed. This way, the user does not need to go to the friend's page to see the profile. In a converted message, the icon to return to the original modality – a rewind button –, has been replaced by all available conversions. When the original modality appears, the button has a different color (blue) compared to the other buttons (grey).

Confirmation messages have been added for the "remove friends" and "exit the system" features, and a dialog box was added with the information that a individual has been added to the friends' list, improving the feedback when adding a person. On the pages, which are not the main (chats, friends and profile), a back button was added to go to the previous page, because many devices today do not have this button. Finally, a tooltip has been added upon the status tag, when the user wants to know the meaning of a badge they just tap it, and the information will be displayed.

On the user profile page it was added the possibility to choose the default modality. When the user selects a default modality, all future messages will be converted to it. In the case of selecting some disability, predefined modalities will be blocked because it makes no sense for some difficulties to have specific predefined modalities (e.g., a person with a hearing impairment precludes audio as a predefined received message modality).

Usability Tests. The focus of the evaluation of the second (refined) high fidelity mockup was to understand how well users learned to use the system and how easily they were able to finish a set of tasks identified as important from the devised scenarios and requirements (see Sect. 2). The evaluation was conducted in a Concurrent Think Aloud (CTA) manner where the users were asked to narrate their thoughts as they went through the tasks. In Table 2, it is possible to see the tasks that users had to complete:

Table 2. Tasks performed by users during the usability evaluations of the second version of the high-fidelity mockup.

Tasks
1. Launch the application with personal settings;
2. Add to your profile an indication that you have hearing problems;
3. Add Liticia Dias to your friends' list and start a conversation with her
4. Remove Fernando Pires from your friends' list
5. Send an audio message to Carolina Dias
6. Send an image to Carolina Dias
7. Josefa Mendes sent you a text message, listen to it
8. Looking at the current screen, how do you know which original format the converted message was sent and go back to that original format
9. Josefa Mendes sent an image, convert it to text
10. Paulo Pinheiro sent you an audio message, see it in image format
11. See Paulo Pinheiro's profile
12. Modify the default modality to always receive messages in text format
13. Change your status to driving and check what the new default mode is
14. Logout

The participants consisted of 10 users (4 male and 6 female) of different ages, from 14 years old to 58 years old. An observation table was filled during the evaluation regarding: (1) if accomplish the task; (2) time take to accomplish the task; (3) critical and non-critical errors; (4) if felt lost; (5) if the user asked for help; and (6) perceived task difficulty. The last one was asked to the participants at the end of each task.

Participants successfully completed all the tasks, some of them with help. Results show that users encountered some obstacles/difficulties in the initial tasks and then learned how to use the application. The third task had a high perceived difficulty (3), which may be because of the complexity of the task. Some participants forgot the second part of the task, to facilitate it could be divided into two distinct tasks. Some participants gave the idea to provide an option to start a conversation from the friends' profile, right after adding a friend. In the current mockup users need to go to the friends/chats page to start a conversation. Some users had difficulties changing the profile, in task 2: even though they were able to add the hearing problem to the profile, they failed to save the change, at the end. Further analysis revealed that this was caused by the need to scroll down the page to see the save button. Task 4 was more complicated for older people, they did not know they had to go to the user's profile to remove the friend, since they were not used to social networks. Some participants had difficulty converting the first message, but after converting it, they understood how it worked. Other difficulties that were encountered by some participants, especially among the older ones, included the fact that the prototype was in

English – the users were native Portuguese – and they found the icons not very intuitive to understand. Overall, all feedback considered, we conclude that the inclusion of a tutorial or introduction may be beneficial, providing some support for the application's first use.

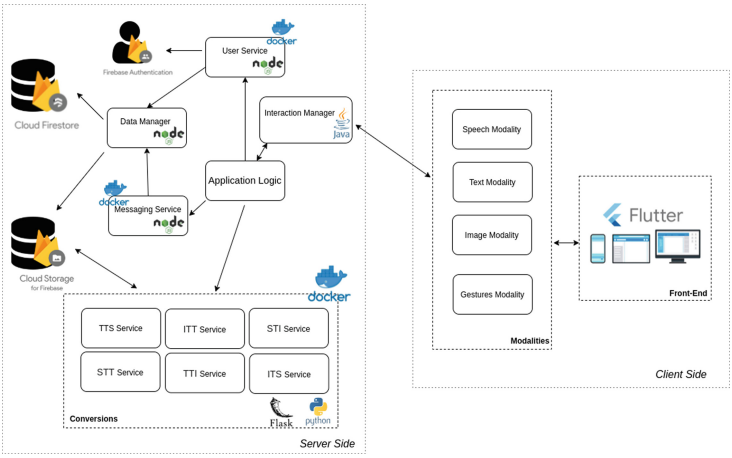

Fig. 3. Overall architecture for the developed multimodal communicator depicting the main modules and considered technologies.

