Voice and avatar face recognition with focus on familiarity and recall accuracy for use in a contact book designed for illiterates - Worksheets

Alex Patrick Hauge & Christian Bødtkjer Jørgensen

May 2013

Contents

1 Introduction

This worksheet is a companion document to the paper "Researching feasibility of using Voice Recognition and Avatar Face Recognition in a Contact Book". It contains explanations for theory used, test design and results and the implementation of the accompanying product. Because this is a companion document, there have not been made an effort to seam it all together which might result in the document seeming messy.

2 Theory

2.1 Introduction

In this chapter the theory behind the choices will be discussed. Each section contains a quick introduction and then some information about the sources used.

2.2 Voice Recognition Theory

A feature in the contact book includes using a button which plays the voice of the contact, thereby helping them recognize the person. In this section there is provided a description of voice recognition.

In our daily lives most of our social interactions happens with combining information from both the face and the voice of the person in order to identify him or her. Such as when meeting a person on the street or seeing one in the television.

One form of interaction that does not use the combination of facial and voice information is interaction through the mobile phone or over the net by using chat programs [3]. There is a reason why the brain combines both visual and auditory input for identifying a person. As is described in the paper by Campanella and Belin [3] there is evidence for the integration of information from face and voice in the human brain. A lot of research has gone into investigating the integration of sound of speech with the visual feedback from the articulating faces. This research is based on the interaction between speech and visual articulation, less is known about other types of facial and vocal integration of information. Using facial and speech input it is possible to learn a lot about a person's identity [3], such features are gender, age, body size(vocal tract length) or even emotional states, information that is of great importance when socializing with other people.

When recognizing a person when socializing, both voice and facial features rely on each other for the most precise identification of the person. As described in the paper by Joassin et. al. [7] it showed that when identifying a person, voice recognition was easier when simultaneously presented with the correct face and of greater difficulty when presented with a face not sharing the same identify. This shows that when listening to a voice, the listener cannot ignore a face when trying to recognize the voice [7], this was only the case for familiar voices as some experience is needed to identify a certain person. In [7] they performed a test using a setup of faces(F) and voices(V) and voice-face association(VF), each test-participant had to identify the persons included in the test after having familiarized themselves with the four identities. The results of this test showed the relationship between facial and voice recognition in relation to time. Voices proved to be identified at a slower rate than faces or voice-face associations; the pattern was the same in response to error rate [7]. Even though voice recognition is slower, it provides important information for identification and for the combined identification using face and voice.

When you talk on the mobile phone, you no longer have facial features for identification but have to rely on auditory information from the speech alone. When there no longer is any visual feature, you will start by trying to identify the callers by gender, young or old, familiar or unfamiliar and even the emotional state [9]. This information is derived from one single audio stream, it can be divided into two sections the recognition of the person and perception of what the person is saying. In order to figure out if sound is usable for identification over a phone it's also important to look into the importance of changing content of a sentence and most important how the sentence is said. In the paper by Lander et. al. [11] they investigate the effects of what is said and how it's said when identifying an unfamiliar person. When a person speaks, the mechanics behind the speech does not only produce the sound of the voice but also the movement of the face, which the brain takes advantage of [11]. This makes the visual information of the face important for speech perception, for both cases of loud noises and emotion perception. This said it is also possible to identify a person only from the face or the speech of the person. Some investigation indicates that for faces movement can also serve as an additional cue for face identification [11]. In general it supports that there is a cross modality when identifying a person, using the combination of both speech and facial information, combining them for an identification. The cross modality has been shown in experiments [11] where the participants were tested on their ability to match identify from a silently moving face to a voice and the same for voice to face. It showed that it was possible to identify, as long as the word or sentence used was also used during training. This suggests that the sequence of the speech is an important cue for matching face and voices [11].

The fact that the sentence is of importance when matching face and voices is necessary to keep in mind when design the tests, in order to keep the variables for the recognition as close to constant as possible. It is still possible to identify a face to speech even though the sentence is not the same as described in [11], this suggest that identify matching is not solely tied to the linguistic content, but that the nonverbal variation of the speech is sufficient to determine the identity. Features such as pitch, loudness and speaking rhythm provides important information for identifying as well as if the sentence is a question or not and the punctuation of the words and last attitude and emotions [11]. These features will all aid in the identification of the person at hand, some more closely tied to the facial expression than others. In [11] they perform a number of tests for these features in relation to face to voice matching and visa versa, these test procedures are used as inspiration for testing the feasibility of identifying an unfamiliar avatar with an unfamiliar voice.

For their test setup they use two sets, one with face-to-voice(FV) identity matching and one with voice-to-face(VF) identity matching, but using videotapes of the face and voice, giving motion to the target. This is not used for the paper which this chapter is written for, instead static avatar faces are used for the visual information. In order to get both sides, the use of voice-to-face matching and face-to-voice matching is used. In the paper [11] they do a total of six tests on the different features mentioned above, to find its impact on identity matching. The results found by [11] is useful for the design of the application, as it helps give a better understanding of the impact on identity matching by different features. In the first experiment they found no sign of impact by changing sentences, but changing the manner had an effect. Sentences spoken with the same manner had a significant better identity matching. This shows that the cross-modal identity matching is more tied to the manner than the content of the sentence [11].

For this test there were no differences between the matches FV and VF. The third experiment was regarding the effect of conversational speech and clear speech, they found no significant difference between the two [11]. They found that a change in manner at this stage also had a great impact on the identity matching. The fourth experiment was made to find the effect of conversational speech and casual speech, as it can change their speaking rate. In the fourth experiment they again found the manner of the spoken word to have a great impact on the identity matching.

It showed that matching conversational to conversational sentences was significantly better than in the case of casual to casual matching. In the fifth test the look into the effect of artificially slowing the voice, which showed to have no effect on the identity matching. The artificial slowing was done in such a way that the fundamental frequency was kept the same. The final experiment involved speeding speech and conversational speech. From this experiment it showed that there was no effect of changing tempo. These results give a better understanding of the impact different features can have on identity matching.

A place that has used identification by voice is the court of law, when trying to identify a person by voice from people that was close to the incident or in cases of threats over the phone [5, 12]. The interesting factor in use of voice recognition in court is the effect of time between the crime and when the witness has to identify the accused. In a case described by McGehee [12] a positive identification of the defendants voice almost three years after the incident, was accepted by the court as material evidence, but what is the effect of time when recognizing an unfamiliar voice. When hearing it can be very difficult to identify a sound of a specific character and hard to assign them to what has caused it. Several features of sound can be misinterpreted, such as the direction of the sound as well as quality and intensity. A simple experiment described in [12] shows how hard it is to identify the character of the sound as well as cause.

The experiment was performed by the teacher striking a tuning fork beneath the desk, then the students one by one had to try and identify the sound source. Only 2 out of a hundred guessed correctly, showing that the judgment when determining a sound source is highly prone to error. Franches McGehee [12] made an experiment to find out the impact of time. A number of students were tested, each first got a paragraph of 56 words read to them from behind a screen and told to listen carefully as they will be tested later. The voice presented for them is used for training the test participant. The intervals of time that is used in the test is 1, 2 and 3 in days, weeks and months. The interesting results from trying to identify the voice after certain time intervals are shown below [12] in Table 1. The test participant was presented with 5 voices, reading the paragraph from behind the screen, and then having to identify the previously heard voice. For the test a total of 189 men and 155 women were tested. These results are achieved using unfamiliar voices, but give a good view of the decline in recognition of unfamiliar voices over time. When talking to a person over the phone, you sometime will en-

Time Interval	Percentage Correct
1	83.0%
2	83.0%
3	81.0%
Weeks	
1	80.8%
2	68.5%
3	51.0%
Months	
1	57.0%
2	35.0%
3	13.0%

Table 1: Table describing the correct recognition of a voice with different time intervals.

counter several voices depend on the place and situation. When there is no longer one single audible track, stream segregation takes place in order to focus on the voice needed. So far the theory has only covered cases with unfamiliar targets, but as this project will use familiar target primarily its necessary with further information. In the paper by Rochelle S. Newman and Shannon Evers [14] they look into the effect of familiarity of a person when trying to recognize the voice in a stream.

The test was setup with 67 students that participated in the test, 24 of these were following a class with a professor that is not the target. The last 44 students will have the target speaker as the professor for the course they were following. The test is then performed later in the semester [14]. This is done to learn and thereby familiarize them with the target. The stimulus for the test was the professor reading both a list of words and a story passage in the style he normally spoke when teaching. The group of test participant would be divided into two, one where they are told who they are listening to (explicit knowledge group) and the other half were not (implicit knowledge group).

As is stated in [14] the familiarity effect is not transferred from fluent speech to a list of words. Beside the target professor, a second male recorded the same passages and words. The target and non-target voice was then intertwined, for the test participant to track. They found that people in the explicit knowledge performed the stream segregation with fewer errors in form of missed words compared to the implicit group. They found no significant effect from familiarity that the one group had as they did not perform significantly better than the group that was not familiar with the professor when tracking the voice. This shows that simple familiarity that all gained in the test as they were given one voice in the beginning to follow is enough to aid in the stream segregation compared to a whole semester of familiarity [14].

2.2.1 Conclusion

In general when identifying a person from the voice, it is a process that is closely tied to the visual input in form of facial expression and movement. It is possible to identify without facial input, but with sound there are several features that play a role in how accurate the identity matching is. Such as the manner of how a word is expressed which include pitch, tempo, timbre all part of the change in the voice that can occur from emotional states and more. Beside these features the effect of time between hearing the voice and trying to identify, is crucial as the accuracy declines to around 50% within 3 weeks for unfamiliar targets. This is important to take into account, as this project is about creating a contact book for illiterates. But since the voice is not to stand alone, but is aided by an avatar image of the person thereby using cross-modality, it should decrease the error rate on identifying the person in the contact book.

2.3 Facial Recognition Theory

In this section some relevant parts of face recognition will be discussed. The most relevant ideas throughout the current research is the ideas of feature-specific perception and holistic perception. These two functions as different explanations of how human beings transfer the information of the collection of feature in the face. The feature-specific perception is explained by the observer scanning a face explicitly for local features, such as eyes, eyebrows and mouth, and using these features as individual parts to recognize the face. Holistic perception is the other way around, by having the observer see every local feature as a whole percept, which is then a collection of all the features that makes a face, including hair, chin, etc.

Belle et. al. [2] took a look at the differences of these two views and tested a brain damaged patient who suffered from prosopagnosia (impaired face recognition skills) and a control group. The test measured performance in different situations where certain parts of the face was masked out, as to render the holistic perception inert or vice-versa with the feature-specific. They compared average accuracy and speed of the control group with the patient, which showed that the patient performed clearly worse than the control group in two cases. The patient was much slower in each task and had less accuracy when showed a full face or a masked face. They were shown either a full face without hair and ears, a windowed face which only showed one feature or a masked face in which one or more features is masked out. The interesting thing about this test is that the control group drops to the same accuracy as the patient when tested on the windowed part. They also did an experiment to see if the size of the mask mattered which they found out it did not.

The different features found in the face also combines to express emotions, show attention, gender, age and other social information. Yueting Sun, Xiaochao Gao and Shihui Han [15] experimented with this extra social information by trying to see if there was a difference in performance when measuring this information between gender. They used Event-related potentials to test for performance in both female and male participants, by showing them pictures in which they had to tell face orientation(low-level perceptual feature) or e.g. gender(high-level social feature). The results show that females has more attention towards the social features and performed better throughout the trials.

Steven G. Young et. al. [17] also looked at the social information and looked for cross-race/same-race differences when recognizing faces. The hypothesis is that people who have never seen another race of humans will not have been trained to be as good at differentiating between faces as people who have had exposure to that race. They review current research and argues for future research in the area. In conclusion they argue that a collected framework which seeks to unify the current theories about the cross-race/same-race problem will further enhance the research and might provide clear results.

Robert G. Franklin Jr. and Reginald B. Adams Jr. [8] also examined the relationship between facial features and deeper emotional meaning being given to these. The participants that took part in their experiment were shown a random series of faces that they would have to remember for a later recall task. The participants had to find a face they had seen before in a series of a mix of distractors and faces they had seen before. The results was supporting their hypotheses, a) individual differences in the ability to decode emotional messages from expressive faces would be positively associated with the ability to encode and subsequently remember a separate set of neutral faces in the same participants, and b) that stimulus-level differences in the extent to which a separate group of raters ascribed emotionality to these same neutral faces would also be positively associated with face memory.

2.4 Illiterate Theory

This project is concerned with illiterates and the difficulties that they encounter, when interacting with new technology through the standard user interface. As most user interface applies text to inform the user of the different functionalities, this means that illiterate people have great difficulty with understanding the interface as they cannot read. During this chapter the focus will be on illiterates and the challenge of designing user interfaces, which are independent of text.

When learning to read and write, it also adds to one's ability to match both basic units of language to words as well as the smallest semantically units in the written language, to internal system for the language [4]. As described in the paper [4], it seems that the beginning of the oral language is also affected by the process of learning how to read and write, indicating that both the oral system and that responsible for reading and writing interacts. For illiterates this mean as indicated by studies [4] that their ability to deal with phonetics, that is units of speech, is not dealt with automatically but is a result of learning to read. With this disability most illiterates has to find another way to deal with the problem of interacting with computers, than literate people.

A case that most illiterate people have to deal with, especially in the western world is withdrawing money from an ATM machine, as most interaction with ATM is through a text based interface. In the article by A. Thatcher, S. Mahlangu and C. Zimmerman [16] they investigate how an icon based interface can be created for illiterate on the ATM. As they describe there have already been suggested a number of alternatives for the normal text based interface. An example is a system for blind people that also have a difficult time interacting with the interface that they cannot perceive, as well as systems for aged people [16]. The problem for both these groups is the written interface just as it is for illiterates, some investigation has gone into using voice recognition as an alternative interface for blind people. This speech based interface could also prove useful for the illiterates. The problem of a speech based interface is when used in a country that is multi-cultural, causing a lot of different dialects. Beside a speech based interface another possibility is an icon based interface for illiterates. As is explained in the paper [16], icons prove to generally lead to a faster recognition when learning a new system and are remembered easier than the equivalent sentence that would have replaced the icon. This is due to that the icons may be stored both in the visual and verbal memory.

In the paper [16] they chose to follow a user centered design, focusing on the most common functionalities of an ATM machine, such as withdrawal. In order to test their interface design a test was made testing three different ATM interfaces with two groups, one literate and the other illiterate. The three different designs are a text-only interface, icon-only interface and a text-and-icon interface [16]. Each design is randomly assigned to each test participant, and asked to perform a task. For the test 25 illiterates were used and a group of 29 literate subject. Both of these groups are familiar with the use of an ATM machine and have had a bank account for a year or more. None of the people in the groups speak English as the native language [16].

In the test of the three interfaces [16] they found that a number of people withdrew from the task before finishing it, both from the illiterate and literate group. For the test of the performance results from the task, the result was measured as the time it took from they started until they finished the task at hand.

When looking at the result from the paper [16] regarding transaction performance of the three different interfaces and two groups, there is a significant difference between the literate and illiterate group. The two literate groups that had text design and text-icon showed to be significant faster than the three illiterates groups with text, text-icon and icon. Within the illiterate group, people performed fastest in the interface with text and slowest in the icon interface [16]. These difference was not statistical significant but worth noting. For both the illiterate and literate groups there was no significant change between the text and text-icon design. As might have been expected the literate group proved slower at icon interface compared to text interface.

The reason why illiterates proved faster in the text design might be due to the design of icons and the conflicts with the general icon design used in ATMs. The reason could also be that the illiterates has learned to recognize certain words, relating its function to the visual construction of the word. This is useful for the design of the app, that it might not be bad with text in the interface, although the ultimate goal is to create a design efficient enough to translate the information without words.

Beside ATMs, illiterates also have to deal with the interaction design of mobile phones, which has a very text based interface design. In the paper by Kristin Dew Et.Al [6] they research the subject of how to create an alternative interface for mobile phones, for illiterates. Normally when illiterates deal with the standard interface designs from mobile phones, they find workaround in order to cope with the text based interface [6], for example by using memory patterns. The goal for Kristin Dew Et.Al [6] was to create an interface for illiterate mobile devices, within the United States. The product that they develop [6] is a program that runs in the background not visible to anyone, thereby leaving the interface as standard instead of modifying it. By not modifying the interface, it helps to not stigmata the illiterates and at the same time help support learning of reading. The program is an on-demand playback, meaning that it plays back text-to-speech, activated by using gestures. This feature will read words allowed from text messages and more. This helps the illiterate to navigate and use functionalities easier at a discrete manor. This is still a work in progress, but the way of dealing with illiterates problem of navigating a mobile interface by using a playback of text is a both useful concept and discrete. Sound can help with closing the gap that illiterates have with text, in the project this is used to help them identify persons in the contact book of a mobile phone.

