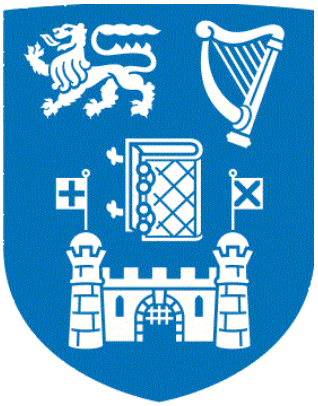
SpeechIsHard - A Serious Game in Aid of Speech Recognition

Brian Maguire

M.A.I.

Supervisor: Dr. Saturnino Luz



Trinity College Dublin

submitted to the University of Dublin, Trinity College,

May 21, 2015

Declaration

I, Brian Maguire, declare that the following dissertation, except where otherwise stated, is entirely my own work; that it has not previously been submitted as an exercise for a degree, either in Trinity College Dublin, or in any other University; and that the library may lend or copy it or any part thereof on request.

May 21, 2015

Brian Maguire

Summary

This project aimed to research, design and build a serious game that would aid in speech recognition research. The end product is SpeechIsHard, available on the Google Playstore. It is a two player gamification of a map tasks, a popular experiment used in speech research. The report outlines the research into speech recognition, mobile game design and serious games.
Research into speech recognition focused on areas where a serious game, or game with a purpose could be of help. The map task design was chosen due to the ease at which it could be converted into a game and the experiment's future use in the field, as a method of collecting realistic speech data. The report also looks into the designs used in the current popular mobile games. This research focuses on the aspects of the design which have been described as addictive. It was the aim of this research to pinpoint the key features of a mobile game that makes them widely popular and make use of it in the game design. The report includes a look at the field of serious games, or more generally gamification. It looks at what serious games are, and what their being used for today. The report outlines the final design chosen, as well as an earlier design. The designs were based on what was learned from the research undertaken. The final design of SpeechIsHard was then implemented. This report describes some of the technical challenges which the design presents and the solutions that were found for them. The finished implementation

was then evaluated by way of a survey. The survey's results showed that the game concept was an enjoyable one. The survey also highlighted several flaws. Among them was some issues with the games poor graphics and frustration caused by slow communication offered by the speech recognition. The survey also asked participants about their experience with speech recognition. One comment given in the survey provides interesting further work, of how people change their speech when communicating through a recognition system. The Report finally concludes by saying that in the aims of producing an enjoyable game concept, that might improve speech recognition technology through its play, this project has been some what successful. This comes with the caveat that SpeechIsHard's ability to function as a research tool is not evaluated. The project only evaluates SpeechIsHard on its merits as a game. The report goes on to suggest further work, including further development on the game to address some issues that came up during the evaluation, as well as some different fields in which SpeechIsHard, or a game like it may be of use.

Abstract

This report describes the design and build of a serious game to aid in speech recognition. It covers a brief review of speech recognition technology, how it works and how it might be improved by a game. The final game is available on the playstore under the name SpeechIsHard, and is a two player game that connects players over the Internet. It is loosely based on a gamification of a map task, an experiment used in speech recognition research. The game should allow for the collection of a corpus of speech data as people play. The report includes an evaluation survey on the enjoyability of the game. This evaluation suggests that the game concept has potential as an enjoyable game. SpeechIsHard has the advantage over normal map task experiments in that it provides a much greater scale of use.

Acknowledgements

I would like to thank Dr. Saturnino Luz for his help and guidence on this project. A special thanks to those that gave their time and took part in the survey. I would also like to thank my family and friends for their love and support.

Contents

List of Figures

Chapter 1

Introduction

In the following chapter I will introduce the project, by describing some of the background and the motivation for choosing this subject. I will then go on to give an overview of what is to be expected throughout the rest of the report.

1.1 Background

1.1.1 Treadris

This project is based on the serious game Treadris. The aim of the game is to correct results produced from a speech recognition system. The results were shown to the player as word lattices that would start at the top of the screen and move steadily back and towards the bottom of the screen. If not corrected by the time it reached the end of the screen, the word lattice would stay on screen, reducing the amount of time for each subsequent sentence. The player corrections which were likely accurate could then be used to adjust the speech recognition model so that it might not make similar errors in future. The game could also be used as a check for automatic transcriptions of video

or podcasts. To progress the ideas started by Treadris, I chose to develop a mobile game for android smart devices that would continue the work of Treadris of being a game for a purpose, that improved speech recognition.

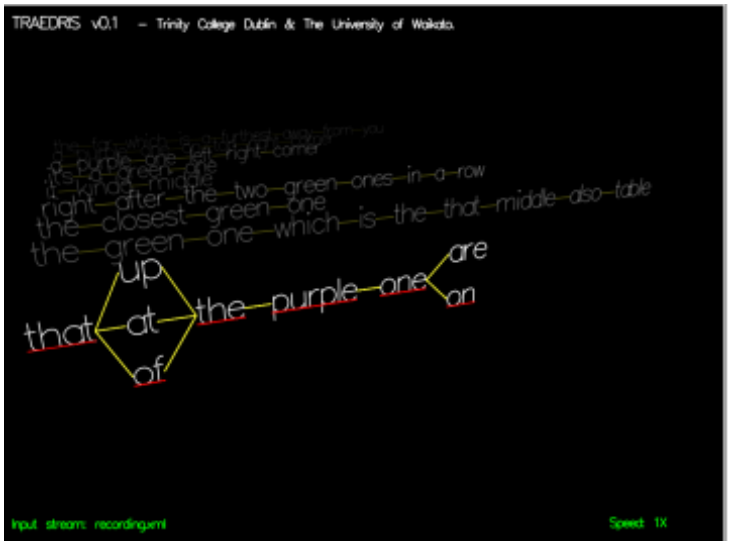


Figure 1.1: A screen shot from the original Treadris game (20)

1.1.2 Mobile Gaming

I chose to develop a game for the mobile platform in particular. The mobile gaming market has been one of the fastest growing games markets. It has introduced a new type of gaming. Mobile games have taken off with the spread of affordable smart devices. More and more people take a potential gaming machine with them in their pocket as they go about their day. Mobile games are developed to take advantage of the short but many moments of boredom people experience throughout their day. It is my aim with this project to put some of the hours used in playing mobile games to productive use.

1.1.3 Speech Recognition

Speech recognition has become an important technology for most people. Google Now, Siri, Microsoft's Cortana are all dependant on quick accurate speech recognition. With the advance of wearable technology, such as smart watches, the need for oral interface has become even more important due to their small screens.

1.2 Outline

Here I will give a brief outline of what is discussed throughout this report.

- **Research**- This chapter goes over the research that was undertaken for this project. It starts by giving a very brief description of how speech recognition is done today, and how a game might improve on this. It then give an overview of what Gamification or serious games are and what they are being used for today. Finally it gives a detailed look at some of the current top mobile games on the market today, with particular interest in the physiological techniques they implement to keep their audiences engaged.

- **Ethics**- This chapter examines some of the ethical questions which this research encounters. In particular the use of some of the addictive qualities in mobile games, as well as the use and storage of players information for research.

- **Design**- This chapter illustrates the early designs, their pros and cons, before describing the final game design and the reasons that they were chosen.

- **Implementation**- This chapter describes the technical challenges with which my chosen design comes, as well as some of the possible solutions to these challenges. This includes the solution that was used and why. The implementation is broken into three main sections, these are communication, speech recognition and the 3d Game.

- **Evaluation**- This chapter describes the method chosen to evaluate the game, a survey of test players.

- **Discussion**- This chapter discusses the significance of the results that were collected in the survey and highlights the most interesting feedback that was received.

- **Conclusions**- In this final chapter an overview of the project in terms of meeting its goals and also some possible areas of future work are explained.

Chapter 2

Research

This section describes the research that I undertook for this project. It includes a look at the current speech recognition technology, to give some insights as to where this game could be useful in the field. I introduce Map tasks as used in speech recognition, and why they are useful. I then look at research into serious games, games that attempt to do useful work, be it education or research while users play them. Finally I review the current mobile gaming market, and look for the successful design features that might be used in the final design.

2.1 Speech Recognition

In this section I will give a brief overview of how speech recognition is achieved at present and some of the work currently being done to improve it.
Automatic speech recognition has been brought into normal life recently by services like Apple's Siri or Google Now. People are more accustomed to speaking into their phone or computer, and they have come to expect accurate results from such encounters.

Conceptually, speech is understood through the perception of definite sounds, called phones. The combination of phones make up words, which consequently can produce sentences. The reality is that all speech contains elements of probability. Rather than constituting well defined sound frequencies, phones are loose context dependent patterns. Similarly, words are rarely separable in a sentence spoken at a normal pace. Speech recognition is a process of running sound waves through statistical models to return the most likely meaning. The modern recognition engine can roughly be separated into 3 steps.