System Usability Scale Questionnaire. After all tasks were completed, participants were asked to fill a System Usability Scale (SUS) questionnaire. The questionnaire consists of 10 questions with a scale ranging from 1 to 5 where 1 is strongly agree and 5 is strongly disagree. One of the benefits from using a SUS questionnaire is because this can effectively differentiate between usable and unusable systems [7]. The resulting mean SUS score, considering the 10 users, was 82.5 points, which means the system has excellent usability.

The results shows that potential users, needs a briefing explanation of the flow of the application to learn some features.

4 First Functional Prototype

After the evaluation of the second high-fidelity mockup, the following iteration initiated the development of a first functional prototype. To this effect, an overall architecture was defined and a first stage of the core element of the system – message conversions – was addressed.

4.1 Architecture

The architecture described below generically adopts the AM4I framework [1]. Designed to support the design and development of multidevice multimodal

interactive systems, the adoption of this framework takes advantage of its decoupled nature where the interaction modalities and interaction management are decoupled from the application logic, supporting a variable and a dynamic number of input and output modalities. Furthermore, it is a scalable architecture so that, in the future, it will be possible to expand the context of use of the communicator into more complex interactive ecosystems, e.g., the smarthome, as a more integral part of the user's environment, overall, and important for its integration as an assistive technology.

Figure 3 presents the architecture of the proposed system. The architecture is decoupled, divided in a set of modules. They are divided by the ones running in the client-side and in the server-side.

The **client-side** provides the cross-platform front-end, able to support several different platforms to reach the most significant number of users. To this end, we choose to work with Flutter[3] since it allows creating native compiled applications for mobile devices, web, and desktops, which can potentiate the use of the system in a wider range of devices and contexts. Nevertheless, at this early stage of the work, the main focus will be on mobile due to the fact that it is an easy-to-carry device and people already use it as the primary communication choice with other users. As a simplification, in this iteration the systems' modalities are all on the client-side, but these and other modalities can run in other devices or the cloud, if required.

The **server-side**, encompasses all the logic and services for the Tell It Your Way system. Overall, it has six modules (see Fig. 3): the message conversion services, interaction manager, application logic, user service, messaging service, and data manager.

The *message conversion* module aims to translate from the modality received from one user to the modality selected by the other user. The *interaction manager* is a logical component and is responsible for receiving the events generated by the input modalities and producing new messages to be delivered to the output modalities [1]. This system proposes a single interaction manager located in the cloud, to which multiple devices can connect. It is directly connected to the *application logic*, which is the system's "brain", responsible for calling the service depending on what it receives from the interaction manager. For example, the user wants to convert a received text message to audio: this information is sent to the interaction manager, who forwards it to the Application Logic. This module will, then, call the Conversions Service to obtain the desired message conversion that will be sent to the client.

The *user service* is responsible for all user authentication, it will be possible to access and create a user's data. The user authentication is performed via Firebase authentication[4] supporting authentication using passwords, phone numbers, and well-known federated identity providers, such as Google, Facebook, and Twitter. Regarding the *messaging service*, it provides all the logic for the chat, allowing message exchange among users, in real-time. It adopts a stream protocol, and

[3] https://flutter.dev/.
[4] https://firebase.google.com/docs/auth.

an event-driven management making use of the EventSource/Server-Sent Events protocol[5]. Finally, for both User and Messaging Services, the components interact with the *Data Manager*, which is responsible for all database management. The database is a Cloud Firestore and the multimedia files are saved on a Cloud Storage. These services were developed in Node JS and Express JS.

As the Fig. 3 show, most of the components/services are deployed on docker containers, taking advantages of the simplicity and faster configurations.