The study made by Zereh Lalji and Judith Good [10] is concerned with investigating how to design a mobile phone for illiterates, that include their expectations and a user designed interface and what they might need for support. Performing a user centered design is useful as a designer does not know what an illiterate might need as well as the impact the experience of an illiterate person has on interacting with an interface. The study by Zereh Lalji and Judith Good was divided into two sections one concerning the different features that illiterates would like as well as design, the second is related to the use of the phone by illiterates and how the features can be incorporated.

By interviewing a group of illiterates they try to identify the needs of illiterates when using a mobile phone. The illiterates found that they were not educated enough to use a phone, some even sold it when given one [10]. When asked for reasons for wanting a mobile phone, they tell that they want to improve businesses, contact with families that might live apart from them. When dealing with words and numbers in the case of using a phone, the illiterate's ability to associate and recall was well developed. To deal with saving phone numbers, the most important numbers was memorized and the other written down in a telephone diaries, or used the recently called function. The way they identified the numbers to a person was by features like the style of the handwriting, page number, markers or ink color [10]. Some of the illiterates was not able to tell time but used the arms on the watch to tell when they needed to be somewhere. This gives a better understanding of how they deal with the different task when interacting with text and numbers on a phone, information useful when designing a contact book for illiterates. As is explained in the paper [10] they found a big gap between the western icon design and what was understandable when shown to illiterate people not living in the western civilization. So dependent on what country or region the design is made for it has to be adapted to that culture. The people in this study are not only illiterate but also illiterate in technology causing some difficulties that might not be the case for illiterates in more developed countries, but need to learn this. An interesting feature that was incorporated into their high fidelity prototype to deal with the difficulties that illiterates had with learning and coping with the use of the phone, was a voice instruction that helped them navigate and use the phone.

In the article [13], the problem of creating an interface for illiterates on the mobile phone is dealt with by creating an interface using icons, audio and cartoons to communicate the information. The interesting section of this article, is how the problem of assisting the illiterates handling new programs, is handled. When an application is opened a smart video dramatization is launched, which illustrates how it works. This might prove to give a better adaptation for different cultures as video is a powerful tool, and better can convey different cultural expression.

Another aspect of how illiterate cope with using a mobile phone memory, in this study by Syed ishtiaque Ahmed, Maruf Zaber and Shion Guha [1] they test and interview 15 illiterates from Bangladesh who are not trained in digital technologies as well. Each of these 15 subjects is required to have used a mobile phone for at least 1 year. They found that all the participants would use the call list, which include the out and ingoing calls, and when saving the calls it was saved on the memory of the phone [1]. There were 13% of the participants that would not make call for them self's and of these there were just 47% that saved the numbers. By looking at this it shows that they use a system of the last called as the current contact book is not efficient enough for illiterates, as they use the call list instead. Besides using the call list to locate a contact, they also found a number of other methods that the illiterates applied. One method was to memorize the last three digits of the phone number; this causes some problem since more than one number can have the same three digits [1]. Others were dependent on getting their contact saved in the contact book, often with help from others; by having them in the contact book they would remember positions of the contact in the list. A very different strategy that was applied was trying to memorize the contacts name as an image, using the way the names was put together with letters as a figure. These also had help save the contact in the contact book before applying the tactic of creating images from names. The last tactic that they found was to use the frequency of letters, by memorizing the image of different letters. The illiterates using this tactic would then search the contact book for a contact with a specific frequency of letters. An example of this could be the name "Lotte" with two "T"'s in it [1].

2.4.1 Conclusion

The information of how the illiterates deal with the phone [1] is useful when designing a contact book, as it gives an idea of how the illiterates deal with the different obstacles of the interface, and at the same time what problems they encounter using the phones memory. Finding a way of creating an interface that would make these problems obsolete would increase the usability of the way the illiterates use the contact book. In general the icon interface that needs to be created needs to be adjusted to each culture as the norms are different, as the semiotics are different within cultures. If a pure icon based design is better for illiterates then one combining the two is hard to decide, ideally an icon based interface should be able to convey all information needed if created right according to culture and more. But at the same time using text combined with the symbols also help with the learning process of words for the illiterates.

3 Implementation

When creating the contact book for the illiterates, the knowledge of how illiterates use mobile phones has to be combined with the design of the contact book. During this chapter there is given a description of the design that lead to the prototype as well as the implementation of the design.

The success criteria for the implementation are described below:

- Functional interface.
- Basic functionality of a standard contact book.
- Creatable avatars.
- Record incoming/outgoing calls.

To implement the contact book for mobile phones, it is necessary to choose what operational system (OS) to develop to. For this project the OS that is developed for is Android. In order to write code for Android OS, the eclipse environment is used as a base for coding Java and Xml. To develop efficiently, the code is accessible through GitHub, allowing several people to code simultaneously. For the development the Android SDK is used together with the ADT plugin for Eclipse. This setup will serve as the base for the development environment.

3.1 Design

In the process of designing the application that should work as a contact book it is necessary to keep in mind the situation that illiterates are facing, not being able to read words. This means that an ordinary interface with text is insufficient. The design during this chapter is based on the knowledge of how illiterates interact with mobile phones. The first step is to determine the needs and how to deal with the problem that illiterates have. The way that the illiterates gap is tried fixed, is by using an avatar created by the user to work as replacement for the normal text interface with just a name and number. To further improve the recognition of this avatar a playback function is tried implemented, but due to hardware restrictions it was not possible, playing a period of conversation from the last call to that contact. These two ideas is the base of this prototype contact book.

To better understand what is needed to design the contact book, a brainstorming of different functionalities was performed in order to determine the needs. Inspiration is also derived from looking into how the contact books of newer smartphones are designed. Below in figure 1 is a flow chart of the different functionalities that was derived in the process. The functions displayed in figure 1 are those intended for the contact book, as this is a proto-type the modify contact is not implemented. These functionalities needs to be incorporated into an interface that can manage the needs required for the functionalities to work as intended. The functionalities and their relations are going to serve as a base for creating the interface. The design of the interface containing the functions described above, is divided into four main screens:

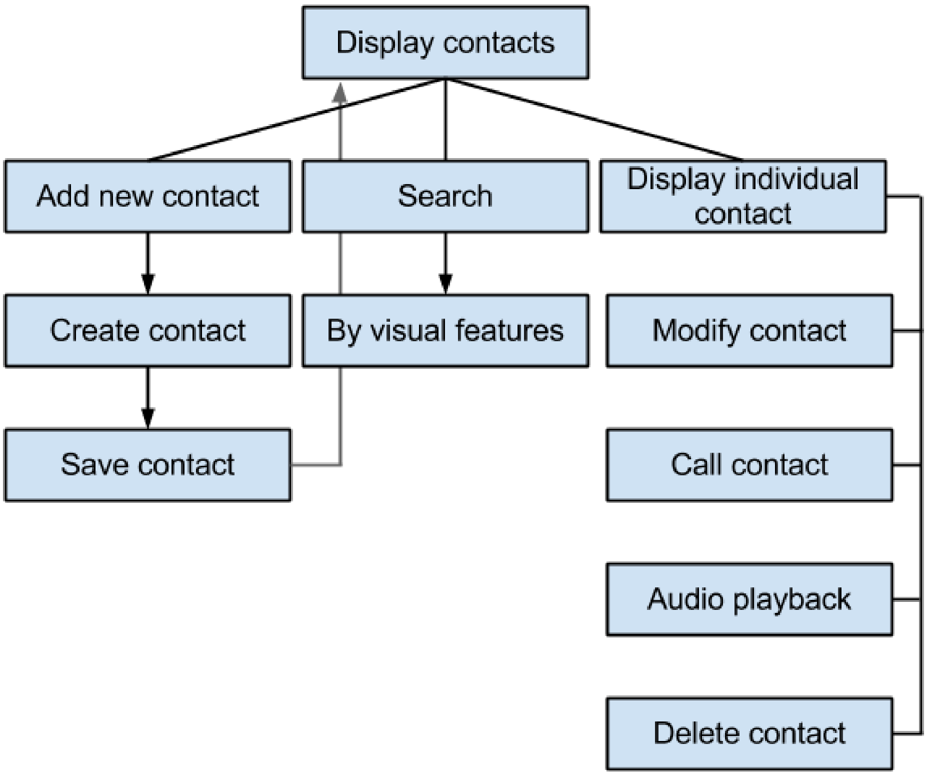


Figure 1: Flow chart of the different functionalities.

- Display of contacts.
- Add new avatar.
- Contact information.
- Search.

This first screen is the one you will enter when opening the contact book, containing all the contacts and two buttons leading to the "search" and "Add new avatar" screen. The structure of this interface is visualized in the concept drawing in figure 2. As is displayed in figure 2, the display of the contacts is made into two columns with the possibility of scrolling up and down through all the contacts. Each contact in the grid has the avatar displayed for that person as well as number and name. The reason for this is that as described in the paper by Ahmed et. al. [1] one of the strategies that illiterates use when navigating a contact book is remembering the last three digits. In the final design the names are not shown in the grid only the phone numbers. By using the number the design supports their strategy used when navigating a contact book. The next screen is the one containing the information of a specific contact, which will show when pressing the contact on the grid, you want to view. The design is shown in figure 3. As is seen in figure 3,

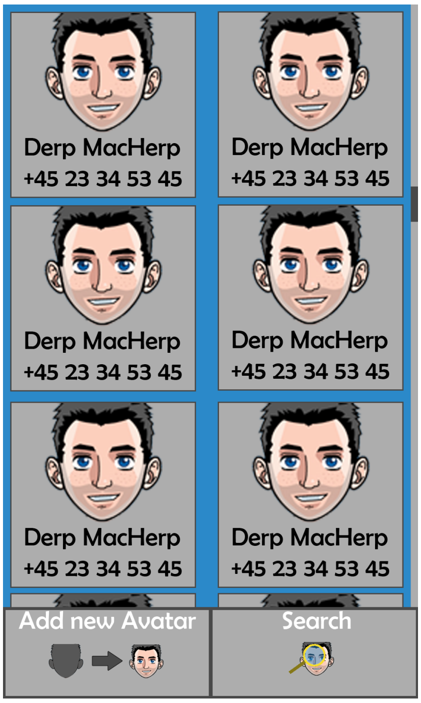


Figure 2: The Start screen of the contact book

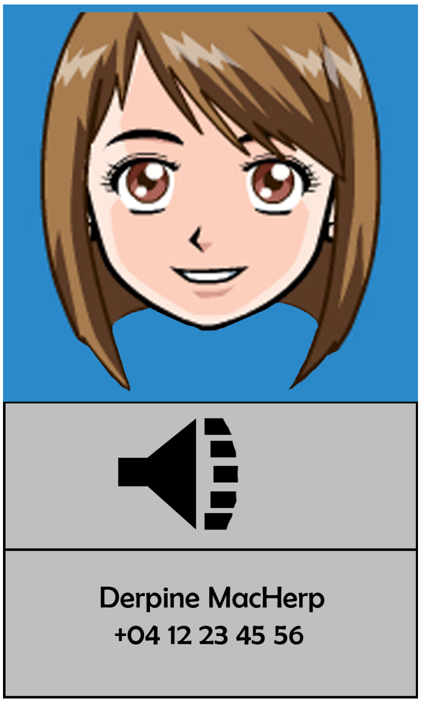


Figure 3: The information screen for the contact

this screen is divided into three sections. The first holds the avatar image, next followed by the button for playback and then the name and phone number. In the final design there is added a button with a cross, positioned to the right of the sound button. This button will handle deleting the contact. The name and number when pushed will perform the call to the contact. The next screen design is for the "Add new avatar" interface, this is shown in figure 4. The structure of the interface in figure 4, is divided into a number of sections. In the top, the ability to create the avatar of the contact is placed. Each arrow will change a feature of the avatar thereby making it possible to create an avatar representing the subject. In the final design the accessories are displayed in a grid instead of one box. The number of features in each category will help make it possible to create an even more unique avatar. When the save button is pushed, a follow up screen will emerge. This screen will allow the user to enter name and phone number and finalize the creation of the contact, which will be saved and showed in the entry screen. The last screen is the one concerning the search function, this interface is displayed in figure 5. In figure 5 the search interface is displayed, similar to the entry screen, it contains a scrollable grid with the contacts in. This grid will update as the user select and deselect different features in the search for the specific contact. An example is when clicking the button controlling the feature that is glasses; it will

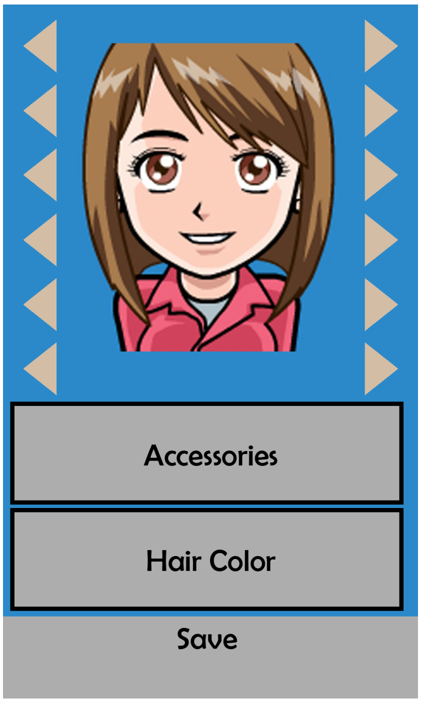


Figure 4: The "Add new avatar" screen, when saving a follow up screen will show allowing the name and phone number to be entered.

either remove all contacts with glasses or all those without dependent on the state of the feature. This way of searching should help the problem that illiterates have when searching a list of contacts in the standardized contact book of smartphone. As illiterates can't search for names as they are not able to read, it allows them to search for the visual features of the avatar they have created to represent the person in mind.

3.2 Implementing the Design

In this section of the chapter, the implementation of the design described in the previous section is described. A general description of the code design is used to better describe the code and structure of the program that supports the functionality mentioned in the previous design chapter. Interface The first element that is covered is the code design for the interface implemented in the project. The interface is coded in two manners, one using Xml language supported by Android SDK and the other in Java alone. Java and Xml can also be combined. The Xml alone provides a fast and easy way of creating an interface using different layouts, the drawback with pure Xml is that it does not support dynamic interfaces on its own. For a dynamic interface you need Java alone or combined with Xml, this

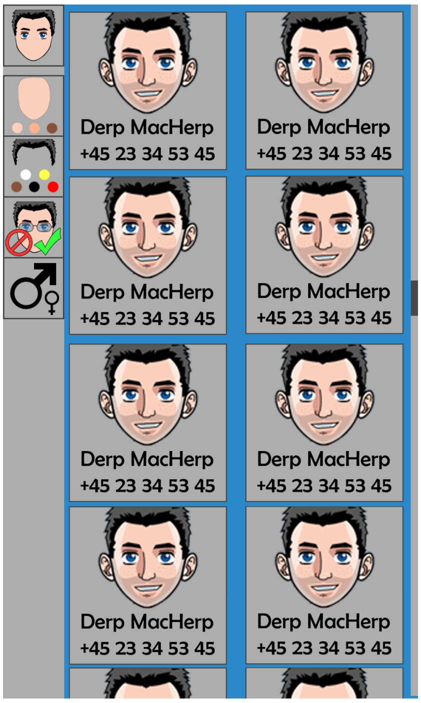


Figure 5: The interface of the search screen, containing a scrollable grid and feature selection for search.

allows changes in real time. The interfaces are constructed using a number of different layout managers that allow for arrangement of icons, buttons and more into the layout needed. The interface is built within each activity, each activity works as a screen giving a surface to display the buttons and graphics on for the user. To get a better understanding of the entire frame of activities and the related interfaces, which provides the base for the functionality, a flow chart has been created as displayed in figure 6. As is shown in figure 6, it's possible to go from one activity the other, providing the general structure of the application. Within each activity is the interface programmed, connected with an underlying code that supplies the functionalities for the different buttons in the interface.

3.3 Avatar Creation

In the avatar creation screen it is possible to cycle through variations of hairstyle and color, noses, eye color and shape, mouths, chin builds, gender, skin color and if the person should have glasses or not. The amount of variations differ to increase variety though optimally, more variations is better. Given the current amount of feature variations, a very high amount of combinations become possible. It ended up looking as in figure 8. In figure 7 a quick flow chart over the

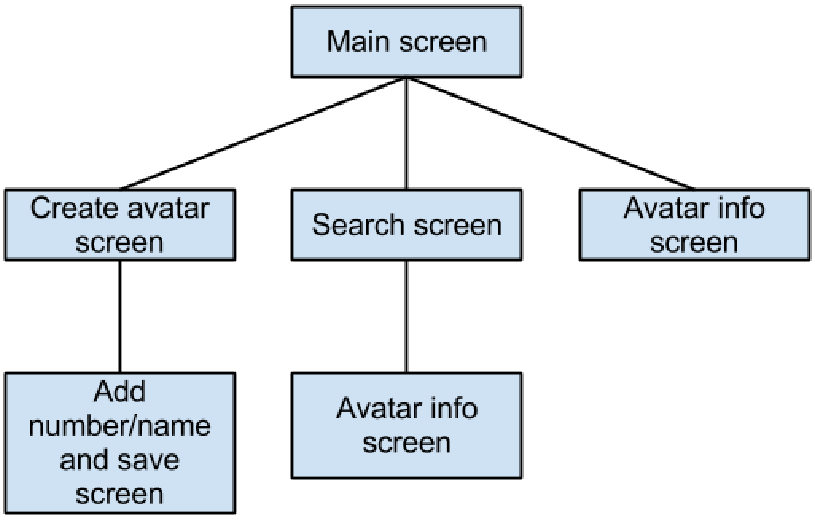


Figure 6: Flowchart of the activity setup.

components included as functionality can be seen.