2.1.1 The Speech Recognition Process

Speech recognition can be broken down into three distinct processes.

- First is feature extraction. Taking the continuous sound signal over a short time period and coding it into a feature vector.
- The second process is to take this feature vector through an acoustic model, which determines what phone or triphone the feature might represent.
- The third process is the language model. This takes the possible sounds, and puts together coherent sentences, based on the grammar laws of a language, and the statistical frequency of words used.

Feature Extraction

The continuous sound signal needs to be converted into a discrete feature, to be used by the following models. The extraction is done by taking sound signal over a short period of time, generally 10 ms, and converting it into a

vector of about 40 dimensions. This vector contains information about the frequency and amplitude of the sound signal, weighted slightly to match the frequencies best heard by the human ear, (10).

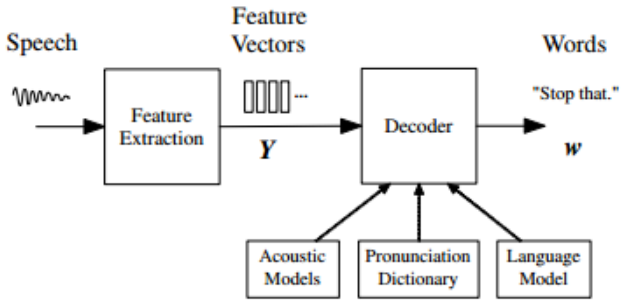


Figure 2.1: A HMM based speech recogniser from (10)

Acoustic Model

A word is made up of base phones. Phones make up the sounds of letters and there are 40 phones in the English language. An acoustic model fits the feature vectors of speech into their most probable phones. These phones are also taken in threes as the context of a phone is often most important in recognition. The phone is therefore considered along with the phone before and after, together as a triphone. It is the acoustic models job to take the feature vectors and return the most probable triphones. Each triphone is modelled using a Hidden Markov Model (HMM). The HMM models the probabilities of state transitions, from one triphone to another. It is hidden, as the spoken phone is not known to the system, only the feature vector which has been recorded for that phone. If each phone is a state in the HMM they produce a certain sound wave with some probability density(24). These models are produced using training data. The training data is made up of sample audio of known text being spoken. This produces the state transition

probabilities and the probability densities of each phone. This training data is produced by collecting a corpus of known speech and recordings of people reading out sentences.

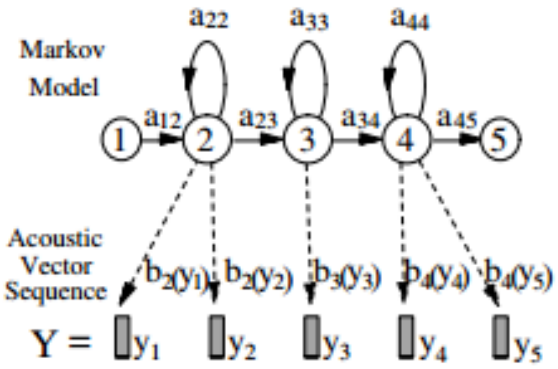


Figure 2.2: Each Phone has a HMM which produces feature vectors (10)

Language Model

The language model takes the resulting most probable phones and restricts them to those probable in language. The language model is generally made up of n-gram probabilities. These are the probability of any word following after n previous words. They are constructed from analysing written text and are the same processes used to produce auto completing search terms on search engines such as Google. The language model keeps the results as real pronounceable words. The most probable words can then be displayed as an n-best list, of possible words or as a word lattice where the translated sentence branches into different possible collections of words.

2.1.2 Speech Recognition Evaluation

The process of speech recognition is inherently probabilistic. People are not expected to hear and understand every word spoken to them, so it is understandable that speech recognition systems will always contain some error. Error is normally measured in Word Error Rate (WER), and Sentence Error Rate SER. The WER is calculated as $WER = (I + D + S)/N$ where N is the actual number of words, I is the number of words added, D is the number of words deleted and S is the words which were altered(5). SER is simply the number of sentences which contained errors, compared to the total number of sentences. The error rate expected from a commercial system can vary greatly, depending on the range of the vocabulary being used and the type of speech(4). With the vocabulary restricted to names in a contact list, letters or numbers, a recognition system can be expected to provide 100 percent accuracy. The rates reported with wider vocabularies using test speech can be as high as 98 percent (4). However, when tested on spontaneous speech, such as normal dialogueue of two people, the accuracy in WER has been recorded at 55 percent(4). Word Error rate may not be the best method of evaluation, to evaluate the recognisers on their ability to transfer information, rather than correctly translate individual words. In (21) Morris describes that other than transcription purposes, a method which measures "the proportion of information communicated" rather than WER would be a more appropriate evaluation. Morris explains some of the difficulties in measuring the information lost, particularly when the speech is heavily context dependent. It is difficult to automatically measure how much information was transmitted in a transcribed sentence just by looking at the words. It is helpful to have another indicator for whether or not a statement was understood.

2.1.3 Map Tasks

Map Tasks are a research technique which can be used to collect genuine speech dialogue, as well as providing an ulterior method of evaluation. A map task is made up of two groups in separate rooms who can only communicate through the use of a speech recognition system. The first group has a map, with labels and features, but no route. The second group has the same map but with a route to follow added. The first group must reproduce, or follow that route, taking instructions from the second group through the speech recognition system. The maps may have some differences, or objects may be missing or labelled differently on the two maps. These differences should force the instructions to be as descriptive and illustrative as possible, and prevents the dialogue from becoming generic directions. The features on the map can be chosen to guide the vocabulary of the dialogue. The use of a rabbit as one of the landmarks will ensure the use of the word rabbit. With this, words which sound similar can be used together as a difficult test for the speech recognition system or to collect a corpus of speech with chosen dialogue.

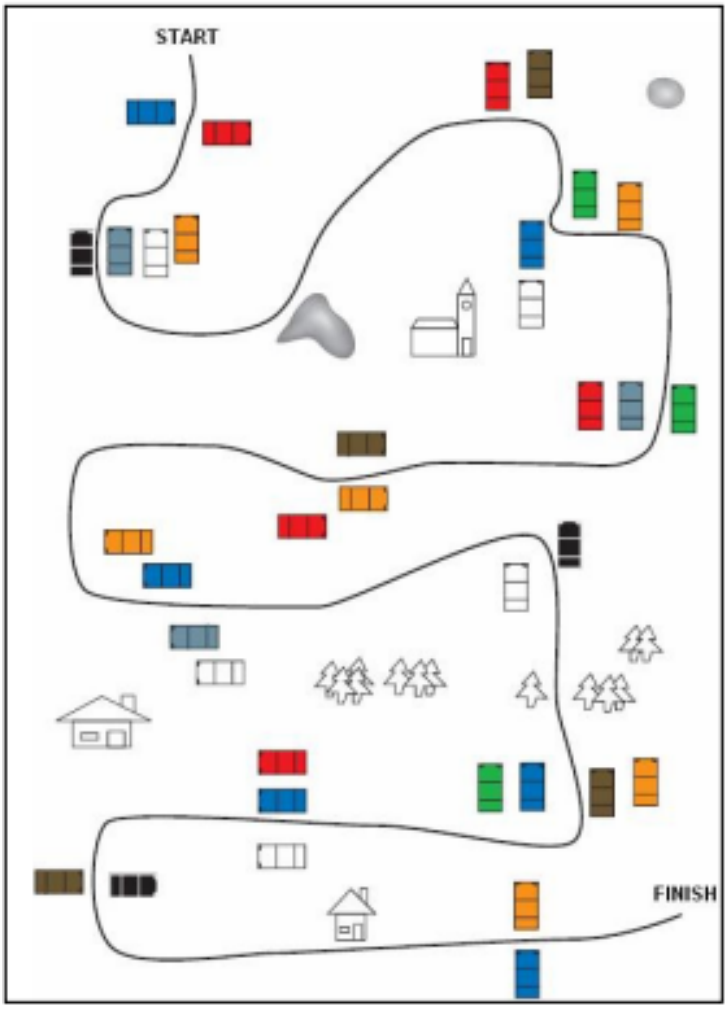


Figure 2.3: The Map used in (19)

In (2) Map tasks are used to collect a corpus of spontaneous speech dialogue that can be used to train speech recognisers, and evaluate existing models on more realistic data. In (25) Map tasks are used as a alternative evaluation technique. The progress of participants through the map and the number of corrective instructions are used as one of the methods of evaluating the recogniser's performance. In (25) this evaluation is compared to that of the normal WER and SER metrics, along with some more advanced

techniques. In (26) map tasks were used to study the error recovering strategies when speech recognition is used. The study looked at the ways people would recover from an error in the recognition, and how their language would change due to the unreliable communication channel. Map tasks can also be used in language translation. In a current study (1) Akira Hayakawa is using a map task experiment to collect a corpus of dialogue between participants speaking in different languages. The map task is modified to include an automatic translation system. This corpus will be used to assess human reactions to an automated speech-to-speech translation system.