4.2 Message Conversions

To provide a first level of message conversions to the first functional prototype, the literature was explored for existing works and libraries that could be considered. While the research did not cover just the technologies described ahead, these were those deemed more suitable for this stage of the development and additional reviewed solutions are not discussed for the sake of brevity.

Currently, the *message conversion* module supports six message conversions: Text-to-Speech, Speech-to-Text, Image-to-Text, Text-to-Image, Image-to-Speech and Speech-to-Image. It was developed in Python given the versatility and the large amount of libraries available supporting the envisaged conversions. To enable communication between the conversions services and the application logic, the Flask[6] library was used, a microframework (does not require private tools or libraries) that, aside from being simple and capable of doing the communication, is easy to set up and easy to start developing.

For *Text-to-Speech* we chose the Google Text-to-Speech (gTTS) library[7] that uses Google Translate's text-to-speech API. This library was selected because it supports a big set of languages, including English and European Portuguese. Therefore, it will be a great advantage, in the future, to support several languages.

The *Speech-to-Text* uses the speech recognition library[8], and it has support for several speech engines and APIs, online and offline. The API chosen was Google Speech Recognition[9], the API is free and does not require an API key to use and supports several languages.

The solution we found for Image to Text concersion was to identify all the objects in the image. There are several object detection methods, such as: YOLO (You Only Look Once), Faster RCNN, SSD (Single Shot Detector), OverFeat, among others [6]. We choose to use YOLOv3 [12] because it has good speed, high accuracy and it is open source. It is is an algorithm that detects and recognizes various objects in a picture (in real-time). The dataset used has 80 labels, coco.names[10].

[5] https://www.w3.org/TR/eventsource/.
[6] https://flask.palletsprojects.com/en/2.0.x/.
[7] https://gtts.readthedocs.io/en/latest/.
[8] https://pypi.org/project/SpeechRecognition/.
[9] https://wicg.github.io/speech-api/.
[10] https://github.com/pjreddie/darknet/blob/master/data/coco.names.

The conversion of text to image can be performed at different levels of complexity. For instance, the conversion of a sentence into pictograms may need to be different than just replacing every action/entity with the corresponding pictogram, since pictograms can have a more complex meaning than just a word or need to be placed in a different order than words. Additionally, recent methods have been proposed that generate images based on a textual description, e.g., DALL-E [11]. Nevertheless, and for the sake of demonstrating the concept, at this point, we opted for a first approach that just tries to replace each word by its visual representation, if it exists. To this end, we go through each word of the sentence and use Text2Picto [14] (which uses the Princeton WordNet [8] databases) to obtain the corresponding images.

Image-to-Speech and *Speech-to-Image* result from reusing some of the different conversion services, for instance, Speech-to-Image uses the Speech-to-Text and then Text-to-Image to achieve the intended result.

4.3 First Proof-of-Concept Mobile App

The implementation of the first mobile application considered the outcomes of the previous evaluations, focusing to solve the usability problems and provide a complete implementation of the provided techniques to validate the application further and enable the first assessment of their impact on users.

The current version of the Tell It Your Way application, developed with Flutter, is being tested on Android systems.

The new implementation has already solved some usability issues mentioned by the participants during the evaluation of the refined mockup. The bottom bar already appears with the icon and label of each option, making it easier for people less acquainted with these tools (in our tests, older people) to interact with it. Adding another user as a friend is no longer mandatory to access the friends' list to start a conversation with them. By adding a user as a friend, the user can quickly start a conversation with them. The save button is always visible to the user while editing his/her profile. Thus, making it possible to save changes at any time without having to scroll down to do so. Overall, the integration of the application with the current version of the message conversion services did not raise any issue and the full conversion flow is working well.

Figure 4 shows several screens for the Tell it Your Way application on Android and depicting, on the left, both sides of an illustrative situation between John and Jennifer. For the sake of simplicity, the image just illustrates how the messages sent by John can be received in a wide range of modalities by Jennifer.

John sends several messages to Jennifer, using different formats. On Jennifer's side, she receives the first message as text because it is her preferred modality. The second message is received in image format since, e.g. Jennifer was in the bus, surrounded by people and without her reading glasses. Finally, Jennifer leaves home and, while she is driving, John's third message arrives and is automatically converted to audio and read aloud.