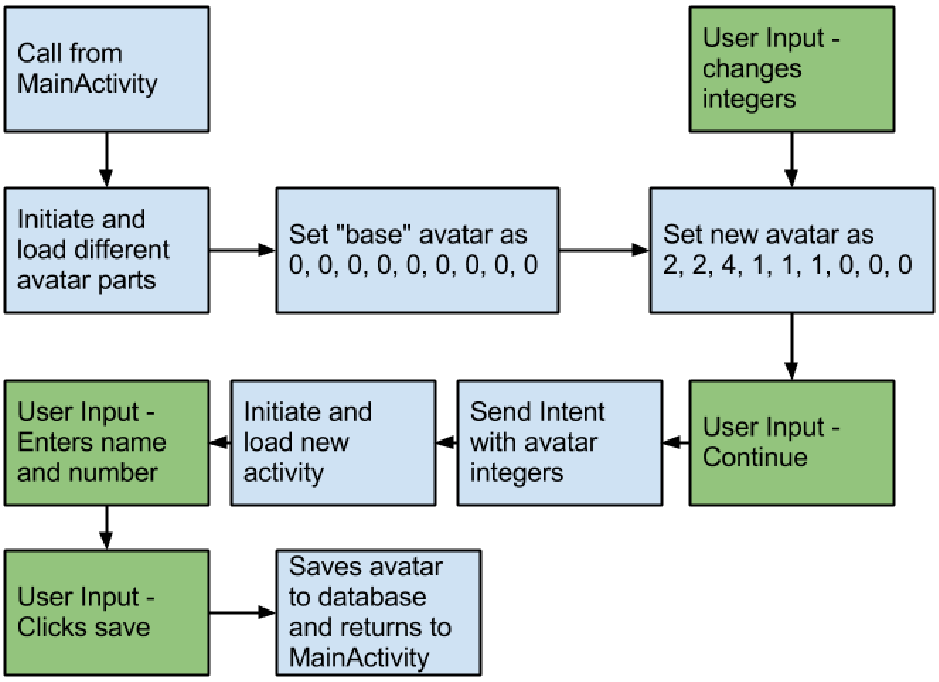

Call from MainActivity
Initiate and load different avatar parts
Set "base" avatar as 0, 0, 0, 0, 0, 0, 0, 0, 0
User Input - changes integers
Set new avatar as 2, 2, 4, 1, 1, 1, 0, 0, 0
User Input - Continue
Send Intent with avatar integers
Initiate and load new activity
User Input - Enters name and number
User Input - Clicks save
Saves avatar to database and returns to MainActivity

Figure 7: Flowchart over avatar creation.

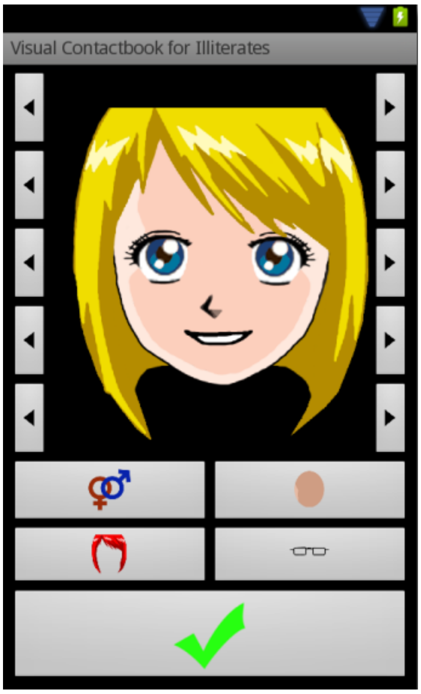


Figure 8: Avatar creation screen in final application.

3.4 Sound recording

In this section the implementation design of the playback feature is described. This design is split in two parts, one concerning the playing of the recorded voice and the other with regards to the recording of the voice during the phone call. Due to restrictions in the API and phone setup, the function created by android for recording phone calls was not accessible on the tested phones. The design in this chapter works as intended but the one hurdle that the tested phones are not permitted to access the android function called voice call. Both the playing audio part and the recording audio are tied together by the database that is explained later in the chapter.

As described in the design section, it is possible to play the voice of a contact by pressing a button within the avatar info screen. This button when pressed will try to locate a sound file in a specific file on the phone; this file will have the same name as the number of the contact. To get a better understanding of the design behind the playback function, a flow chart is created in Figure 9. When setting up the playback function there is first created a button as seen in figure 9, to this button an "OnClickListener" is assign which listens for any activations of the button. If the button is activated it will create a "MediaPlayer" that will handle playing the audio file. Before starting the "MediaPlayer" the recorded sound file needs to be

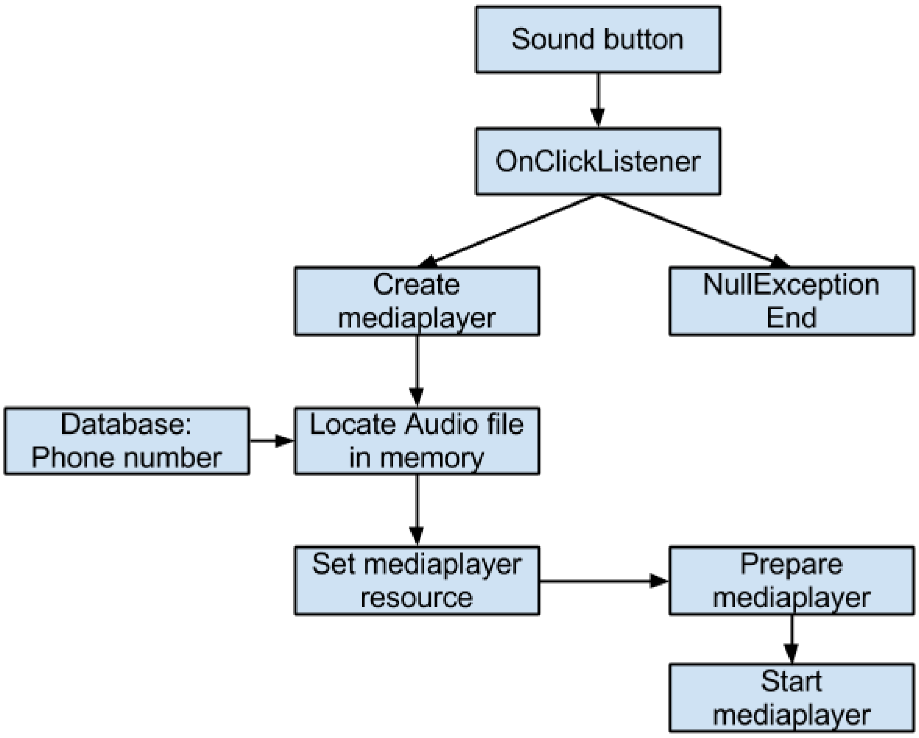


Figure 9: The flow chart of the sound button for the playback function.

located on the phones memory. The audio file when recorded as is explained later is saved as the phone number of the person that is called. Knowing what folder the files are stored in, it's possible to search for the file by acquiring the phone number of that contact through the data base and use it as a part of the address for the audio file. When the audio file is located it is set as the "DataSource" for the "MediaPlayer", this allows the "MediaPlayer" to prepare and start to playback of the sound. Before the audio file can be played it needs to be recorded. This design is made to record directly from the uplink and downlink in the phone call, recording both incoming and outgoing conversation. It was discorvered at a late stage in the implementation that this functionality provided by android was not accesable for all phones as some phones had blocked it. The design is still made with this function as it is what is intended for the product, instead of having to do a low tech and very indiscrete recording over the speaker whenever talking with a person. This is not an option, as there should be no signs when using this application, that can cause illiterates to feel different than the population. To get a better overview of the structure of the record audio design, a flow chart is displayed in Figure 10 explaining the structure. The whole process of recording the voice during the call starts when the record function as depicted in Figure 10 receives an

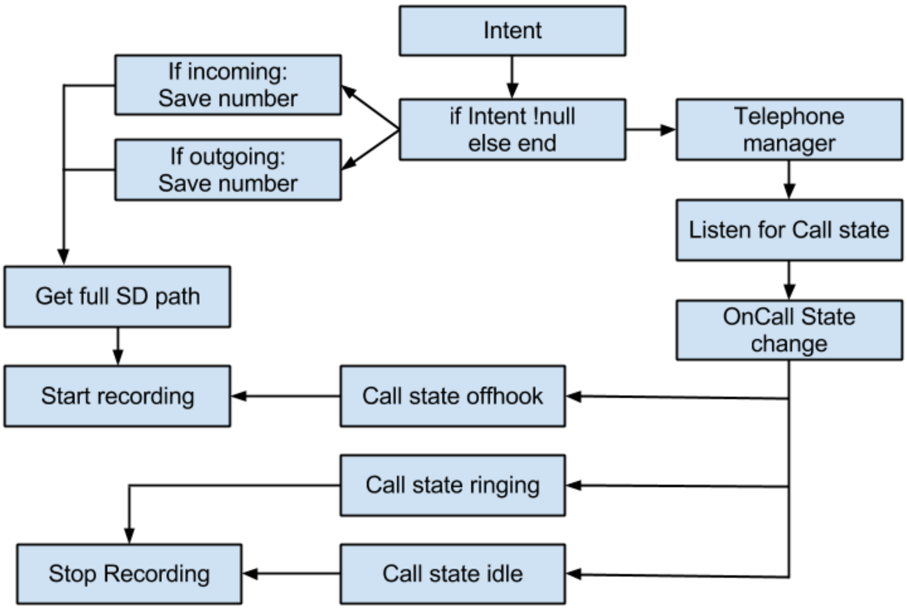


Figure 10: The flow chart of the sound record design.

intent. The intent is specified in the manifest file, where an intent filter is applied to this function only letting it receive specific intents. If the intent is not null it saves the phone number of the person that is calling or called which goes into the function that sets up the path for the memory of the phone, for later to work as the address of the newly recorded file. The essence of the record function is the "Telephone manager" the applies a listener that listens for changes in the call state of the phone. When the call state of the phone enters the "Offhook" it starts recording, as it indicates that the call is in progress. The two other states "Idle" and "ringing" both ensure that the call is stopped as a call is no longer in progress at that state. The result of this function, when the start recording is called, is an audio file saved in a specific directory with a file name similar to the phone number of the called person. This allows for the playback function when in the avatar info screen to play back the sound knowing that the audio for this contact, has the path with the phone number as the name of the file. At the current state the function produces an audio file that is saved with the correct name in the correct folder, fulfilling the recuirements for the playback function. The problem is as the phone it is tested on is restricted, it is not allowed to initiate the "prepare" state when starting the recording as the API and phone restrictions prevent the use of "voice call" as mentioned earlier. This results in an empty audio file in the right position and with the right name.

3.5 Search

The search is based on visual features which is represented by an integer in the database. As such an avatar could be represented by e.g. the series of integers 0, 1, 3, 4, 0, etc. The search algorithm works by selecting those numbers who fits the current search series of numbers. The search algorithm also contains a neutral state in which it sends -1 through the algorithm disabling the sorting of that particular part. Figure 11 describes a scenario in which three avatars exists in the database and the user searches for e.g. hair color black represented by the 0 in the search function. The search algorithm then checks the avatars in the database to see if anyone has a 0 in the corresponding position and disables any avatar in the list who has not got the 0. This ends up resulting in only one avatar in this scenario.

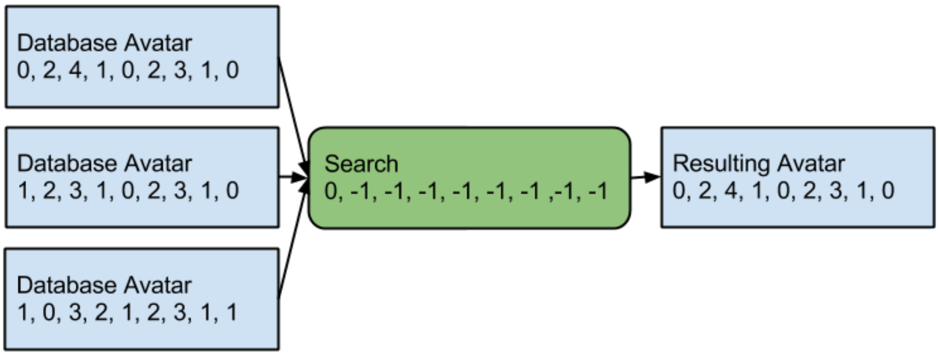


Figure 11: Search algorithm.

3.6 Database

The database is structured to contain information for each avatar. The list below details which items are saved and in what type of data they are stored.

- Integer index used for easy access
- String name of contact
- String number of contact
- String filepath of the sound that has been saved with the contact
- String filepath of the avatar image saved at generation of avatar
- Integer representing hair style

- Integer representing hair color
- Integer representing eyes type
- Integer representing mouth type
- Integer representing nose type
- Integer representing jaw type
- Integer representing glasses
- Integer representing skin color
- Integer representing gender

3.7 The Final Product

During the implementation there have been made a number of corrections to the concept design described in the beginning due to some missing features as well as the results from the usability test performed on the application. The entry screen as described in the design section, still has the same features, though without the names for the avatars in the grid. For the future the "add avatar" icon needs to be modified accordingly to the input derived from the usability test. The current version of the entry screen is displayed in Figure 12. The next screen is the one used for creating the avatar. In the concept design it uses only one button for accessories and one for hair color, using text to add information on the buttons. In the final version this is changed to four buttons with icons on for different accessories and features of the avatar. The save button is also changed from the text "Save" to an icon. The current design of the avatar creation screen is shown in Figure 13. The gender icon and skin color icon was changed due to results from the usability test. After setting up the avatar you continue to the next screen that allows you to enter name and number for the contact you are about to add to the contact book. This screen was not design during the concept design, as its was a small part. The final version is display in Figure 14, it contains a icon of the avatar they just created with a field use for entering the name, and below a image of a numpad with a field for entering the phone number. The next screen if for viewing information of the avatar. In the concept design of the screen it had only three parts, the image of the avatar, a sound button and the button with name and number of for calling the contact. In the final version it has changed both the sound button and name and number button to not show as a button. Beside the sound icon has been changed to a more detailed version. Another change is the introduction of the delete button next to the sound button, visualized by a red

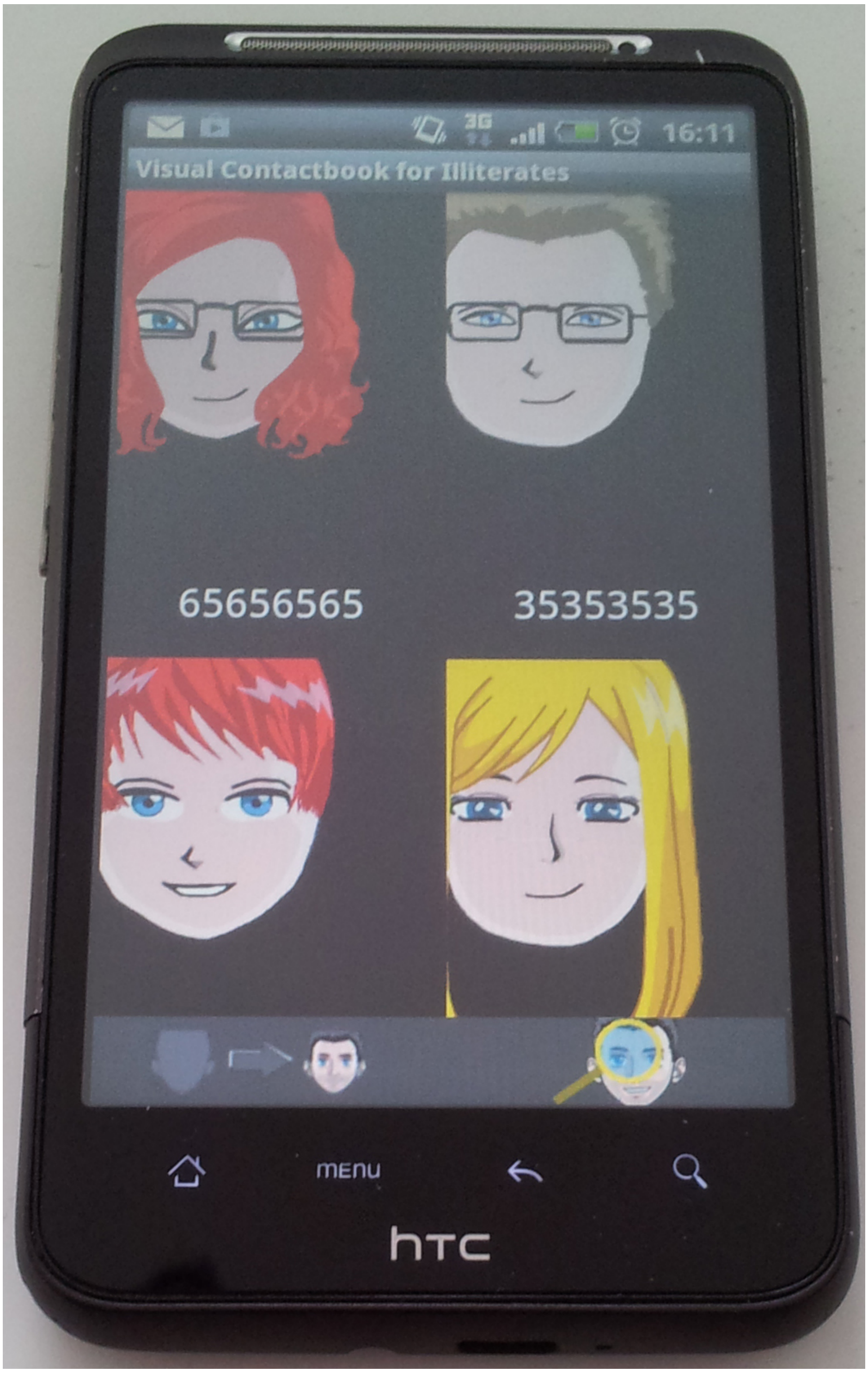


Figure 12: A screen-shot of the current version of the entry screen.