2.1.4 Possible Areas of Work

After my review of speech recognition technology I have highlighted some areas in which a serious game might be of use.

- Transcription corrections - Similar to how Treadris functioned. If a game can be made in which players correct speech recognition errors, these corrections can be used to update the acoustic and language models of that speech recognition system.

- Collecting Speech Data - A game could automatically collect a speech corpus. As part of voxforge, a project aiming to create a collection of open source speech corpus on which to train acoustic models. A game could be designed which, through its play, collected audio of the players speaking a selection of words, which could then be uploaded automatically to Voxforge.

- A Gamification of Map Tasks - A mobile game version of Map Tasks, which could collect dialogue and evaluate speech recognition systems.

Map tasks can be laborious, requiring a good deal of time from participants. If the process could be turned into a game that kept even a small player-ship, speech corpus might be collected at a scale that is usually not possible with a normal map task experiment.

2.2 Gamification & Serious Games

Serious games, or games with a purpose, are games such as the original Treadris, which perform useful work through their playing. This has been used to tag pictures or to transcribe speech. It is part of a wider movement of Gamification. "Gamification is an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience and user engagement"(6). The original Treadris game is an example of gamification, turning transcription into a game by adding a scoring system, and a time limit. Another example of this is HabitRPG. This is a game designed to

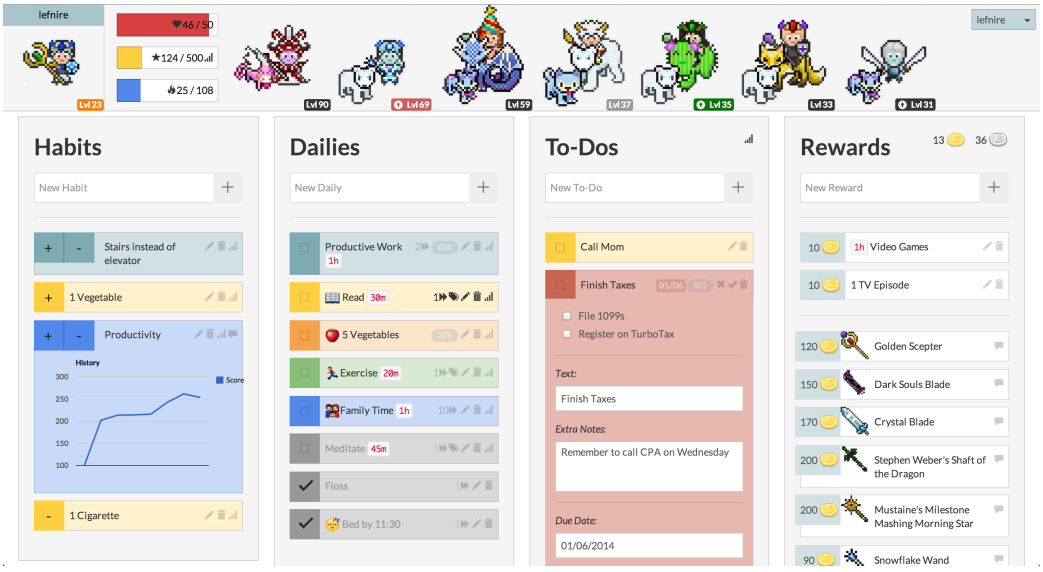


Figure 2.4: Screen shot of HabitRPG showing To-Do tasks as characters (12)

help build positive habits and remove negative behaviour. In the words of its creator "It "gamifies" your life by turning all your tasks (habits, dailies, and to-dos) into little monsters you have to conquer."(12). This game is just one of many in a growing trend of porting game elements into life improvement or other work. Hackerranker is a website that provides coding challenges. It also makes use of game elements such a leaderboard and badges which are awarded after overcoming certain challenges. The hopes of gamification are that the game features, which were developed to keep the investment of their players, can be transferred to real world problems. Hackerrank can be used by potential employers, allowing them to seek out coders who have excelled at relevant challenge. Users on Hackerrank are performing coding tests, but rather than seeing them as a work task, they are being done for enjoyment.

2.3 Game Design and Mobile Gaming

2.3.1 Top Current Mobile Games

As part of my research I looked at the current leaders in the mobile games market. It is my intention to identify the features that make these games so popular in this competitive market.

I looked at the top grossing games on the Google play store to give a list of the games that have got the most downloads and continue to draw players. The top chart list gives preference to games that are gaining popularity fast. This is not something in which I am interested, as it can depend on current trends. I am most interested in the style of game that remains popular.

In the top 20 grossing games, there are three that carry the names of popular movies or TV shows, such as "The Simpsons: Tapped Out" at number 11. These are also not of great interest as its impossible to tell how much of their

popularity comes from the brand that they carry. In a similar fashion, there are at least 4 games in the top 20 that are mobile versions of board games, or otherwise popular computer games. These include "8 Ball Pool" at number 5, or "Zynga Poker Texas Holdem" at number 17(14). These games are also of no real interest, as for this project I am looking for features that make a mobile game popular, these games are popular regardless of their platform. Of real interest to this project are games that have become popular through their mobile use; games that don't seem to have traction on other platforms. These are the games which rely most on mobile specific features.

One name comes up three times in the top 10, "Candy Crush Saga". The original game is ranked as number two, and a spin off game "Candy Crush Soda Saga" is at number 4, followed by "Farm Heroes Saga" at number 7. The basic game mechanic for all three is essentially identical. rearranging coloured objects on screen to get 4 of a kind together. When 4 objects, be they candy or soda, are placed together, they explode into colourful coins and objects on top of them take their place. It is a very simple puzzle, put together with a narrative, taking the player from level to level. Each level requires a higher score, while each level is essentially the same puzzle system.

Candy Crush Saga

Candy Crush Saga has between 100,000,000 and 500,000,000 downloads from the Google Play Store (14). It makes an estimated 1,000,000 dollars per day(11). Candy Crush Saga is clearly doing something right. The puzzle itself seems too simple and boring to gain this response on its own merits. I looked further into the techniques that this game uses to keep its players

Figure 2.5: Candy Crush Saga Game Play

playing and paying money.

Candy Crush Saga (CCS) is free to download and to play. The game gives you 5 lives which you lose if you fail to pass a level. Each life takes 30 minutes to replenish. You have the option of inviting friends to play the game to earn more lives, or you may pay for new lives with money. If a player chooses not to pay money or bother friends, the game is essentially stopping them from playing. Intuitively this seems like a terrible idea as most games try to keep a user's attention for as long as possible. It seems completely counter intuitive to block players out of the game for 30 minutes at a time. This behaviour is actually a potent physiological tool that helps the game keep players coming back every 30 minutes.

2.3.2 Hedonic Adaption

Hedonic Adaption is the process in which experiences be they positive or negative, dull over time so that an individuals' happiness will stay relatively

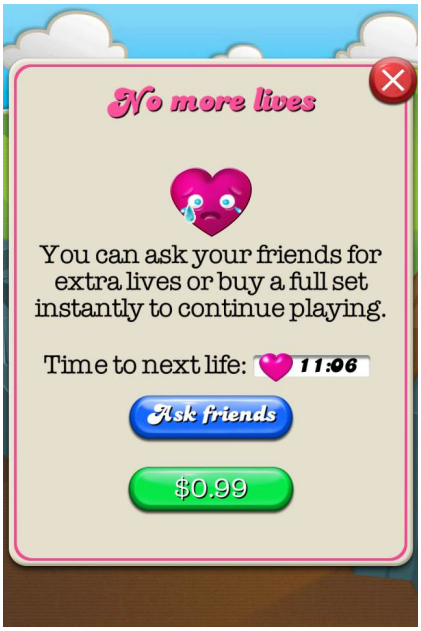


Figure 2.6: CCS forces players to pay for extra lives or wait before continuing play

steady, despite changes in their life. This effect normally means that enjoyable activities become less enjoyable over time and unfavourable experiences will become more bearable with repeated encounters (7). This effect is particularly relevant to simple games, which reward players for passing game stages. The enjoyment a player gets from playing a game such as CCS should inevitably decrease, unless the player is forced to stop before Hedonic Adaption can occur.

In (Quoidbach 2013)(23) researchers experimented with chocolate. All participants were asked to eat chocolate and then rate their experience. Half the participants were asked to refrain from eating chocolate while the other half were given chocolate and told they could eat as much as they wanted over the next two weeks. When the test was repeated two weeks later, unsurprisingly, the group which abstained from chocolate rated the experience of eating it

again far better than they had originally. For the other group, chocolate had lost much of its draw.