Additionally, users can add the disabilities they have to their profile (Fig. 4b) and the messages will be shown to the user in the most appropriate modalities.

Fig. 4. Illustrative screens of the first functional prototype running on Android and depicting: (a) both sides of a conversation between John David and Jennifer Days with different messages conversions (image-to-text; audio-to-image; and text-to-audio); and (b) Amilcar's profile screen showing his preferred modality and disabilities.

5 Conclusions

In this paper we present first efforts towards a novel system to support communication between people with different needs and preferences. To this end, and adopting a user-centered design approach, we identified a set of requirements and reached a proof-of-concept that already supports sending and receiving messages in different formats and converting them to a chosen format regardless of their original modality. Our long term goal, for which this is a first stage, is to make communication mediated by technology approach the efficiency and versatility of face-to-face communication regarding an adaptation to users' preferences, abilities, and contexts.

Despite it only has been tested and evaluated with a few users, in controlled contexts, and the number and quality of the conversions can still evolve greatly, we can conclude that the presented proof-of-concept served its overall purpose, showing the potential and viability of such systems. In this context, other developments will follow that can further enrich the communication system, such as, the inclusion of an onboarding mechanism, the support for gestures, and first efforts on multilingual support.

Acknowledgement. This work was supported by EU and national funds through the Portuguese Foundation for Science and Technology (FCT), in the context of project AAL APH-ALARM (AAL/0006/2019) and funding to the research unit IEETA (UIDB/00127/2020).

References

1. Almeida, N., Teixeira, A., Silva, S., Ketsmur, M.: The AM4I architecture and framework for multimodal interaction and its application to smart environments. Sens. (Switz.) **19**(11), 1–30 (2019). https://doi.org/10.3390/s19112587
2. Amarasinghe, A., Wijesuriya, V.B.: Stimme: a chat application for communicating with hearing impaired persons. In: 2019 IEEE 14th International Conference on Industrial and Information Systems: Engineering for Innovations for Industry 4.0, ICIIS 2019 - Proceedings, pp. 458–463 (2019)
3. Bryan-Kinns, N., Hamilton, F.: One for all and all for one? Case studies of using prototypes in commercial projects. In: ACM International Conference Proceeding Series, vol. 31, pp. 91–100 (2002). https://doi.org/10.1145/572020.572032
4. Cooper, A., Reimann, R., Cronin, D.: About Face 3: The Essentials of Interaction Design, vol. 3 (2007)
5. Daems, J., Bosch, N., Solberg, S., Dekelver, J., Kultsova, M.: AbleChat: development of a chat app with pictograms for people with intellectual disabilities. In: Engineering for Society - Leuven 2016 - Proceedings, pp. 25–32 (2016)
6. Liu, L., et al.: Deep learning for generic object detection: a survey. Int. J. Comput. Vis. **128**(2), 261–318 (2019)
7. Martins, A.I., Rosa, A.F., Queirós, A., Silva, A., Rocha, N.P.: European Portuguese validation of the system usability scale (SUS). Proc. Comput. Sci. **67**, 293–300 (2015)
8. Miller, G.A.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge (1998)
9. Mirzaei, M.R., Ghorshi, S., Mortazavi, M.: Helping deaf and hard-of-hearing people by combining augmented reality and speech technologies. In: Proceedings of 9th International Conference on Disability, Virtual Reality and Associated Technologies, pp. 10–12 (2012)
10. Nielsen, J.: Heuristic Evaluation. Usability Inspection Methods (1994). Edited by: Nielsen J, Mack RL
11. Ramesh, A., et al.: Zero-shot text-to-image generation. arXiv preprint arXiv:2102.12092 (2021)
12. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-December, pp. 779–788, June 2015
13. Samonte, M.J.C., Gazmin, R.A., Soriano, J.D.S., Valencia, M.N.O.: BridgeApp: an assistive mobile communication application for the deaf and mute. In: ICTC 2019–10th International Conference on ICT Convergence: ICT Convergence Leading the Autonomous Future, pp. 1310–1315, October 2019
14. Sevens, L., Vandeghinste, V., Schuurman, I., Eynde, F.V.: Less is more: a rule-based syntactic simplification module for improved text-to-pictograph translation. Data Knowl. Eng. **117**, 264–289 (2018)