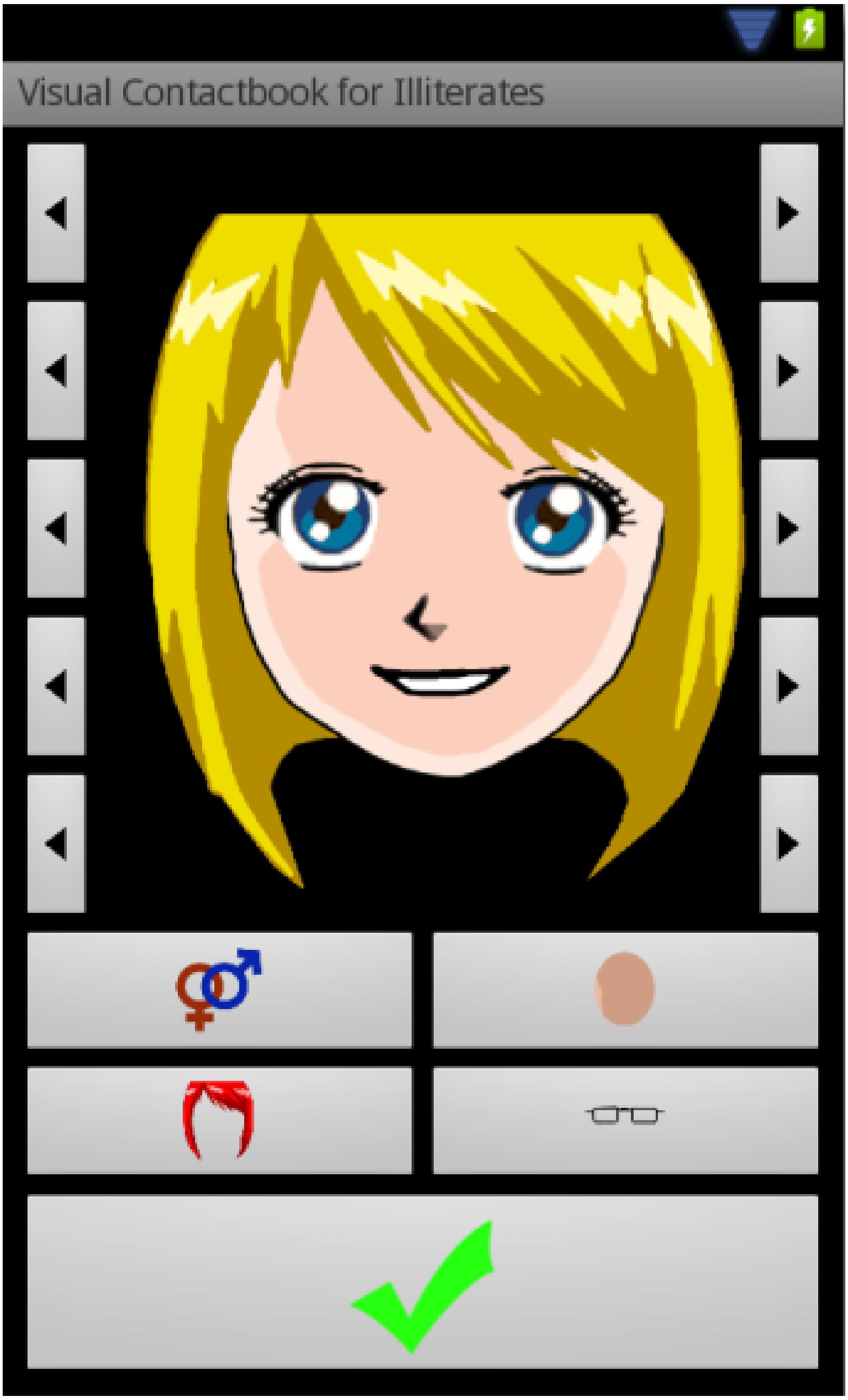


Figure 13: A screen-shot of the current version of the avatar customization screen.

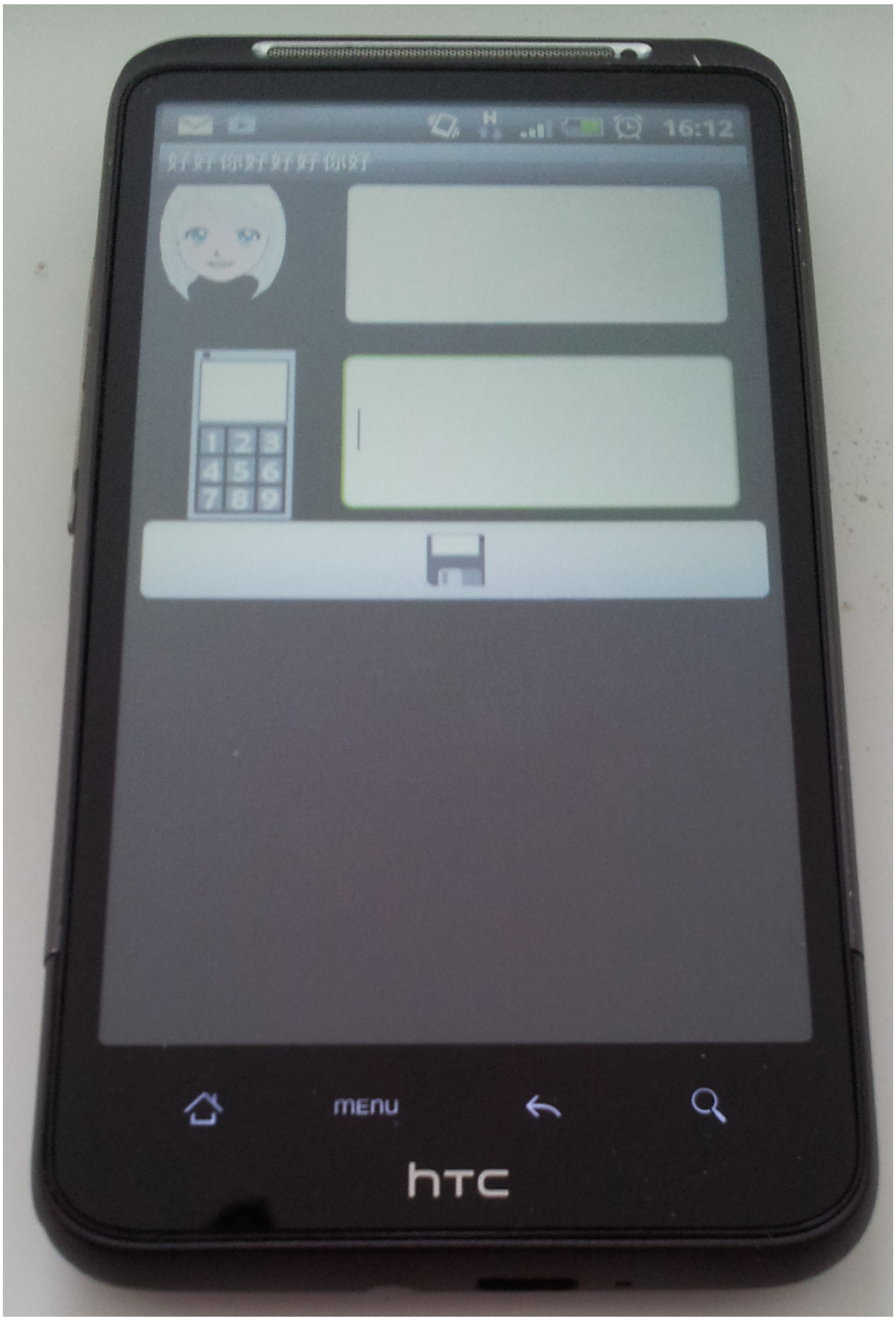


Figure 14: A screenshot of the screen for entering the name and number of the created contact.

cross. This cross will have to be changed in size as the result of the usability test, where it proved to easy to access by accident, deleting the contact. The current design of the screen is shown in Figure 15. The final screen of the application is the one that manages the search function, allowing one to search through the

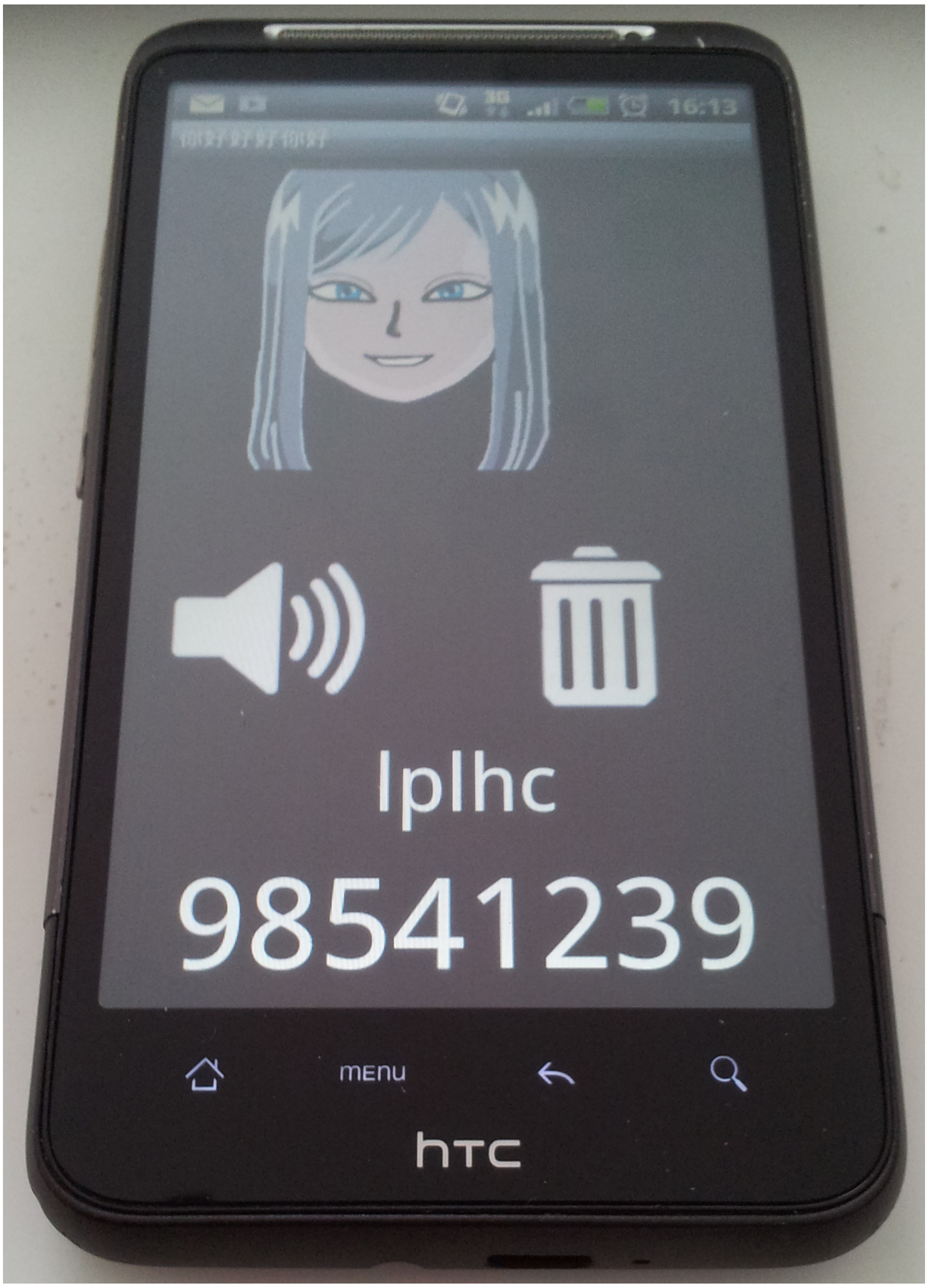


Figure 15: Screenshot of the screen containing avatar information.

database by using a variety of features. Just like the concept design it contains four features for searching such as gender, glasses, hair color and skin color. The were some changes to the icons as is displayed in Figure 16.

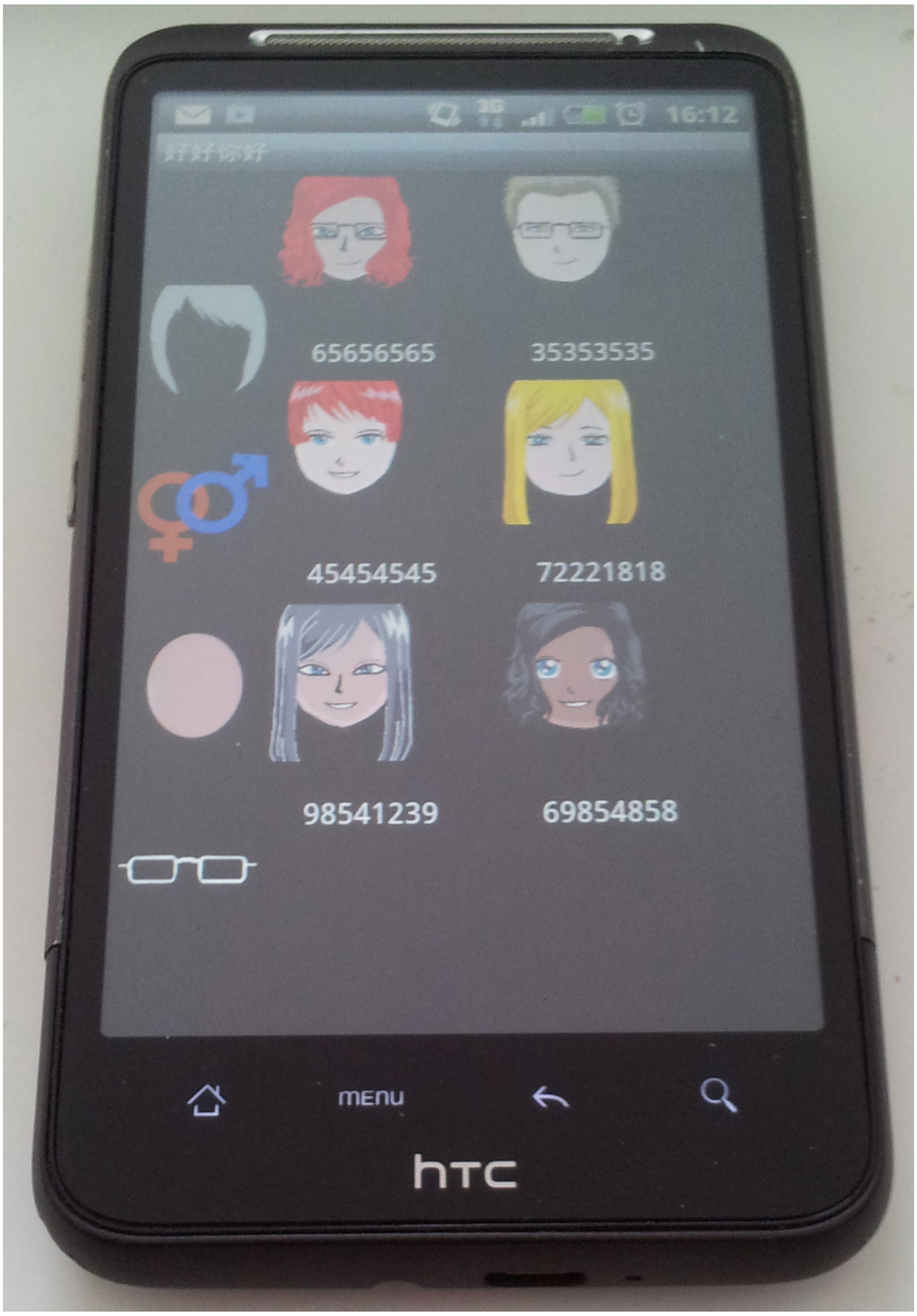


Figure 16: Screenshot of the screen used for searching the database of contacts.