Similarly, in (22) the researcher looks at people's enjoyment while watching a television programme, one group nonstop with the other being interrupted with advertisements. While the participants did not enjoy the advertisements and would rather not watch them, they enjoyed the television programme more because of them. In these situations and in CCS, the interruptions are making each moment of the game or show as enjoyable as the last. The interruptions may also have an additional positive effect.

2.3.3 The Zeigarnik Effect

The Zeigarnik effect states that people remember more clearly those tasks in which they have been interrupted in before completion, then those which have been completed fully without interruption. In an experiment described in (17) a group of students are given mundane tasks to complete. Half are interrupted before completing these tasks while the other are not. The students who were interrupted could remember more details about the tasks, and further, many chose to return to the task to complete it, despite not being required to do so. The parallel for mobile games is clear. CCS purposely interrupts the game play, ensuring the player will be thinking about the game for the cool down period, and will pick it up as soon as possible.

This feature is ideal for the mobile platform. While an interruption on a computer or console could easily drive the player off the machine and on to other tasks, a mobile game has the advantage of being carried with the user. Smartphones are carried with users practically everywhere and their notifications get significant attention. Games such as CCS interrupt players, and then notify them when they can play again. There are no barriers to

playing these games, such as turning on or loading a console. The game is immediately available to provide the next round wherever the player is.

Another important game in the mobile and social media platform is FarmVille. While no longer in the top 50 games, it has inspired a number of the top ten. The game is a simple strategy game, in which players try to run a farm. "Crops" can be planted by the expenditure of in game money, and then the player needs to wait for the crops to grow. Once grown, the crops can be harvested. FarmVille added a social element, in which players could enlist the help of friends to "water" their crops to substantially increase the yield. This social element helped the game spread virally, and kept players in the game due to the need to reciprocate help given. The basic game design, of initialising a build or task, and then waiting real world time for the task to materialise into more points, has been copied by many of the top mobile games."Clash of Clans", currently the highest grossing game on the playstore has very similar game play, only with a military styling. These games also have the feature of requiring constant attention if the the player does not want their character to deteriorate.

2.3.4 Behavioral Game Design and The Skinner Box

In (13) Hopson discusses some of the physiological methods that can make a long lasting game. In the paper he uses research by B. F. Skinner and his skinner box or Operant conditioning chamber. A Skinner box is an apparatus used to study behaviour in rodents and birds. It is a cage or box in which the specimen is kept. It will have a lever, button or different input method and a food dispenser. The floor of the box may also be an electrode. The experiments all revolve around controlling the frequency of the lever pushes,

or other input method equivalent. If a rat inside the box presses the lever it receives food from the dispenser. It may get shocked if it does not press the lever, or may only receive food on every third lever push. By changing the frequency of the rewards and the punishments Skinner was able to get the animal to push the lever at whatever frequency needed. For example, if the rat received food every lever push, it would quickly tire of the food and stop pressing the lever. If the food was dispensed randomly, with some chance of food each lever press, the rat would press the lever repeatedly, even after receiving its fill of food.

Many of the behavioural traits have been found to be species indifferent.

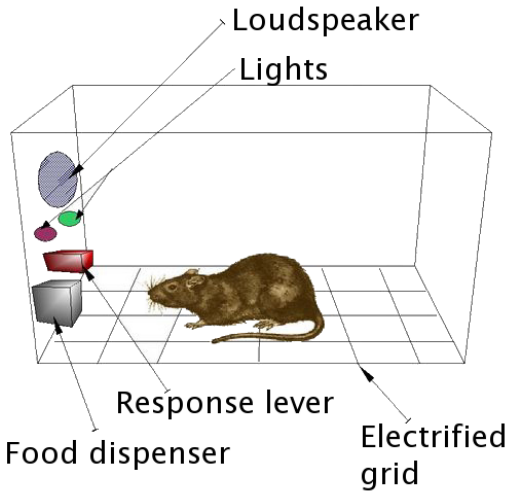


Figure 2.7: A Skinner Box with a rat as the specimen

Hopson describes how a well designed game is similar to a human Skinner Box. The food is replaced with reaching a new level, or gaining a new item in the game, carefully timing and weighing rewards can keep the human playing at the maximum rate. Games like FarmVille and Clash of Clans take full advantage of these physiological traits.

In (16), a collection of essays written by game designers, Jackson points out

FarmVille's creators in particular as following Skinner box like design in their games. FarmVille is a game that is devoid of narrative or skillful play. Most of the game play is simply clicking on objects to upgrade them with game money, and then waiting for game money to accumulate again. Jackson compares FarmVille to a satirical game "Cow Clicker" in which players may click on a cow each 6 hours, each click delivers a point. The game also gave the opportunity to reduce the wait by paying money or inviting friends. There was an in game currency which could buy new cows or other aesthetic objects for the game. The game was designed by Ian Bogost to call attention to the growing number of mobile and social media based games which lacked "meaningful opportunities for achievement, social interaction, and challenge". In Ian's words "Cow Clicker is Facebook games distilled to their essence"(3). In an ironic turn Cow Clicker became surprisingly successful, gaining over 50,000 players before being removed(16).

Cow Clicker makes an interesting design guide for any mobile game. It managed to hold a significant playership despite being designed to have no content. Careful design of rewards and punishments can be all a game needs to keep players attention. It is important to note at this point that all the games that rely on these techniques are also visually very well designed, with colourful animations and careful sound engineering. The rewards that FarmVille deliver would likely not have the required effect on players if they were not so carefully designed.

Chapter 3

Ethics

This section addresses some of the ethical questions that this project encounters, namely that of addictive qualities in games, and protection of user data.

The project that I am undertaking involves the designing of a mobile game, one that will be collecting user data, and could use known addictive techniques in its design. This report is meant as a discussion on the possible ethical dilemmas and pitfalls that this project could fall into.

There are two main questions that this project brings about. First is Game Addiction. The mobile game industry is made up of some of the most addictive games ever to be made. The fast paced short life span development environment in which these games are made has moulded them to be as addictive as possible. 60 percent of the revenue from in app purchases came from 0.23 percent of the users. This brings up the question, is it ethical to purposely design a game to be addictive? Further, is it ethical to do research into addictive game techniques with the intention of building a game that uses them? The second major question of ethics is that of user data. To gain any improvements in speech recognition technology, the final game

that is developed from this project will need to collect data on each game. Information such as speech data, start time, finish time and what moves are taken and when is collected. This information is needed to analyse for studies but could also be a breach of privacy. In particular, the collection of the results from the speech recognition system could contain sensitive information about the user and some consideration will need to be taken to ensure that this information is anonymised correctly.

I will now go into further detail with these two areas, including examples of their misuse.

3.1 Addictive Properties

As part of my research for this project I looked at what features made the current most popular mobile games so successful. I found that much of their appeal may be based on their addictive nature. Candy Crush Saga makes use of Hedonic adaption delay effects, and FarmVille can be described as a carefully crafted Skinner box. The designers of these games may not have deliberately used research into addictive behaviour to build their games, but it brings up an important ethical question, is it ethical to build games as traps for addicts? It would seem obvious that any research done with profit in mind and with full knowledge that its effects will be harmful to people is unethical. However, my project has the aim of using people's idle game time to do useful work and to help research that may improve people's lives. Do these good intentions make the research ethical? One could make the argument that addictive games exist and will always exist and it is surely ethical to replace these games that are for profit with games that could do some good. A similar line of logic can be used to justify research in the weapons industry. More

accurate weapons will hopefully lead to less innocent victims and therefore it is ethical to do research for the development of predator drones or cluster bombs. It is a difficult question of ethics that I will deal with in my project by looking at a more ethical styles of game design.

3.2 Data Protection

In January 2015, mobile games creator Big Fish informed their users that personal information, including names, addresses and in some cases bank details had been stolen from their servers. In February it became apparent that Samsung's smart televisions were recording everything spoken within earshot and sending it back to Samsung, to potentially be sold to advertisers. Data protection has become an important ethical question of late and it is of key concern when designing any system that relies on collecting data. Collecting inappropriate data, like Samsung or the NSA has, could easily be considered unethical. Just because a system can collect data does not mean that it should. I have kept the data that I collect as concise as possible, only keeping metrics and information on each game and not keeping information on names, ages or other personal information of users that would not serve a purpose. Keeping the data anonymous is another key step to ensure that no data laws are broken. One design choice I made in this project is to use Google accounts as a sign in device. This leaves the duty of storing passwords and collecting personal information out of my hands. Rather than information being stored on any server I set up, it is being held behind the security of Google.

Chapter 4

Design

This section describes early designs, their pros and cons. It then illustrates my final design for the game, and the reasons for choosing it.

4.1 Early Design

My original designs for this game were closer to that of Treadris(20); a simple transcription game against the clock and borrowing where possible from the design techniques used by popular mobile games. The game would be single player. Players would hear a section of speech and then be asked to type out the words correctly, as fast as possible. The player would be given suggested words to help speed up the process. The suggestions would be based on the results from automatic speech recognition. If a word was not present the player could bring up a standard keyboard to type out the word. The game would be separated into levels, where a player would need to transcribe a group of sentences under a certain time to progress. Additional difficulty could be added by including misspellings of suggested words, about which the speech recognition system was most confident. This would provide a

simple method of grading. An incorrect sentence would result in a time penalty, bringing the end of the level closer. A rough mock up of the early design can be seen in figure 4.1. In the image, the red section of the screen is moving towards the green side. Once the whole screen is red the player would be dead and the level failed. Submitting a wrong sentence would push this red line ahead by some random amount. The design would incorporate some addictive qualities used in current mobile game design. A Hedonic adaption delay effect would be included, limiting players play time artificially. A credits system that could be earned through passing levels, sharing the game on social networks, or possibly recording audio of known sentences to help develop language models would also be included. These credits could then be exchanged for more lives, or to increase the allowed time for a level.

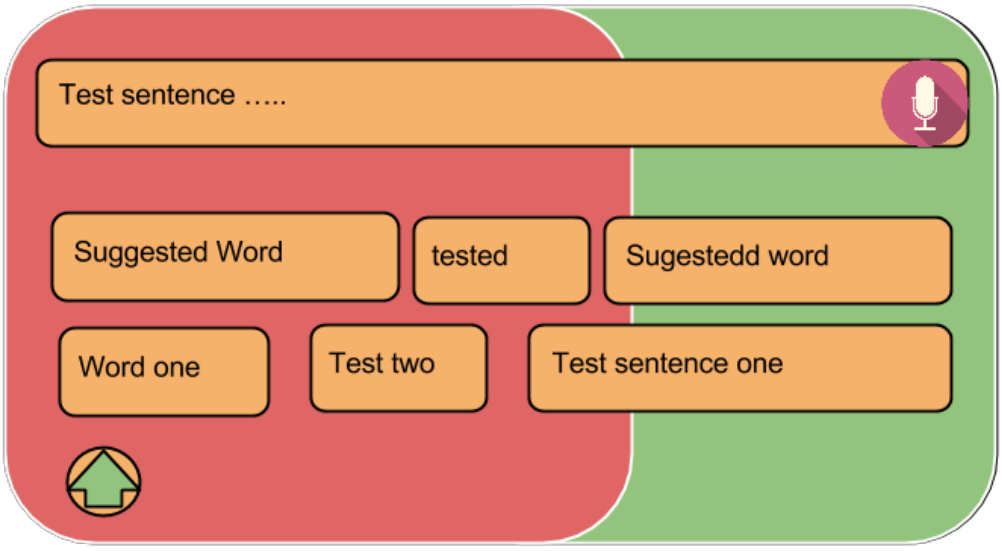


Figure 4.1: Early design mock up

Early Design Problems

- Screen Size - Typing on a small touch screen can be slow, error prone and frustrating. If the suggested words were often wrong, the keyboard needed to be used and the game would devolve into transcription. This may work with a large tablet screen but a normal smart phone would be at a significant disadvantage. Players may find that their ability was entirely based on the room within which they have to type.

- Scoring - The game must reward a player when they enter a correct sentence and punish them when they submit an incorrect one. The game does not know what the sentences are. That is the point of the game, to collect the correct transcriptions. Other than using misspellings to catch out errors, the game must use its recognition results to gage whether an answer is correct or not. If the recognition was badly wrong, giving a high confidence for an entirely incorrect sentence, there is no way for this to be noticed. Players who attempt to correct the game will be marked incorrectly.

- Design Quality - This design is based on some of the more popular mobile games at the moment. Those games tend to rely on the quality of their production and aesthetics. This project is unlikely to match the quality and production value that those games can produce. I am uncertain that the addictive techniques used by mobile games today can be successfully implemented in the time that is available.

4.2 Final Design

The final design moves away from the simple style of Treadris. Instead of using quick reactions and skill as a draw it uses social aspects. Inspired by the map tasks used in speech experiments, this design uses communication

between its two players as the main theme. There are two players and two defined roles; a guide and a guided. The guided player one is in a maze like map. They must get to the finish mark as fast as possible. The task is possible on their own, but monotonous and time consuming. The guide, player two has a top down view of the map. They can scroll across the maze, find the finish and inform player one. The only form of communication between the two players is an automatic speech recogniser. The concept for this design originally came when noting how funny the incorrect results from the speech recognisers could be. The recognisers combination of random, unintelligible sentences and completely inappropriate or out of context phrases can make for fantastic entertainment on its own. Rather than see recognition errors as a problem, this design uses them as an entertainment feature.

Cooperation is another key element of the game. It relies on the fact that players will continue to play through difficult or dull game moments if they know their partner needs them. Leaving the responsibility of one players game performance on another will hopefully ensure player engagement.

Figure 4.2: Player 1 (left) & Player 2 (right)

This design has a less direct effect on the speech recognition improvements. Treadris succeeded in getting players to do useful work, to correct transcriptions. Those transcriptions could be used, the audio could be taken from untranscribed videos and used to provide subtitles or improve the search ability of its contents. This design does not match that level of utility. Instead it will improve on research, possibly replacing or improving experiments that make use of map tasks. It will evaluate recognition systems, or collect new dialogue on chosen vocabulary. The game could also include a correction aspect, asking players to correct the messages that they sent after the game. This could act as a bonus round, allowing for extra points to be earned on a public leaderboard.

Figure 4.3 shows a flow chart describing the different states of the game, starting from a player opening the application and going on to play a full round, then changing roles.

4.2.1 Limitations

As already mentioned earlier in this section, yhis design does not provide the same utility as Treadris, or other earlier designs. It is limited to collecting speech corpus data, or an evaluation platform. In addition to its limitations as a tool, it is limited as a game in some respects. The most popular mobile games are generally used in public places, people use these games while commuting, in a crowded train or bus. The requirement to speak into the device will limit the possible audience for this game. Users will be unlikely to play SpeechIsHard in public places. The two player nature of the game also creates a barrier to play, as a match will need to be found. This feature is an advantage in some respects, as it will persuade players to invite friends to download the game and play them. Despite this advantage, it also provides a barrier to someone trying to play as users may easily be drawn to other games while they wait for a game match. Finally, the game relies on the players being unfamiliar to the maze. Even though the map is randomly produced, there is a limit to how much the map can vary. Players may find that they can remember parts of the map, making their guide unnecessary.

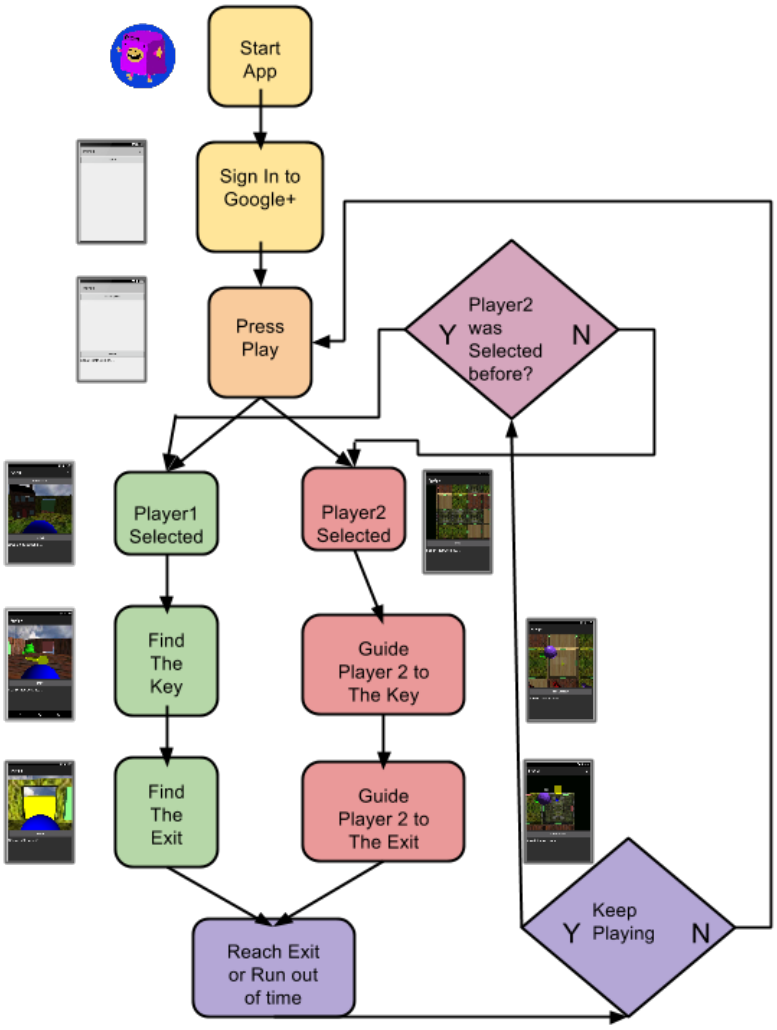


Figure 4.3: Flowchart explaining how to play the game

Chapter 5

Implementation

This section outlines the implementation methods, the technical challenges faced and the solutions used to overcome them. The project is broken down into three separate sections; back-end and real time communication, the speech recognition system and finally the 3d game itself. This section also includes a brief description of the Android app work and a breakdown of the schedule of work.