4 Test Design and Results

4.1 Introduction

In this chapter there will be a detailing of how each test is structured and designed, and which results the tests gave. It has been split up as voice, avatar and user interface tests was conducted seperately.

4.2 Voice Recognition Test

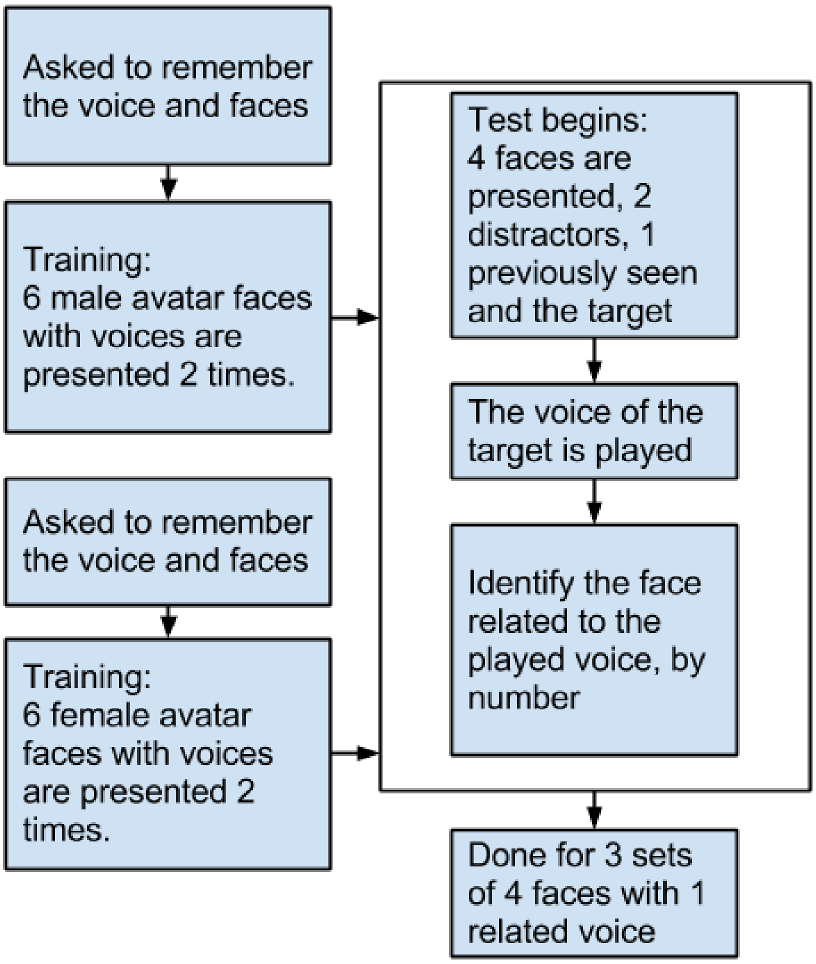
During this section a description of the two low fidelity test and high fidelity test for voice recognition is given. The first low fidelity test will serve to give a better understanding of identity matching between voice and face, using avatars for unfamiliar targets. The second low fidelity test is made in order to investigate the effect of sound quality on people's ability to identify voices . The high fidelity test is done in order to investigate the impact that the relationship between the participant and the target has on identification over time.

4.2.1 First Low Fidelity Test

During this section a description of first low fidelity tests is given, this test is concerned with voice recognition. The reason for performing this test is in order to find out how well voice recognition can be used in aiding illiterates navigating a contact book. The first test is regarding voice recognition, for this test a total of 16 subjects were tested all students from Aalborg University at the department of Medialogy, 3 women and 13 men. The test was design to test the participant's ability to recognize unfamiliar voices in relation to an avatar face; therefore this test was split into two. One section was concerned with voice-to-face recognition with 8 subjects and the other face-to-voice recognition with the remaining 8 subjects in order to test both aspects of recognizing a voice from an avatar. For the test a range of female and male avatar faces was created as well as a number of female and male voices was recorded saying the same sentence. These voices was recorded using a Samsung Galaxy S2, recording at 44 Khz, the environment for the recording was people's homes, giving the setting that would be the case for the application. The reason for keeping the sentence the same is in order to keep focus on the voice, by removing queues that might arise if people say different sentences. The sentence that is used is "Hvordan har din dag været?" which translates into "How have your day been?".

Test Setup

The test setup is made up of one person controlling the visual feedback to the test participant, using a laptop while sitting next to him, and another facilitator sitting across the table controls the audio feedback. Before each test starts the test participant is told that his objective is to remember the avatars and their voices as for later recall. For both the voice-to-face recognition and the face-to-voice recognition the test is divided into two sections one containing the female and one for the male. This test design is inspired by Joassin et. al. [7]. The test participant is first shown 6 female faces with voices, running through them twice in order for the participants to familiarize themselves with the avatars and their voices; the same is done before doing the test regarding the male avatars. The avatars are



Asked to remember the voice and faces

Training: 6 male avatar faces with voices are presented 2 times.

Asked to remember the voice and faces

Training: 6 female avatar faces with voices are presented 2 times.

Test begins: 4 faces are presented, 2 distractors, 1 previously seen and the target

The voice of the target is played

Identify the face related to the played voice, by number

Done for 3 sets of 4 faces with 1 related voice

Figure 17: Test setup for face-to-voice, dependent on the whether the participant starts with male or female avatars they begin training with the gender they are chosen to start with, this setup is the same for voice-to-face but revers with 4 voices and 1 face.

named with a number instead of a name; this is done to eliminate any relation the participant might have to that name, which could help him/her remember. For the face-to-voice they are shown 3 sets of 4 faces for male and 3 sets of 4 faces of female, two of the 4 are distracters that they haven't seen and the other two is hence the target and another of the 6 avatars shown in the beginning. They then have to assign the correct face to the one voice that is played. They only hear the voice once. For the voice-to-face the same setup is used but giving them 4 voices per set, where they have to recognize the correct voice that belongs to the one avatar face that they are shown per set. For both face-to-voice and voice-to-face there are 4 subjects that start with the male targets first and the last 4 subjects start with the female targets. The order by which these sets are organized is randomized in order to rule out any pattern that might become visible for the participant. A visualization of the test setup can be seen in Figure 17. In Figure 18 a simple sketch is given of how the test setup was done. As the facilitator sitting next to the test participant was showing the visual feedback in form of the avatars, the avatars was organized in files using unrelated names preventing them from relating names between the first shown targets and the sets of 4 avatars in the case of face- to-voice recognition. These files are arranged in the test order so that the two facilitators had a simple pattern to follow. Each time the test participant utters the avatar or voice that he/she claims to be the correct one, the result is noted by the facilitator sitting across the table in an excel ark. The entire test setup and results can be found in an excel ark on the attached DVD.

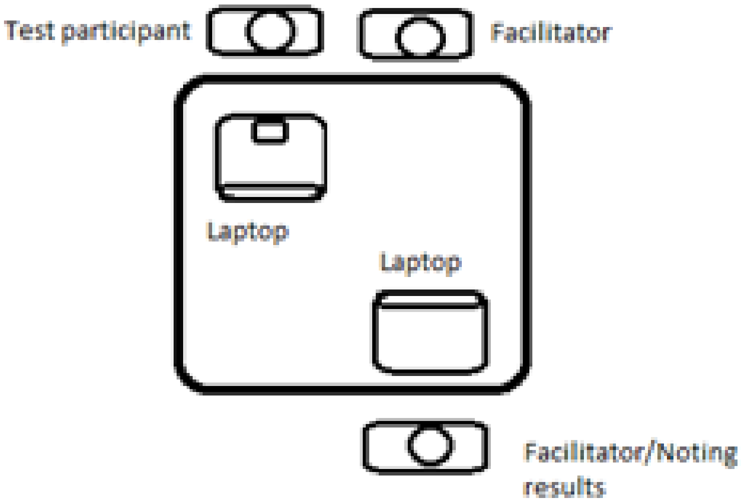


Figure 18: Physical test setup.

Test Results: Voice to Face Recognition

In these diagrams, Figures 19, 20 and 21, the "wrong-distractor" means that the test participant has made a wrong recall by identifying a distractor. A "wrong-tar" means that the participant has identified a previously seen/heard avatar, one of the 6 avatars shown with in female and male.

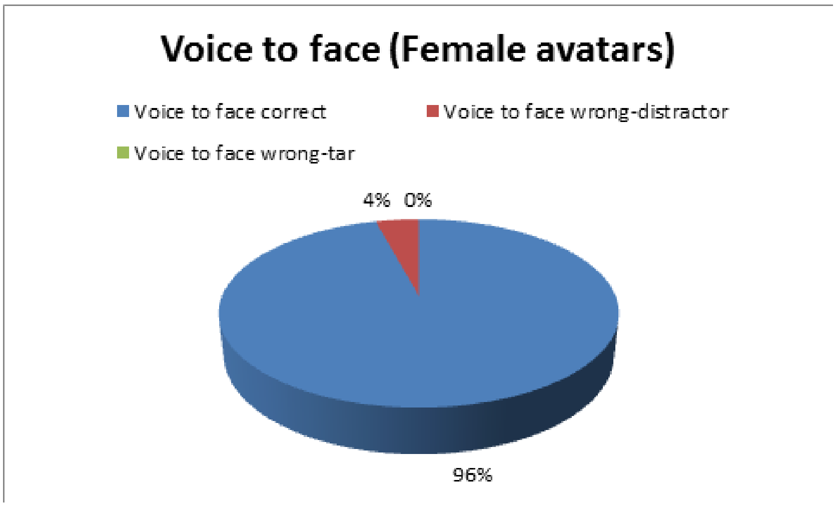


Figure 19: Female avatars: voice to face. (23 correct, 1 wrong-distractor, wrong-tar 0)

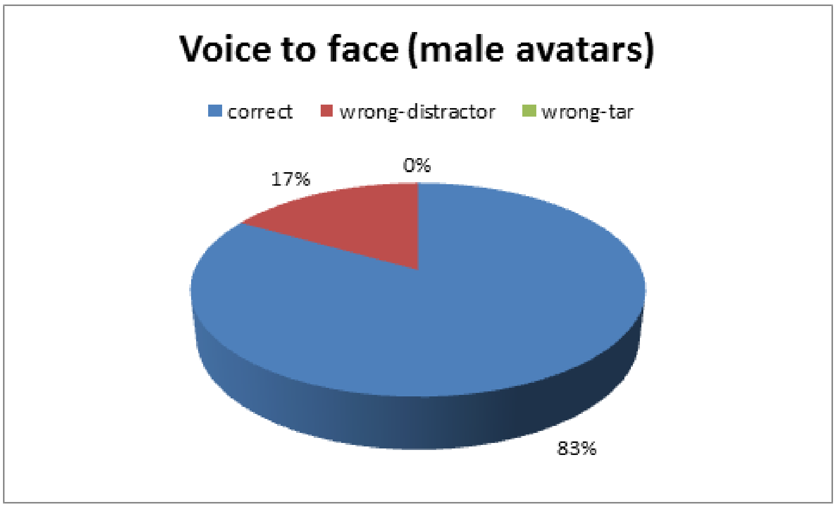


Figure 20: Male avatars: voice to face. (20 correct, 4 wrong-distractors, wrong-tar 0)

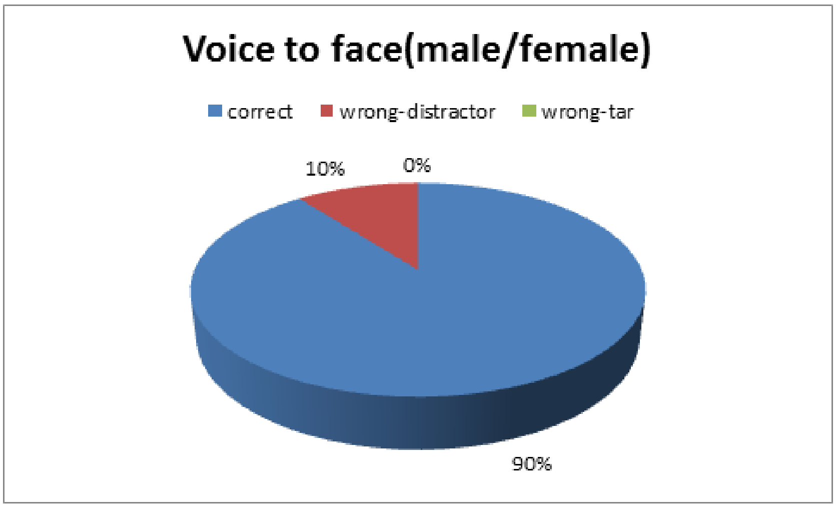


Figure 21: Female/male avatars: voice to face. (43 correct, 5 wrong-distractors, wrong-tar 0)

Test Results: Face to Voice Recognition

In these diagrams, Figures 22, 23 and 24, the same applies as for the previous diagrams, except that this if for the face-to-voice test.

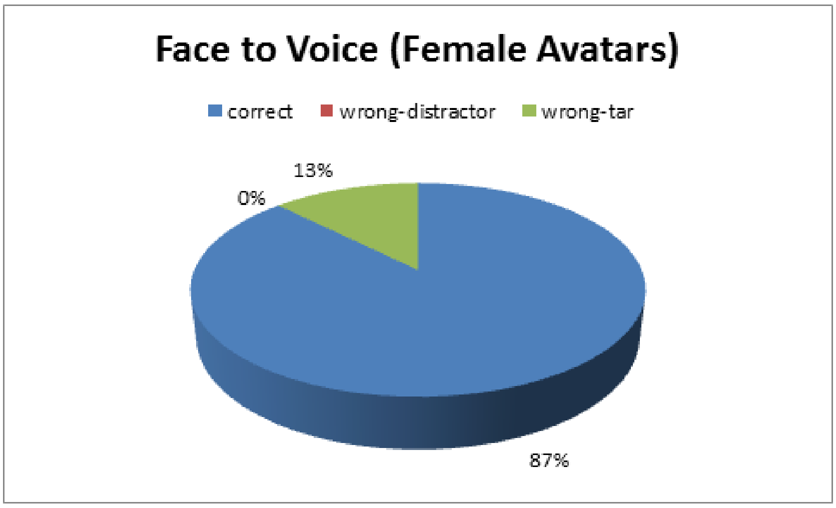


Figure 22: Female avatars: Face to voice. (21 correct, 0 wrong-distractor, 3 wrong-tar)

Part Conclusion

When looking at the data from the voice to face test and the face to voice test for

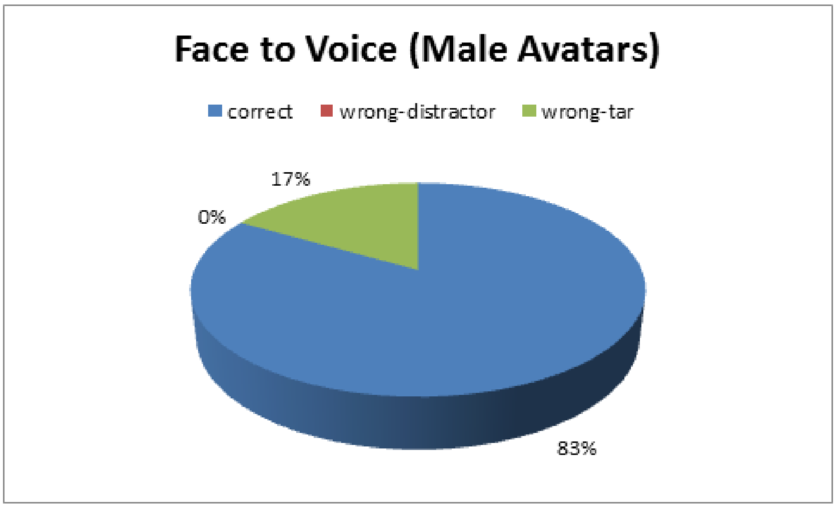


Figure 23: Male avatars: Face to voice. (20 correct, 0 wrong-distractor, 4 wrong-tar)

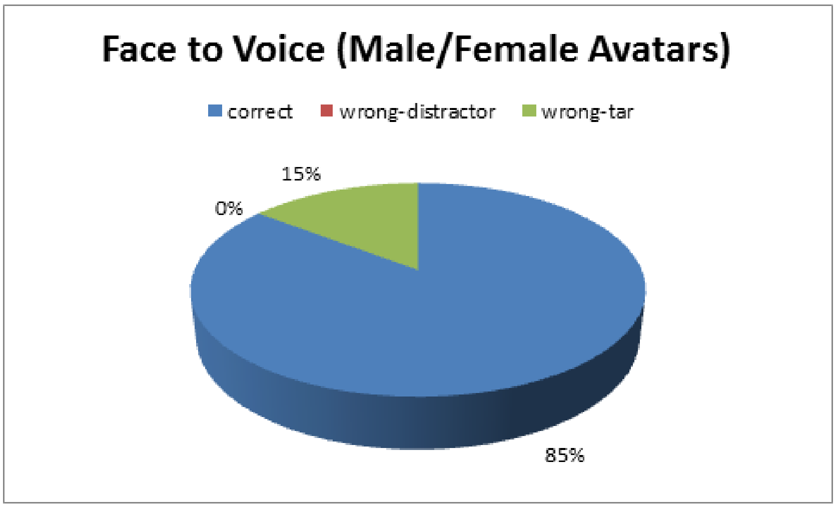


Figure 24: Female/male avatars: Face to voice. (41 correct, 0 wrong-distractor, 7 wrong-tar)

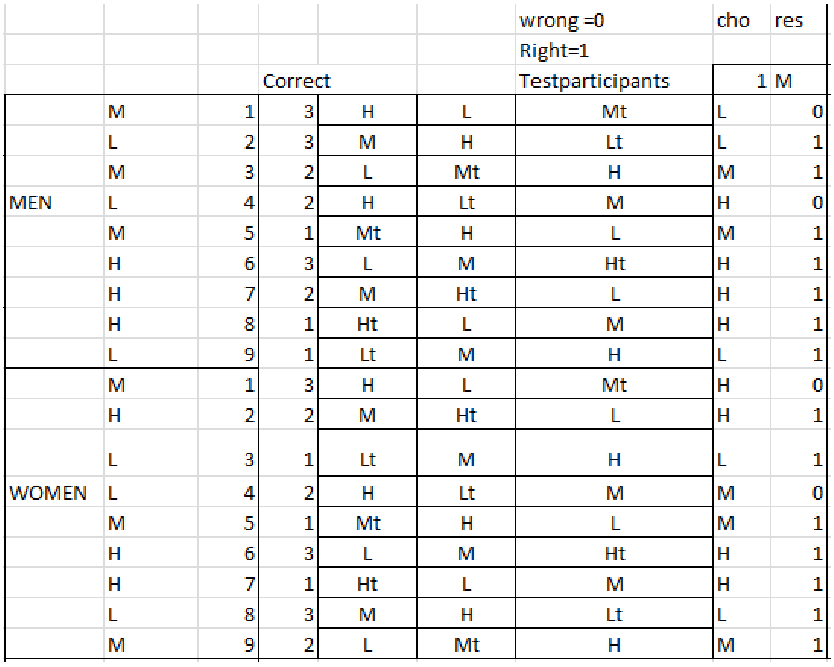
recognition of voices, an interesting pattern emerges. When the test participants in the voice to face test perform a wrong recognition they recall a distracter. On the other hand when participants in the face to voice test perform an error they recall a previously seen target which is not the target for the voice they were told to recall. Although there is this difference when performing errors, the general error rate is lower than 20% and lowest for the voice to face. The fact that the test participants in the face-to-voice test choose a previously seen target when choosing wrong could indicate that the information from the faces is easier recalled. In contrast, in the voice-to-face test, they never choose a previously heard target when failing to recall the target; this shows that the participants find it harder to recall voices than faces. Keeping this in mind, the voice-to-face had a lower error rate than the face-to-voice.

4.2.2 Second Low Fidelity Test

The focus of this test is to find how big the impact of sound quality is on people's ability to match identity from recorded unfamiliar voices. The reason for this is that not all phones purses the same sound quality when hearing the caller through the phone or on loudspeaker. The voices used in this test are the same as recorded for the first low fidelity test. As this test is for finding the impact of sound quality, the recorded samples have to be down sampled to a lower quality for the test. This is done through Audacity where the files are saved in an mp3 format at a lower sample rate. The three different qualities that is tested are 8, 24, 32 Kb/s with the original being 44 Kb/s from the Samsung galaxy S2. As the audio is taken from the first low fidelity test, the sentence spoken is the same "Hvordan har din dag været?".

Test Setup

In this test the test participants will have to identify different voices, played at different qualities. To eliminate the effect of different genders, the voices that will be identified are arranged in a group of male voices and one with female voices. The test is arranged as a within subject test, each subject will be exposed to each sound quality within both male and female voice, at each position in the sets. The table below in figure 25 shows the organization of the male and female sets with the three different qualities. As is seen in figure 25, there are 9 sets of male and female voices. For the training before the test they are familiarized with 6 voices of men and then females. These voices are not given a name but a number, just as in the first low fidelity test. For both men and female there are 3 targets out of the 6 voices, each played with the three sound qualities and with each quality at a different position in the order of play. Together all the male and female targets ensure that each quality at one point is played in each position. For a better visualization of the test procedure a flow chart is created, this can be seen in figure 4. Each test participant is trained with 6 female/male voices twice, within each set as shown in figure 25. In each set there is the target voice, a distractor and previously heard voice, which is one of the 6 voices that are not targets. As is seen in figure 26 the test participant is trained in first 6 male voices and then tested, then trained in the 6 female voices and tested with those. The physical test setup was with one person playing the sounds from a laptop and a facilitator that



						wrong =0	cho	res
						Right=1		
			Correct			Testparticipants	1	M
	M	1	3	H	L	Mt	L	0
	L	2	3	M	H	Lt	L	1
	M	3	2	L	Mt	H	M	1
MEN	L	4	2	H	Lt	M	H	0
	M	5	1	Mt	H	L	M	1
	H	6	3	L	M	Ht	H	1
	H	7	2	M	Ht	L	H	1
	H	8	1	Ht	L	M	H	1
	L	9	1	Lt	M	H	L	1
	M	1	3	H	L	Mt	H	0
	H	2	2	M	Ht	L	H	1
	L	3	1	Lt	M	H	L	1
WOMEN	L	4	2	H	Lt	M	M	0
	M	5	1	Mt	H	L	M	1
	H	6	3	L	M	Ht	H	1
	H	7	1	Ht	L	M	H	1
	L	8	3	M	H	Lt	L	1
	M	9	2	L	Mt	H	M	1

Figure 25: The organization of the sets for both male and female voices. The L, M and H is the different qualities of sound. L = 8 Kb/s, M = 24 Kb/s, H = 32 Kb/s

noted down the answers of the test participant as well as any expressions by the participant.

Test results

For this test a total of 10 participants were tested, all male. To get an understanding of the effect of quality, a point for each correct answer is awarded. All point for each quality is summed up for a value reflecting each qualities success rate.

- High (H) = 51 correct (85%)
- Medium (M) = 46 correct (76.7%)
- Low (L) = 42 correct (70%)

The reason for the relative small difference might be due to characteristics of the voices used, but this reflects what you would encounter in the real world as people talk with different amplitude, punctuation and more. It has not been possible to find the cut off quality where it is no longer recognizable, but a graph depicting the tendency is shown in figure 27. As is seen in the graph in Figure 27, it is possible to see a downward tendency assuming that a quality of 0 Kb/s is equal to no recognition. Given the high percentage of success rate it can be said

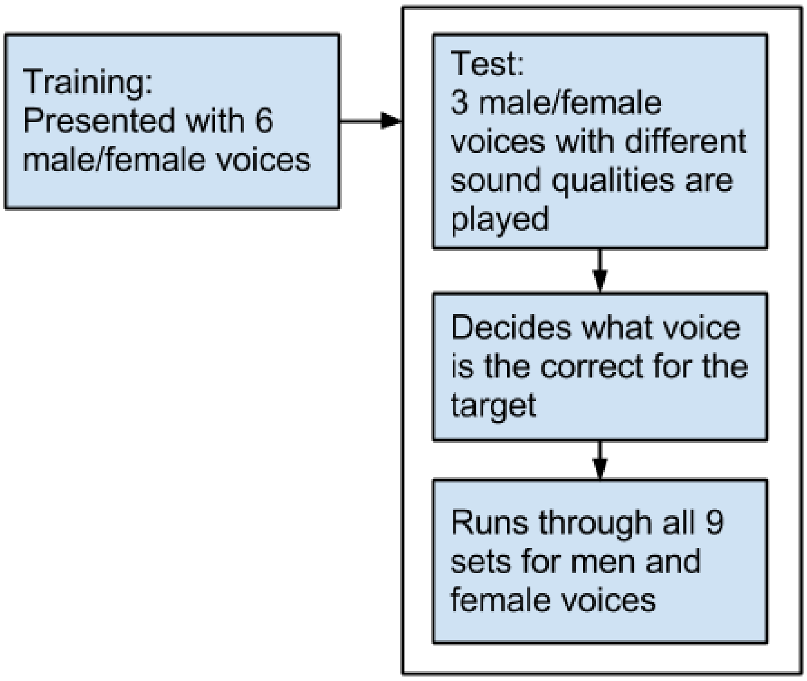

Training:
Presented with 6 male/female voices
Test:
3 male/female voices with different sound qualities are played
Decides what voice is the correct for the target
Runs through all 9 sets for men and female voices

Figure 26: A flow chart of the test procedure for the test participants

that it is possible to lower the sound quality to 8 Kb/s and still have a success rate of 70%.

4.2.3 High fidelity test

During this test, the hypothesis will be answered. The goal is to investigate the impact of the relationship between the participant and the voice they are trying to identify, over time. The hypothesis is displayed below.

- Null hypothesis: when matching identity from audio input there is no difference over time between participants that are familiar with the targets and those who are not.

- Alternate hypothesis: when matching identity from audio input there is a difference over time between participants that are familiar with the targets and those who are not.

The audio stimuli used in this test is retrieved in two different ways as one group of the test participants is tested with familiar voices and the other group with unfamiliar voices. The group with unfamiliar voices will hear the same voices that were recorded for the low fidelity tests, recorded with the Samsung Galaxy

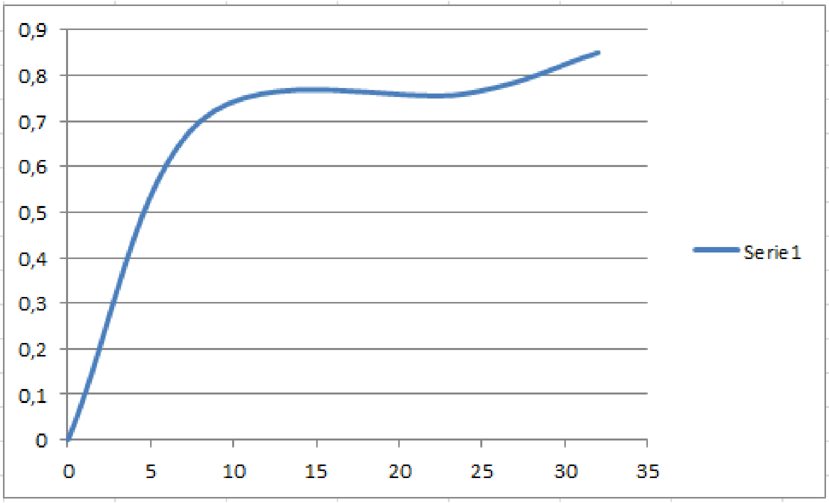


Figure 27: A graphical representation of the data from the three different qualities, the y axis is the percentage of correct answers, and the x axis is the quality of the sound. This graph assumes that a quality of 0 Kb/s equals no recognition

S2. The group with the familiar targets will have to call 4 friends or family over the phone. The phone is then put on speaker and the friend or family member will say the same sentence as for the unfamiliar that is "Hvordan har din dag været". This is then recorded on the Samsung Galaxy S2. By having them speak over the speaker we get the worst case scenario for sound quality as it is lower quality then the audio directly from the phone. This causes a difference in sound quality between the audio for the group with unfamiliar targets and those with familiar targets, which has to be taken into account.

Test setup

The test is divided into two; one group of the test participants is tested with unfamiliar targets and another with familiar targets. For both groups the test is a two part test, the first day the participants are familiarized with the targets through hearing all voices 5 times. Each participant is trained on 4 targets, 2 male and 2 female, familiar or unfamiliar. For familiar this was not always the case as it was not always possible to reach 2 male and female friends or family at the time of recording. Two days later the participant is tested for his/her ability to identify the voices. The test is a within subject test, each participant is run through 14 sets, 2 of them with no targets in. Each target is rotated around so it is placed at all positions at one point. Each set contains three voices 1 which is the target and two distractor voices. Each time the test participant has been presented with the 3 voices, he/she will tell at what location the person is that they recognize and that person's name or number. The group with familiar will use the name of the person and the group with unfamiliar target the number of the person. A visualization of the test design is displayed below in figure 28. When recording the voices for the group with familiar targets, it was not always possible to attain 2 female and 2 male voices of their friends or family. This meant that some test participant had only men and others a mixture. The physical test setup was placed in a room at Aalborg University. One person handled playing the audio in the correct order according to the sets, and the other noted the results while taking notes of occasional remarks from the test participant. After identifying the voices in the 14 sets, the test participant is asked to answer a questionnaire as displayed below. The two first questions are related to the group with familiar targets.

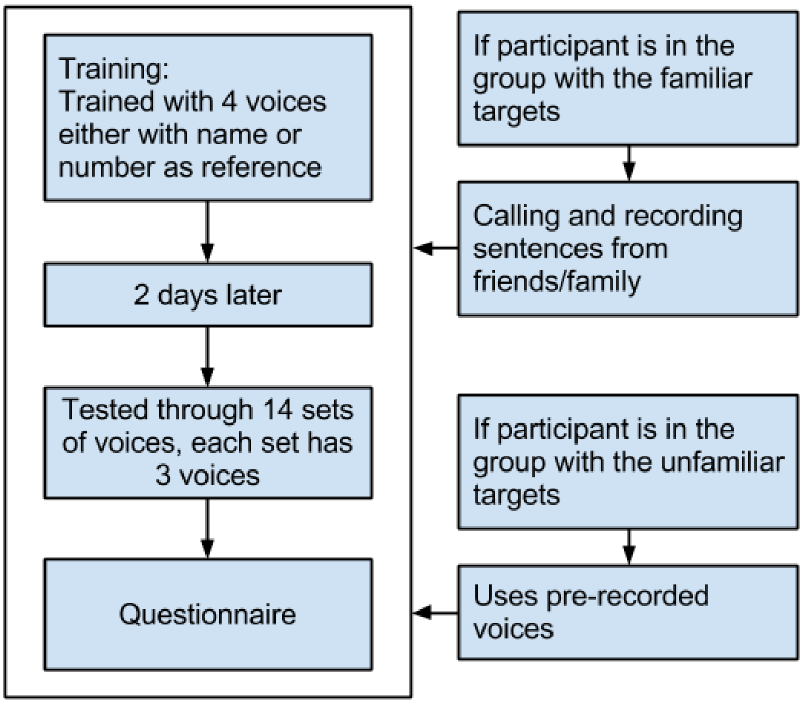

Training:
Trained with 4 voices either with name or number as reference
2 days later
Tested through 14 sets of voices, each set has 3 voices
Questionnaire
If participant is in the group with the familiar targets
Calling and recording sentences from friends/family
If participant is in the group with the unfamiliar targets
Uses pre-recorded voices

Figure 28: A flow chart of the test design for the high fidelity test.

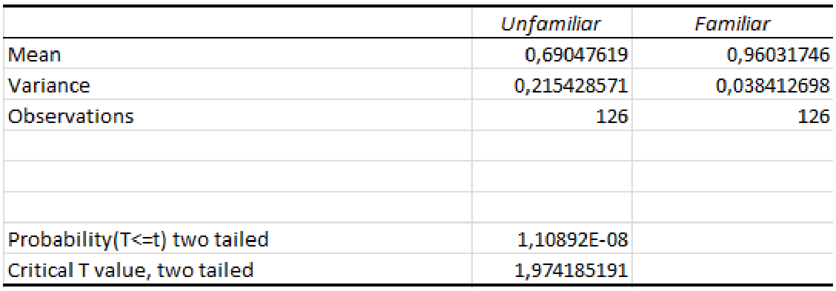
- Name:
- Have you talked with any of the subjects from the first part of the test since it was recorded?
- If you have, how often and how long?
- What helped you recognize the voices?

- What proved hard to recognize and why?

These questions are asked to get a understanding if the participants in the group with familiars has talked with the targets in the two day period and if how often, as it might not be possible for them not to talk to them. The remaining two questions are asked to try and figure out what they attend to when identifying the target. The questionnaires and excel arks containing the test setup and results are possible to find on the attached DVD.