5.1 Communication

A key technical challenge of any real-time multiplayer game is keeping players in communication with each other. Data relating to the game state must be sent many times a second to keep both players up to date. Games rely on speedy responses, and are very susceptible to long latency times. In SpeechIsHard, both players need to see the same information for any instructions to be useful.

5.1.1 Communication Requirements

Below is a list of requirements for the connection used in this app.

- World wide - SpeechIsHard depends on the two players being in different rooms to work at all. A local network connection, or short range systems such as Bluetooth would not be acceptable. People are unlikely to be close enough to receive messages, and also far enough to make the verbal instructions inaudible.

- Low Latency - Consistency of the game state must be maintained across both devices. High latency could lead to one device going behind the other when it comes to game events, e.g. death of a character.

- Reliability - The connection must be maintained for the entirety of the game. If the devices were to loose connection at any point the game would need to be abandoned. This is contrasted by most communication over the Internet, where a brief interruption in communication only results in a delay in content delivery. Each game set-up represent a significant investment in time and resources.

- Scalability - The communication system must be capable of providing acceptable service regardless of the number of games that are being played concurrently. Mobile Game uptake can be sporadic in nature, and the system must be capable of dealing with a sudden increase in use.

- Inexpensive - The system chosen shouldn't depend on expensive services. In short this means that the resources of severs and other rented machines should be kept to a minimum.

- Secure - To properly moderate games some form of authentication will be required.

5.1.2 Peer 2 Peer

Peer 2 Peer (P2P) Is a direct connection between two or more devices. Rather than the Server Client model used for website requests, P2P requires the clients to communicate directly with one another. In the case of this project, it means two Android devices communicating with each other with no external server in the middle.

The advantages of this in terms of speed are obvious. Using a middle Server adds to the message travel time, first going from the device to the server, to then be sent on to the intended player. Using P2P allows messages to be sent directly to the intended device. This can mean an important change in latency, especially if the devices are far from the server geographically. P2P also removes the single point of failure that a server presents, and significantly reduces the running costs of the final game. The alternative to P2P is providing a server to handle all messages to and from all active players. While relatively simple to implement this would come with considerable time penalties, and the cost of maintaining a server. Also important to consider is the scalability of the design. This app is unlikely to encounter high levels of use but it is important to choose a design that could. A Server based communication system would present a single point of failure. All messages must go through the one server. High demand could lead to slower message times, and/or additional costs, in paying for more machines. P2P is largely unaffected by large traffic. Regardless of how many people are playing the game, the two devices will get the same response times. There is a down side to P2P and that is the complication of implementing it over the web and

between Network Address Translation gateways.

NAT Gateways

Network Address Translation (27) is the process of translating a devices local IP address to a public IP address, done generally by a network router. It may be done for security reasons or to conserve public IP addresses. The process goes as follows, device A is on a local network connected to a router R, which in turn is connected to the rest of the Internet. A's IP address on the local network will be a 192. or 10., not valid public addresses. These addresses can not be used as a return address in an IP message packet. The router instead overwrites the sender address and port number of any message sent from A to the rest of the Internet. It uses its own IP address, and selects a free port number. It saves the original data in a table, along with the newly chosen port number that was sent with the message. When a reply is sent back from the message sent by A, the router receives it at the chosen port, and from that looks through the table to retrieve A's local address and the original port. The issue with this for P2P is that any device behind a Nat gateway is unaddressable. The only messages that can be sent to A, are replies to it's requests. If two android devices attempted to send messages to each other directly they would be addressing the routers. Almost all android devices would be connecting to the Internet through some form of Nat gateway, making this an essential challenge to overcome to allow P2P over the Internet and not just local networks.

NAT traversal

Also known as NAT punch through is the general term for any method of connecting two peers that are behind a Nat. It requires a publicly addressable

server to set up(8). The two devices send messages to the Server, which makes a note of the port that their messages came from. Using this port number it is possible for the middle server to predict the port number that will be assigned to the next connection that device A and B make. This prediction depends on the type of Nat protocol being used. Often it is simply a sequential port number. For example, if Device A sends a message to Server S, S gets the message with the return port set to 30. It now knows that the next connection A makes to the Internet will be sent through port 31. After A and B send messages to the server, it can respond with the opposite devices contact information. A receives the public address for B and the port that is likely going to be used in the Nat table to send to B. Both devices then send messages to each other using this information. If the port prediction was successful B will receive A's message as a reply from its own message and a P2P connection will have been set up.

5.1.3 Google Play Games Services (GPGS)

Google Play Games Services Is a service designed by Google to facilitate multiplayer games on their app store, Playstore. It is a free service up to a large usage limit and it utilises P2P connections by implementing Nat traversal. GPGS has an Api for Android, making it a simple process to implement in an app. The association with Google also provides the service with an authentication method in the form of a Google Plus sign in credentials. This is both an advantage and a disadvantage. It provides security and moderation abilities, but at the cost of restricting players to those who have Google Plus accounts. Google Plus accounts are free and quite popular however, which makes this an acceptable cost for the benefit it brings. GPGS also provides for Leader-boards and achievements, facilities for public world wide

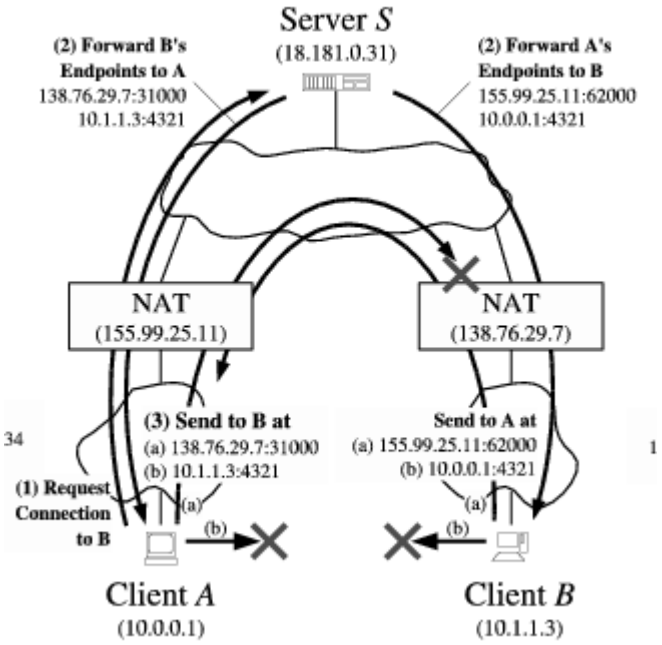


Figure 5.1: Nat punch trough from (8)

high scores, as well as the achievements which can be small rewards peppered throughout the game to keep players involved. These facilities come with the reliability of Google servers which provide massive world wide scalability. GPGS provides an elegant solution to connecting the two players and providing secure authentication.

5.2 Speech Recognition

In future development this app may be used with a developing speech recognition system to help in evaluation. For the current development the speech recognition needs to prove the concept of the game and so must have the following qualities.

5.2.1 Recognition Requirements

- Accurate - Part of the fun of the game will be the inaccuracy of the recognition. However, the results must match the quality that users find in other speech recognition systems that they use in every day life. The instructions must be of some use to ensure the game is playable.

- Computationally Inexpensive - This app runs on android devices and so is limited in processor power and memory. The System chosen must not grind the phone to a halt whenever used, especially when considering the large resource drain that the 3d game and the multiplayer connection will have on the device's resources.

- Fast - For any instruction to be of use in this game it will need to be processed fast. A long wait for instructions that are being recognised for several seconds will leave both players bored.

- Inexpensive - In the same ways as the connection, the speech recognition must not come with a large running cost, as the game is designed to be free and doesn't have any revenue generating abilities.

With these requirements there are two recognition systems which I considered; CMUSphinx and Google speech recogniser.

CMUSphinx

CMUSphinx is an open source system developed by Carnegie Mellon University(18). It uses a Hidden Markov Model and is a highly customisable recognition system. A number of different language models are available and new ones may be used with ease. The system is written in Java and there is also a mobile

version of the software. This lightened version takes up significantly less resources at a cost to accuracy. This customisation could certainly be useful in the future to evaluate new language models, however, for the purposes of this project it is more important to have a working system. The mobile version is built to tie into android operating system. It requires a significant download of over 150 MB to function and provide acceptable results on a restricted vocabulary, falling short when the vocabulary is varied such as it will be with this game. The sizable download will create a barrier to many for trying the game, as well as the increased phone resources that the program must use. The full CMUSphinx program would be most appropriate. The program could run on an external server. This would reduce the resources needed on the phone, but add some latency while the sound data is sent to the server and an answer received. The device will record the speech, and decode it into the frequency vectors used by CMUSphinx. This could then be sent to the server for processing. While this solution would be sufficiently accurate and light on the device, it may run into trouble when considering scalability and expense. Receiving and processing audio files makes this heavy work for any server. Large numbers of players could again increase costs, or the speed to unacceptable levels.

Google Speech Recognition

Google Speech Recognition is built into the Android operating system. It is up to commercial standards. Using Google servers, it has none of the scaling or speed issues that Sphinx could suffer from. It's use on Android is encouraged, and is provided free of charge, with no usage limit. The only significant downside to using Google's recogniser, is that it takes control out of the hands of the developer. The recogniser simply returns a list of the

best chosen words and their confidence figures. There is no way to adapt the language model. This loss of control is acceptable, since the aim of this project is to prove the game concept. Google's recogniser provides a fast and reliable solution which is easily implemented.

5.3 3d Game Engine

The design of this game calls for a 3d first person view. In comparison, most mobile games, such as Candy Crush Saga, are 2d side scrolling games. Building a 3d game for a mobile platform, the resources used become a primary issue. Smart devices tend to have smaller, slower CPUs and GPUs, while often having similar resolutions to desktop and laptop computers. Given the already considerable draw this game will have on a device's hardware with the speech recognition and game communications running, a light weight graphics solutions was needed that could also provide suitably impressive graphics and room for further development.

5.3.1 Android Graphics

The Android operating system uses the OpenGL ES library to communicate with the devices graphical processng unit. OpenGL ES or OpenGL for Embedded Systems is a sub set of the OpenGL Api. OpenGl is an Api for interacting with a devices GPU directly, to achieve hardware-accelerated rendering. It is possible to write code for a game directly with OpenGL ES, however this would require a substantial amount of coding and would fix the game to the android platform. A better solution is a Game Engine that can be developed for Android.

5.3.2 Unity Game Engine

Unity is a cross platform game engine, written in C++ and C#. It has support for 15 different operating systems as of its latest release(15). It is free of charge for any company making less than 100,000 dollars in gross annual revenue. Unity game engine brings together shader technology, 3d modeling software, physics engine and C# scripting. Unity provides a work environment to create 3d games fast. 3d models can be imported and added to a scene. C# scripts can then be added to these models to control the game objects behaviour. Newtonian physics can be applied to game objects by adding a ridged body to them. The lighting and textures of all objects can be adjusted in game. Unity can then compile the resulting project for use with a huge variety of graphics interfaces including the Android interface, OpenGL ES. Using Unity also leaves room for growth in different platforms. It is compatible with ios, Windows, Mac and there is a Unity Web player that makes it compatible with most browsers, as well as Facebook apps. It is this wide range of compatibility, alongside the powerful game creating opportunities that make Unity appropriate for this project.

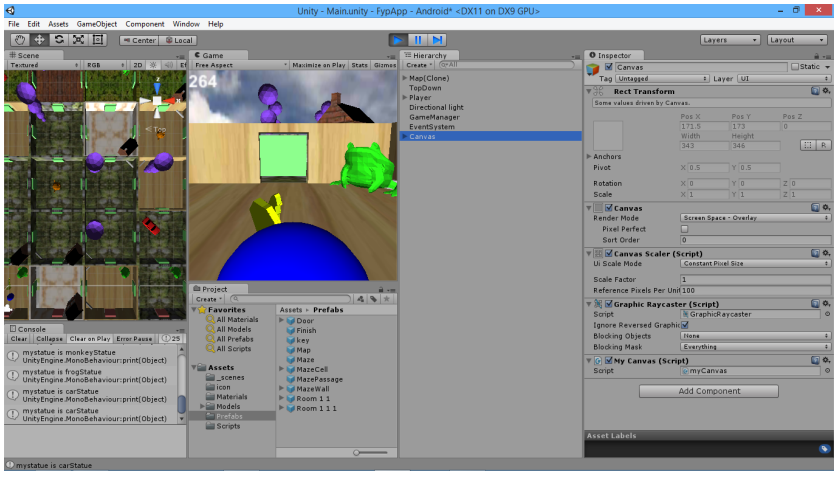


Figure 5.2: Unity work environment

The Final Game has a player character which is simply a blue sphere. This character is placed inside an automatically generated maze at the start of each game. The Maze is made up of a series of rooms with four doors as shown in figure 5.3. Each door is assigned at random one of 5 different textures. In addition to the textures, the rooms are assigned different statues that they contain. These range from a House, as seen in figure 5.3 to a Frog statue. These statues and random textures help give a reference for directions, as well as guiding the dialogue that will be collected. The Random nature of the rooms makes it impossible to simply remember which way to go after the first play through. After creating the maze, a key object is placed in one of

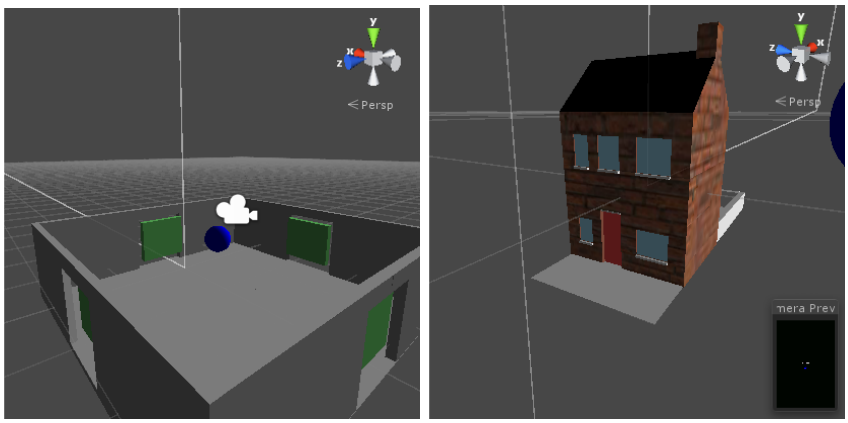


Figure 5.3: The Rooms Model (left) & The House Model (right)

the rooms at random. This key, when found unlocks the door to the exit, which is at the top of the map.

5.4 Android App

The Android app is written in Java using the Android SDK. Figure 5.4 gives an overview of the different parties the App communicates with to run the game. The App, shown in light red, includes code that accesses the Native

Unity code that produces the game. It also communicates with Google Play Games Services through a separate app on the device and directly with the Google speech recognition system. Finally, it communicates with a backend server to send game information.

Communication between the two devices during a match is handled by direct connection. Instructions are sent over TCP, while the player position information is sent unreliably over UDP to improve speed. GPGS provides a call back function for when messages are received over the games P2P network.