Test results and discussion

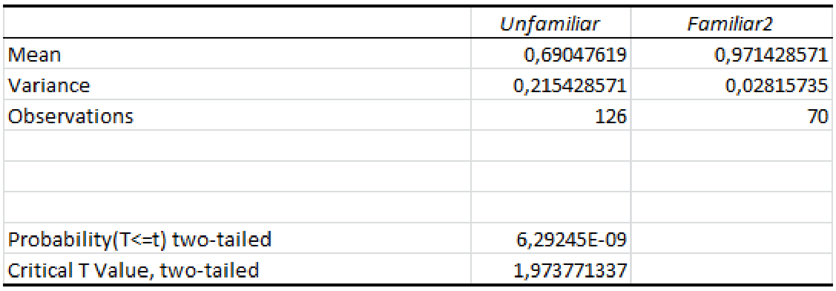
To verify or reject the null hypothesis a T-test is performed on the data. The T-test is a two tailed T-test with one group of data from the participants with unfamiliar targets and the other with participants with familiar targets. The result of the T-test is displayed below in figure 29. As is seen in figure 29 the p value is below 0.05 meaning that the null hypothesis is rejected, thereby accepting the alternate hypothesis. This shows that there is a significant difference between the two groups, showing that the test participants are significant better at identifying the voices from familiars than unfamiliar.



	Unfamiliar	*Familiar*
Mean	0,69047619	0,96031746
Variance	0,215428571	0,038412698
Observations	126	126
Probability(T<=t) two tailed	1,10892E-08	
Critical T value, two tailed	1,974185191	

Figure 29: The data from the T-test.

By looking at the questionnaires it is possible to see that one of person from the group tested on familiar targets did not hear the targets voice in the two day period. This might improve the difference between the two samples. At the same time the voice recorded for the group using familiars, was of poor quality compared to the quality of the unfamiliar group. The recording of the voice from the speaker of a phone meant that the quality was low compared to recorded audio direct from the human voice. But as they were able to recognize the voices even though the lower quality, strengthens the result of the familiar group being better at identifying the voices. The one person that did not talk with any of his friends and family recorded for the test had a success rate of 100%. To get a better description of the relationship between the two samples, familiar and unfamiliar and the effect the two days had, another t-test is performed only using the targets that the test participants had not talked with over the two day period. The result of this t-test is shown in Figure 30.



	Unfamiliar	*Familiar2*
Mean	0,69047619	0,971428571
Variance	0,215428571	0,02815735
Observations	126	70
Probability(T<=t) two-tailed	6,29245E-09	
Critical T Value, two-tailed	1,973771337	

Figure 30: The t-test results when only using the targets not talked to during the two days period

As is seen in Figure 30, the probability is even smaller when using only the targets that were not spoken to during the two day period. This confirms that the conclusion made using all the data also applies for the reduced data with only targets not talked to. This shows that there is a significant difference between familiar and unfamiliar showing that the familiar are better at recognizing the targets over a three days period.

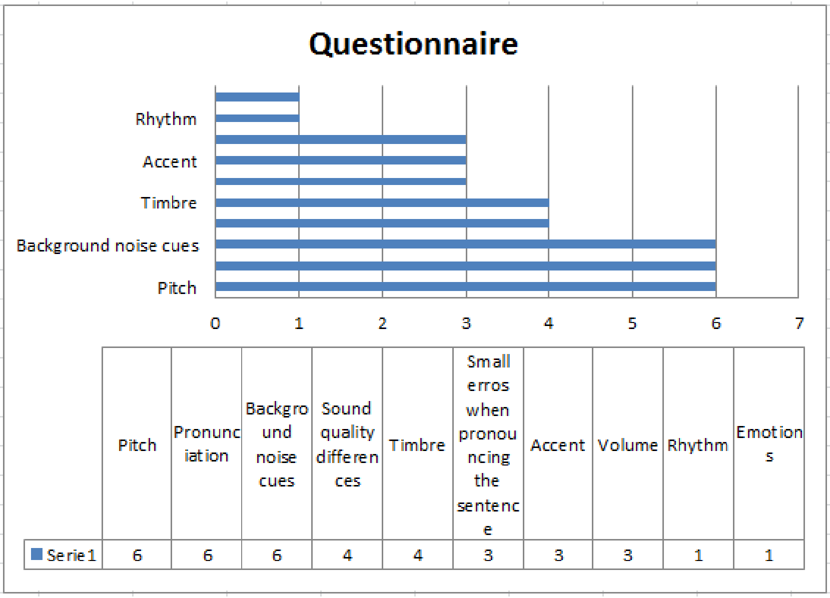
As is concluded above, there is a significant difference between the group with familiar and unfamiliar targets, proving that the group with familiars performed significantly better. To get a better overview of the precision within both groups, the success rate in percentage is displayed below.

- Group with familiar targets: 97.24 percent
- Group with unfamiliar targets: 69.04 percent

Even though the group with unfamiliar did not know the targets they still performed above chance at a success rate of 69.04 percent. With familiar targets the success rate is 97.24 percent giving an almost 100 percent accuracy, showing that this function is usable as a feature for recognizing contacts on a phone.

From the questionnaire a number of different features were given that helped them recognize the voices when tested. What helps the test participants according to the questionnaire is the Pitch, Emotions, Background noise cues, accent, Volume, pronunciation, small errors when pronouncing the sentence, rhythm and quality differences. The features that the test participants found most helpful was the pitch, pronunciation, background cues. To better visualize the results from the questionnaire, a diagram is created showing how important the feature is. The diagram in figure 31, is a result of counting how often each feature is mentioned when answering "What helped you recognize the voices?". As is seen in figure

31, the most important features are pitch, pronunciation and background noise cues. The fact that there was a difference in quality can have biased the results,



	Pitch	Pronunciation	Background noise cues	Sound quality differences	Timbre	Small erros when pronouncing the sentence	Accent	Volume	Rhythm	Emotions
■ Serie1	6	6	6	4	4	3	3	3	1	1

Figure 31: Diagram of the occurances of the different features when answering the question "What helped you recognize the voices?"

both in the sense that it could prove harder to recognize the voice and that it would be easier to eliminate the distractors.

4.3 Avatar Recognition Test

This chapter explains the design and method of the two tests which used visual avatars. The first test was used to determine if the approach from face recognition known as holistic perception also worked with virtual cartoon characters faces. The second test was about recalling avatars presented in three different ways over a two day period of time. In the end of this chapter a discussion will explain the results from the tests.

4.3.1 Low Fidelity Test

There have been a lot of research regarding how humans perceive and understand a face but not if this research also works for virtually created avatars which only has features that resembles human features. This first test was a test to see if feature-specific perception or holistic perception would be strongest with virtually created avatars. The design of the test follows that of Belle et. al. [2] in which the participants was shown a face masked at certain locations to enable feature-specific perception or enable holistic perception. An example of this can be seen in figure 32. Then in a mix of distractors and previously shown faces the participants would have a very short amount of time to decide which of the faces they had seen before, if any. Our test used this test as a starting point by also having the participant seeing 6 avatar faces in a series which the participant then had to remember, an overview of the test setup can be seen in figure 33. The series was shown 3 times in which each avatar had around a 2 second period. After this remembering part the participant was shown another series of 6 avatars masked accordingly to either feature-specific recognition, see figure 34, or holistic perception recognition, see figure 35. Each participant would receive 12 of these series in which

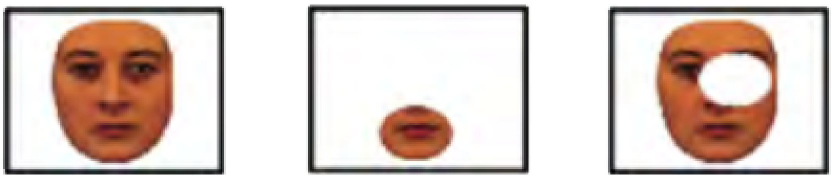


Figure 32: Example taken from [2]

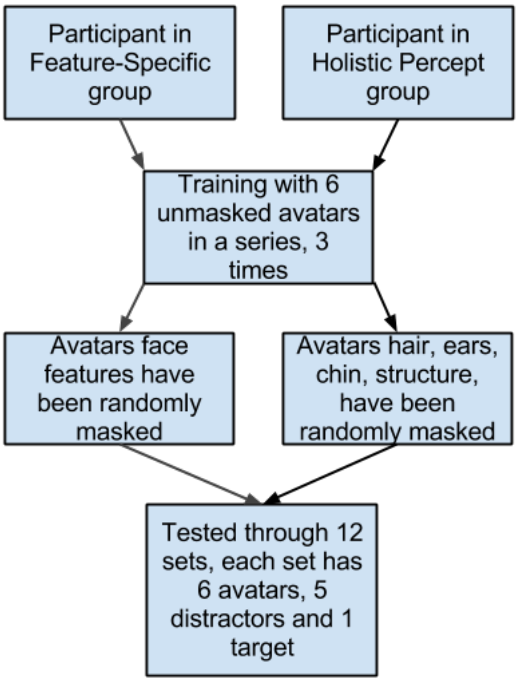

Participant in Feature-Specific group
Participant in Holistic Percept group
Training with 6 unmasked avatars in a series, 3 times
Avatars face features have been randomly masked
Avatars hair, ears, chin, structure, have been randomly masked
Tested through 12 sets, each set has 6 avatars, 5 distractors and 1 target

Figure 33: An overview of the test setup of the avatar low fidelity test

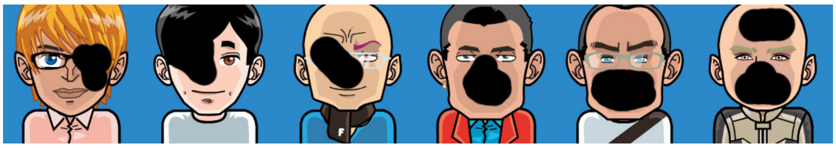


Figure 34: A sequence from a participant who was shown feature-specific masked avatars.

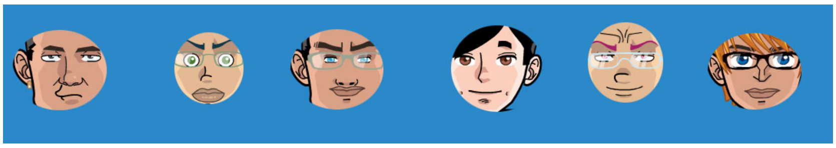


Figure 35: A sequence from a participant who was shown avatars masked to fit a holistic approach to face recognition.

they would have to find one of the previous remembered avatars in a mix of previous seen avatars and distractor avatars which they had never seen. The tester would ask the participant to find e.g. Avatar 1 from the remembered set, making this a recall task in which they had to remember each avatar by a number and the avatar visual features. In total 8 participants had the feature-specific, 4 had male while the other 4 had female avatars, and 8 participants had the holistic approach, half and half on both male and female avatars again. In figure 36 a brief overview of the test setup can be seen. The hypothesis for this test was to find out if the holistic percept is more important in avatar recognition than the feature-specific approach. The results rejects the hypothesis as the participants who were tested with the holistic approach did not have any significant differences compared to the ones tested with the feature-specific approach, the correct answers per group were used as data and the P value for the test is 0.053. A summarization of the numbers used can be seen in table 2. Further information about results and notes during the test can be found in appendix "Facial Recognition.xlsl"which contains the raw data from the test.

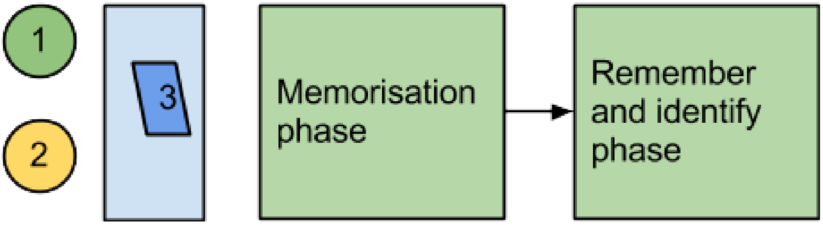


Figure 36: 1) Test facilitator changed images on screen. 2) Participant. 3) Screen.

Write the correct "link" to the file and name

Feature-Specific		Holistic Approach	
Correct	Failed	Correct	Failed
75	21	85	11
Mean Correct	Percent Correct	Mean Correct	Percent Correct
9.25	78.1%	10.125	88.5%
P-Value			0.053

Table 2: Correct and failed answers from the two groups.

4.3.2 High Fidelity Test

This test was designed to see if there is a difference in the ability to remember an avatar over time if you had a certain level of familiarity with the subject you made an avatar for. The test included three different groups, a control group, an unfamiliar group and a familiar group. The control group was supplied with pre-made avatars from the application the other groups would use to generate and customize their avatars with. Each participant would be presented with two female and male avatars which they would have to remember over a period of two days at which point they would have to point out which avatar they remembered, much like the first test except here they would not be asked to find a certain avatar but just tell if they found one they remembered and the number associated with it. This time it was 16 sets of four avatars that included distractors, multiple correct targets and one set that only had distractors in it. In figure 37 an example of what participants in the control group would have to remember can be seen. In the unfamiliar group the participants was supplied with photos of two women and two men which they would have to make an avatar from and remember them over the same time period as the control group. This group also had 16 sets of four avatars that included distractors, multiple correct targets and one set that only had distractors in it. This group was used to see if the act of creating an avatar for an

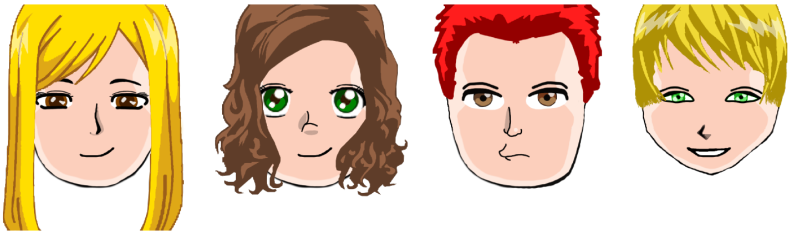


Figure 37: An example of what participants in the control group would have to remember.

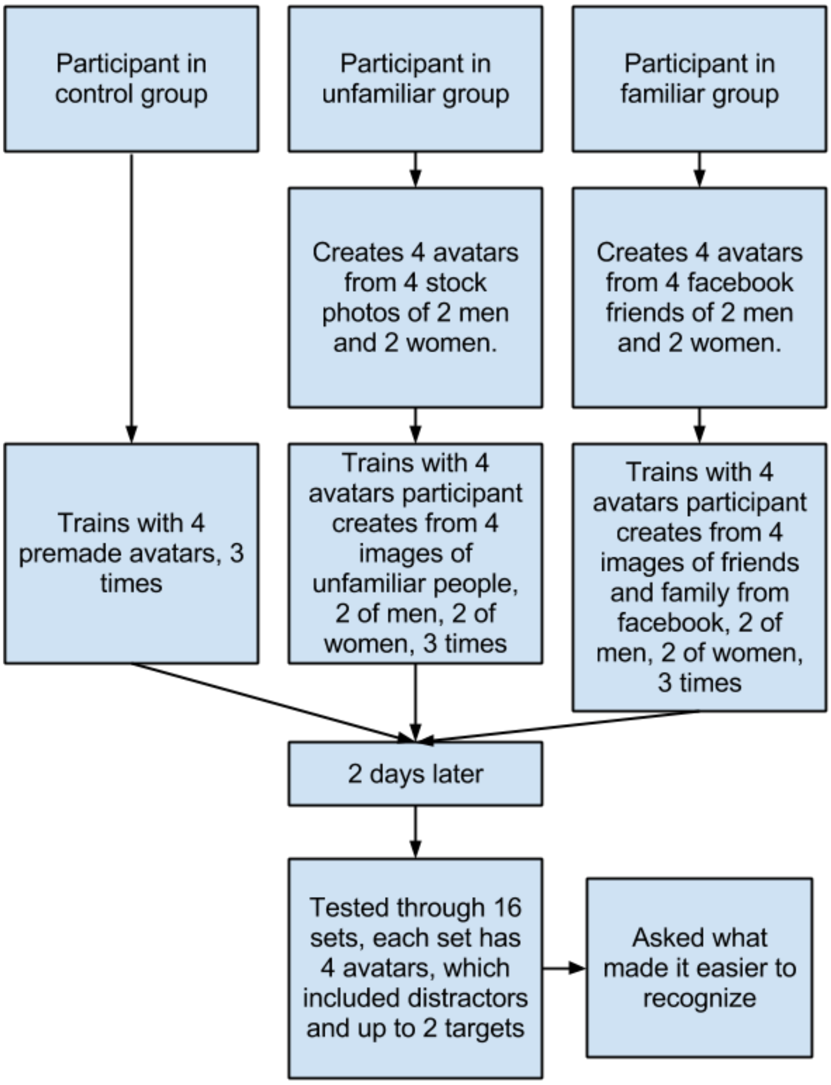

Participant in control group
Participant in unfamiliar group
Participant in familiar group
Creates 4 avatars from 4 stock photos of 2 men and 2 women.
Creates 4 avatars from 4 facebook friends of 2 men and 2 women.
Trains with 4 premade avatars, 3 times
Trains with 4 avatars participant creates from 4 images of unfamiliar people, 2 of men, 2 of women, 3 times
Trains with 4 avatars participant creates from 4 images of friends and family from facebook, 2 of men, 2 of women, 3 times
2 days later
Tested through 16 sets, each set has 4 avatars, which included distractors and up to 2 targets
Asked what made it easier to recognize

Figure 38: An overview of the test setup of the avatar high fidelity test

unfamiliar person would increase the accuracy of the memory or if it would be the same as the control group. In the familiar group it was the same procedure as the unfamiliar group, except the photos would be of two female and two male friends of the participant which, together with the participant, was found on facebook. The participant was instructed to create an avatar that resembled the person as they remembered the person and not as they looked on the profile picture on facebook. The name, picture and avatar was recorded as they would have to recall name and avatar at the later stage. This group was used to determine if a person was familiar, we categorized familiar as a good friend or family, it would be easier to remember and as such the accuracy would be higher than the other groups. An overview of the test setup for this test can be seen in figure 38.

As such we end up with three hypotheses for this test:

1. The unfamiliar group has a higher accuracy than the control group.

2. The familiar group has a higher accuracy than the control group.

3. The familiar group has a higher accuracy than the unfamiliar group.

The results from the test can be summarized as seen in Table 3. As seen each hypothesis has some support by a difference of at least 10% in correct targets which is also shown to be significantly different through a t-test as shown in Table 4. The concept behind correct position is if the participant found the correct position of the target but could not remember which avatar corresponded. This was clearly a problem with the control group in which one subject switched around all the avatars but still got 18 positions correct. Adding the correct positions together with the correct targets delivers a new set of values in which the control group actually scored higher than the unfamiliar group. This could also support that when you create the avatars yourself you are in a better position to remember the associated name or number, because you associate the picture or person with the avatar. A graph showing the correct answers by group is shown in Figure 39.

	Control Group	Unfamiliar Group	Familiar Group
Participants	8	8	8
Total Targets	160	160	160
Correct Targets	111	127	145
Correct Position	28	1	0
Percent Correct	69.38%	79.38%	90.63%
Percent Correct Position	86.88%	80.00%	90.63%

Table 3: Summarization of results from the second test.

After each run through of the sets each participant was also asked which features made it easier to remember the avatars and remember the differences between each avatar. Out of the 24 participants 20 used hair color and style to differentiate while eyes, chin/face build, mouth, glasses and nose had 5-6 participants using them to differentiate. This separates hair as the primary source that helps separate the avatars from each other.

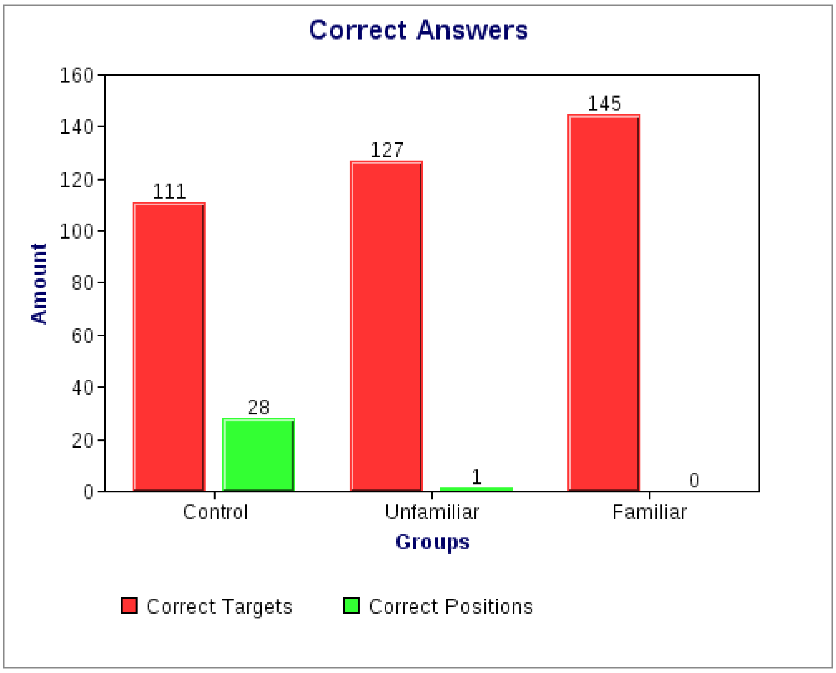


Figure 39: A graph showing correct answers through groups.

T-Test Groups	Control Group	Unfamiliar Group	Familiar Group
Mean Correct Answers	13.88	15.88	18.13
Mean	0.693	0.793	0.906
Variance	0.213	0.164	0.085
P-Value Control-Unfamiliar			0.0406
P-Value Control-Familiar			1.55788E-06
P-Value Unfamiliar-Familiar			0.0047

Table 4: T-Test for the second test.

4.3.3 Discussion

The tests supports that feature-specific perception and holistic perception also works with avatars as a means of understanding and remembering the avatar. In the test some participants also noted that they thought the avatar felt something or looked like they were in a certain mood, giving social meaning to a virtual face. This could be because humans are hardwired to see and understand this deeper social meaning from facial features as shown in the chapter describing facial recognition theory. Because of this, the participants in the second test who had to make their own avatars also gave each avatar an expression that would show a certain state of mind as they perceived that person. Multiple participants also noted that they could remember through different mouth types, ones that looked happy, satisfied or disagreeable. In the second test the results show that a higher percentage accuracy can be retrieved when the participants use familiar subjects to create the avatars from. A reason for this could be that they use more time to create the avatars, but so does the unfamiliar group. Multiple participants that created their familiar persons was also noted to be laughing when creating the avatars at certain points because of the avatar designs as some of them was unexpected in the application. This, as a form of play, could increase the persons accuracy in remembering which the test results supports. So adding play in the form of avatar customization increases accuracy over a two day period compared to a control group that did not get to play with the avatar customization. Adding the concept of using familiar subjects further increases the accuracy, giving increased support to the idea of having a contact book that builds on the idea of having avatar customization.

4.4 High Fidelity Test of Interface

The contact book created in this project is oriented towards illiterates, trying to deal with the problems that illiterates have when trying to use the contact book of a mobile. The purpose of this test is to investigate if the interface is working as intended; this is done with an usability test. The target group of this project is illiterates, but due to the circumstances it was not possible to find illiterate people to test the interface on. Instead a number of literate people are used for the test to test the usability of icons and functions. This can help adjust potential errors that have been made. The fact that the test participants are not illiterate might result in lag of important input, as illiterates have certain tactics they apply when dealing with contact books of phones [1]. To simulate the situation that illiterates are in for the literate test participants, text is changed into Chinese symbols, forcing them to rely on the icons.

4.4.1 Test setup

The usability test is divided into two sections, in the first the participant has to carry out a number of assignments and last answer a short questionnaire. When the test participant enters they will be explained the process, that they first have to carry out a number of tasks on the app, followed by a short situated interview. Beside the test participant there is one facilitator in control of the tasks and another facilitator noting the test participant's actions during the tasks. When the participant is given the tasks he/she is told to think aloud, telling what they will do in order to follow their thought process and actions. After completing the tasks the test participant is asked a number of questions to follow up on the tasks they performed. Below in figure 40 is a visualization of the test design. The test is performed as an iterative process, after each test participant has been tested, the most prominent error is changed to accommodate the participants input. This is done to find as many flaws as possible. The tasks that the participants are given during the test, is displayed below.

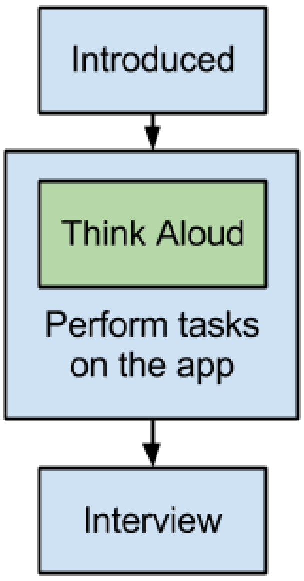


Figure 40: The structure of the usability test

- Create a new contact (Female, blond hair, long hair, green eyes, glasses) with the phone number 38651942.
- Locate the contact in the contact book (to check if they use the search function).
- Use the search function to locate the contact (and enter for information).
- Play the voice of the contact.
- Call the contact.
- Delete contact.

Before the test participants are given the app to perform the tasks on, the contact book has been filled with a number of contacts both male and female. This is done to simulate a more realistic use of the contact book already containing contacts. The tasks presented above are made in order to test the different aspects of the contact book with regard to functionality and icon design. Followed by the task is the interview, below is the questions asked in the interview.

- Gender.
- Age.
- What did you find difficult and why?
- What could be improved and how?
- What would you use the app for, in what situation or case?

The first two questions are asked to get information about the participant, the next two is asked to locate potential problems and suggestions for improvement. Lastly a question to get a more broad aspect of the app, even though the test participants are not illiterates it can help understand in what other situations this app could be useful.

4.5 Test Results

In the test there were three subjects that participated in the test, all students from Aalborg University. The test was carried out at Aalborg University. During this section each subjects test result is presented and the changes that was made in order to get a better view of the results.

4.5.1 Subject 1: Male, 25

Create a new contact (Female, blond hair, long hair, green eyes, glasses) with the phone number 38651942: During this task the subject found the button for adding a new avatar on his first try. When creating the avatar he plays around with the buttons on the side to figure out what part of the avatar it changes. For the last stage of the creating the avatar where the name and number has to be entered, he understands the symbols for them and enters the correct and saves using the disk icon.

Locate the contact in the contact book (to check if they use the search function): When presented with this assignment he chooses the button with an avatar and a loop over because it to him represent search.

Use the search function to locate the contact (and enter for information): First he just scrolls down through the contact book without using the search function, but when told to use the search function he starts the search by choosing glasses as a feature in the search. He miss understands the icon for gender for being something related to the hair. He enters the contact by pushing on the window containing the avatar.

Play the voice of the contact: Uses the correct sound icon as his first choice.

Call the contact: Pushes the phone number to call the contact on his first try.

Delete contact: Pushes the cross to delete the contact on first try.

4.5.2 Subject 2: Male, 24

Create a new contact (Female, blond hair, long hair, green eyes, glasses) with the phone number 38651942: When trying to create a new avatar he instead pushes one of the contacts and when trying to return he pushes the cross and thereby deleting the contact. On his next try he pushes the right button. When creating the avatar he utters that the icon for the glasses should be more clearly displayed.

Locate the contact in the contact book (to check if they use the search function): First the participants just look through the contact book, thinking he found the contact but was not the correct one. On his second attempt he uses the search function.

Use the search function to locate the contact (and enter for information): First he starts by scrolling down, instead of using the search features. When told to use them he uses them flawless.

Play the voice of the contact: Uses the speaker icon.

Call the contact: Uses the phone number as button.

Delete contact: Uses the cross button.

4.5.3 Subject 3: Male, 24

Create a new contact (Female, blond hair, long hair, green eyes, glasses) with the phone number 38651942: This subject has tried the creating section of the app before so performs the entire task correct when creating the avatar.

Locate the contact in the contact book (to check if they use the search function): He uses the button with the loop on as he utters, to search for the contact.

Use the search function to locate the contact (and enter for information): This subject is the first that uses the search features in first try, instead of scrolling as the two previous subjects.

Play the voice of the contact: Uses the correct button for the task.

Call the contact: Uses the correct button for the task.

Delete contact: Uses the correct button for the task.

4.5.4 Main Findings from Subjects

Even though it was not possible to test on illiterates, the test of the interface with literates proved to give some useful information about general changes in the design of the interface. For a more case relevant usability test, a test is needed with illiterates, in order to understand their needs and the effect of the strategies they apply when using technologies like mobile phones. In general subject 1 performed with few errors during the tasks, he understood the way of creating avatars well with no troubles at understanding the functionality and icons within that area. The only icon that he didn't understand was during the search task, which was the icon for gender that was thought to be related to something with hair. During the interview he told that some of the icons in the search was not logical, especially the one with the gender as was also found in the tasks. He suggested using the symbol for men and women instead. Beside the button with the gender, he would have liked if a cross had applied to the glass icon when the glasses was removed from the search. For the creation of the new avatar, he utters that a small green plus sign would had helped him recognize faster that it was the "add new avatar" button. Last he explained that the arrows used when switching among different features when creating the avatar, should be more outstanding. When thinking of cases where this way of ordering your contact could be used, he thinks of people with a huge network. The reason for this as he explains is that it might be difficult to remember all the names of who is who, so could be smart to search by visual appearance instead.

Contrary to subject 1, subject 2 does not get the first attempt right when trying to create an avatar and miss interpreting the cross as an exit button. Contrary to subject 1 he only says during the test that the glasses should be clearer when creating the avatar. Just like subject 1 he does not start out by using the search function when trying to locate the contact, but first in second attempts. But when used it was used flawless. The fact that both subjects do not locate the search features when searching for a contact indicates that the search function has to be more visually noticeable. He uses all the icons within the contact correct, due to the fact that he found the cross to delete a contact earlier. During the interview explains that just like subject 1 that the "create new avatar" was difficult to understand and that a green plus symbol would help him understand this. The reason why he used the Red Cross as a return button was due to his lack of experience with android. He explained that it would be easier for him to understand the delete icon, if it was visualized as a garbage can. This icon might prove to be more universal than the cross. Subject 3 performs perfectly in the creation of the avatar as he has tried it before and generally uses the correct icons for the tasks. He is also the first that uses the search function correct on first try. During the interview he expresses that it would be nice with an indicator that shows how far you are in the list of features. Beside this the skin color icon could be improved with having a face with color on instead. When navigating the interface of the contact, he explains that it would be better to have a smaller delete icon as he might accidentally hit it and deleting the contact as it is placed next to the sound button. The size of the delete button needs to be changed to reduce the risk of deleting a contact by accident. When asked of what other cases or situations this could be useful for, he explains that it might be useful for children that cannot read yet. This could be an interesting aspect to look into. As described above there are a number of features and icons that needs to be changed to further improve the contact book, keeping in mind that it needs to be further improved with regards to illiterates. In general the interface proved usable as the participants were able to navigate and use the functionalities with few errors. As any new interface there is always a learning curve that will cause some flaws when trying to use the functionality of the contact book.

4.5.5 Conclusion

Even though it was not possible to test on illiterates, the test of the interface with literates proved to give some useful information about general changes in the design of the interface. For a more case relevant usability test, a test is needed with illiterates, in order to understand their needs and the effect of the strategies they apply when using technologies like mobile phones.

References

[1] Syed Ishtiaque Ahmed, Maruf Zaber, and Shion Guha. Usage of the memory of mobile phones by illiterate people. *DEV'13*, 2013.

[2] Goedele Van Belle, Peter De Graef, Karl Verfaillie, Thomas Busigny, and Bruno Rossion. Whole not hole: Expert face recognition requires holistic perception. *Neuropsychologia*, 48:2620–2629, 2010.

[3] Salvatore Campanella and Pascal Belin. Integrating face and voice in person perception. *Trends in cognitive science*, 11(12), 2007.

[4] A. Castro-Caldas, K. M. Petersson, A. Reis, S. Stone-Elander, and M. Ingvar. The illiterate brain: Learning to read and write during childhood influences the functional organization of the adult brain. *Brain*, 121:1053–1063, 1998.

[5] Brian R. Clifford. Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior*, 4(4), 1980.

[6] Kristin Dew, Carin Fishel, and Muna Haddadin Apurva Dawale. Karaoke: An assistive alternative interface for illiterate mobile device users. *CHI 2013: Changing Perspectives*, 2013.

[7] Frédéric Joassin, Mauro Pesenti, Pierre Maurage, Emilie Verreckt, Raymond Bruyer, and Salvatore Campanella. Cross-modal interactions between human faces and voices involved in person recognition. *Elsevier Cortex*, 47:367–376, 2011.

[8] Robert G. Franklin Jr. and Reginald B. Adams Jr. What makes a face memorable? the relationship between face memory and emotional state reasoning. *Personality and Individual Differences*, 49:8–12, 2010.

[9] Patricia K. Kuhl. Who's talking? *Science AAAS*, 333(529):367–376, 2011.

[10] Zereh Lalji and Judith Good. Designing new technologies for illiterate populations: A study in mobile phone interface design. *Interacting with Computers*, 20:574–586, 2008.

[11] Karen Lander, Harold Hill, Miyuki Kamachi, and Eric Vatikiotis-Bateson. It's not what you say but the way you say it: Matching faces and voices. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4):905–914, 2007.

[12] Frances McGehee. The reliability of the identification of the human voice. *The Journal of General Psychology*, 17(2):249–271, 1937.

[13] Indrani Medhi. Building interfaces for the illiterate, 2010.

[14] Rochelle S. Newman and Shannon Evers. The effect of talker familiarity on stream segregation. *Journal of Phonetics*, 35:85–103, 2007.

[15] Yueting Sun, Xiaochao Gao, and Shihui Han. Sex differences in face gender recognition: An event-related potential study. *Brain Research*, 1327:69–76, 2010.

[16] A. Thatcher, S. Mahlangu, and C. Zimmerman. Accessibility of atms for the functionally illiterate through icon-based interfaces. *Behaviour & Information Technology*, 25(1):65–81, 2006.

[17] Steven G. Young, Kurt Hugenberg, Michael J. Bernstein, and Donald F. Sacco. Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, 16(2):116–142, 2012.