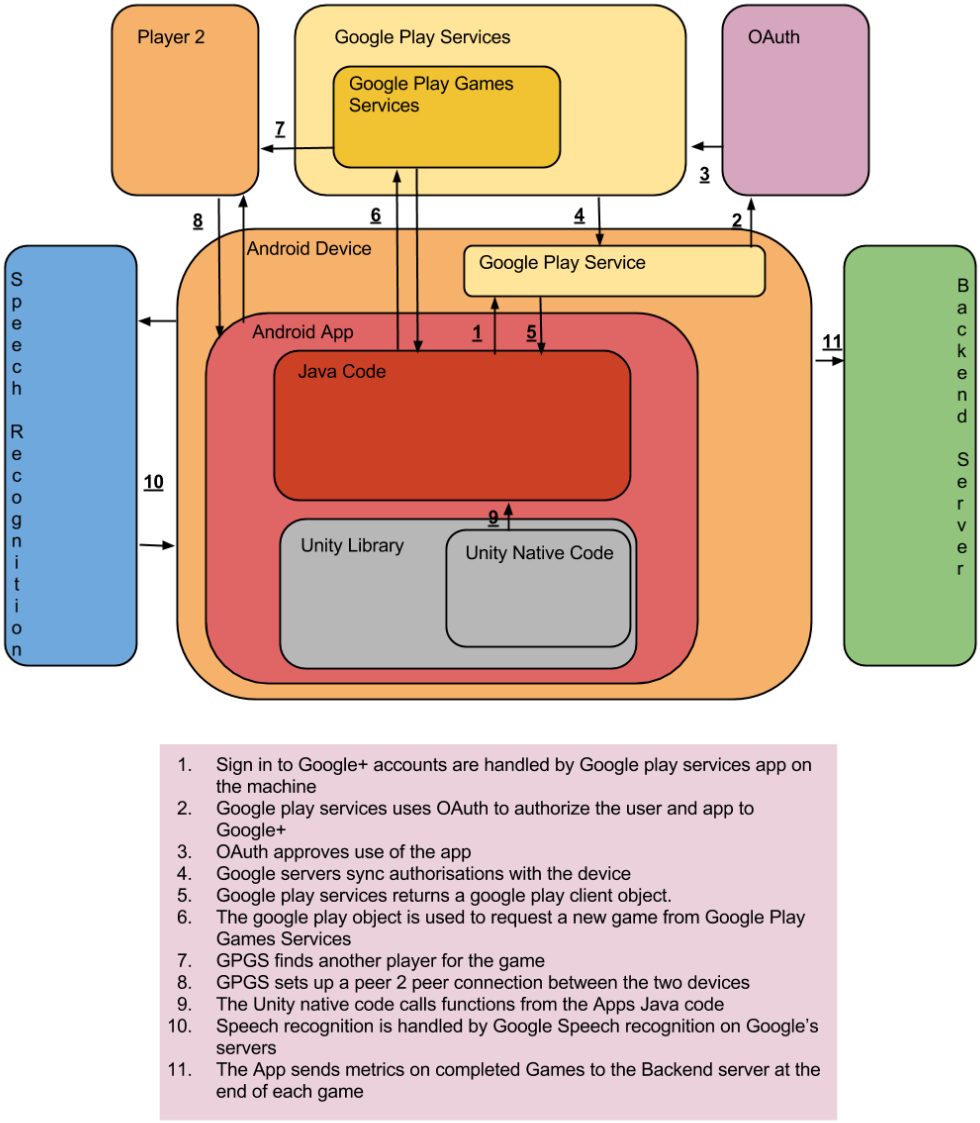


Figure 5.4: Overview of the technologies and how they interact

All code for the game can be found in the cd attached. The applications java code is found in the $SpeechIsHard/app/src/main/java/com/bmaguir/FypApp$

Chapter 6

Evaluation

The Final implementation of the game includes all main features outlined in the Design chapter. The Application successfully pairs two players together for a match, Unity is correctly integrated to provide both players with a 3d game environment, and the speech recognition system provides the players with a means of communication. From a technical standpoint the implementation performs to expectations. The P2P connection between the two devices provides near instantaneous communication, keeping both devices up to date on the game state. The game has a functioning leader board for fastest playthrough. To evaluate the games usability, playability and level of enjoyment a survey was used to gather opinions. The goal of this evaluation is to gauge whether the basic game concept is playable, and if so, what aspects of the game need work on for it to be successful.

6.1 Survey

The participants for this survey were required to be somewhat familiar with casual gaming to ensure they could reliably compare SpeechIsHard to other

games that they have played. The questions for the survey are based on this paper (9). The paper attempts to create an evaluation scale for educational games, and uses a survey as a basis for their scale. The questions on their survey focus on the following areas; concentration, goal clarity, feedback, challenge, autonomy, immersion, social interaction and knowledge improvement. For each section I chose one question that was most appropriate to SpeechIsHard. In a addition to these areas, I also asked players about the speech recognition system. How accurate it was, how much they used it, and how accurate they find speech recognition in their everyday life. The full survey can be found in appendix A.

6.2 Discussion

In this section I will discus the results of my survey, looking at whether the game works as a concept, what aspects of the game are the weakest and what new developments might improve it. The results as a whole showed that the game is fun and enjoyable, with a user friendly interface. The survey also highlighted the lack of clear instructions, player feedback and much frustration with the speech recognition's speed and accuracy.

Limitations of the survey This Survey contains results from just 7 participants. The participants may not be providing entirely impartial answers due to their relationship with the researcher. This survey was undertaken to identify areas of the game that need work, and to give a suggestion as to whether the main game concept is usable. This survey can not be used to make sweeping statements about the quality of the game.

6.2.1 Positive Feedback

Enjoyability and Concentration

These questions were chosen to gauge the level of enjoyment that players experienced and whether the game held their concentration. I consider this an important clue as to whether the games basic concept works.

The game grabs my attention

This was answered positively, with 83.3% of participants either agreeing or strongly agreeing with the statement and 16.7% not agreeing nor disagreeing.

I enjoyed the game without feeling bored or anxious

This question was answered very positively, with all participants agreeing or strongly agreeing with the statement. In addition to these questions, there were many comments on the survey suggesting that the game had been enjoyable. Such as "Quite enjoyable to play actually", "Core idea is very interesting and novel" and "Very good concept". These results show that within the limited domain of this survey, the game concept of two players helping each other out of a maze with speech recognition, was shown to be enjoyable.

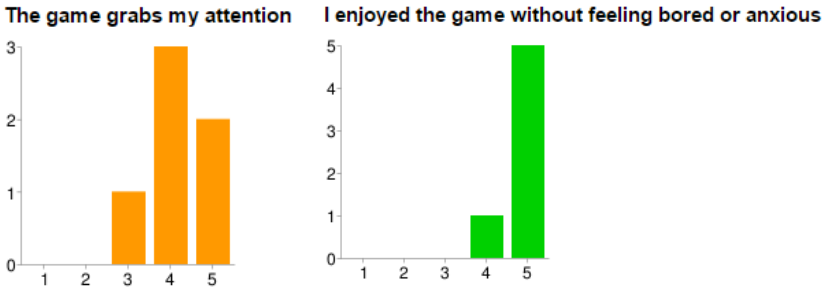


Figure 6.1: Answers from question 1 & 4 of the survey

player controls

This question was chosen to evaluate the quality of user control and the method for controlling the players through out the game.

I feel a sense of control and impact over the game

This question was answered positively, with all participants agreeing with the statement. As well as this question, I received two verbal comments on the ease of use of the user interface in the game.

social interaction

A strong feature of the game was social interaction. All participants agreed that they felt cooperative with the other player while playing. While they did feel cooperative, there were varied results when asked how much help the other player was in the game. Just over 50% of players either disagreed, or were not strongly opinionated either way. This suggest a possible issue in

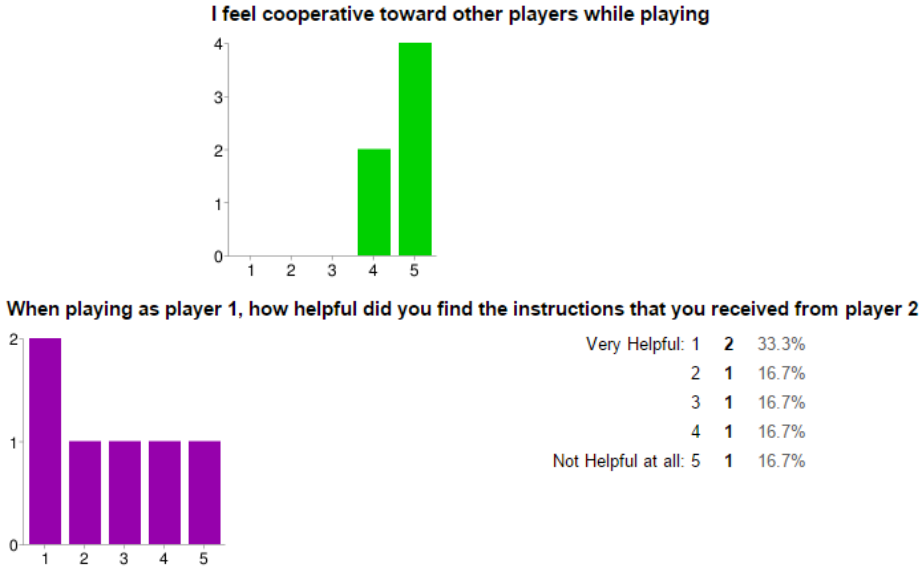


Figure 6.2: Answers from question 10 & 14 of the survey

SpeechIsHard, which is the dependency on two players for the game to work. One player's enjoyment may be strongly linked to others ability to play the game. This trait may be an advantage and disadvantage. The dependency improves the social interaction between players but it also may leave players feeling frustrated at inept partners.

6.2.2 Negative Feedback

The main negative feedback that was received was about the clarity of the in game goal. Many participants did not properly understand what they were supposed to do, or the role that room features had in the game.

Overall game goals were presented clearly

This question got a mix of answers. This may suggest that the question was poorly worded and participants did not understand what it was asking. There were further comments on participants not understanding the game until playing for some time. For example "I was not sure what the effects of the various objects in the rooms would be. So I did not know if rooms were to be avoided or not" and "lack of instructions". The instruction material is not sufficient. The addition of a tutorial round, or a better difficulty progression might be necessary.

I know the next step in the game

This question was answered in a similar fashion. The game would appear to lack visual cues of the next task to complete.

speech recognition

The Speech recognition was highlighted as a point of frustration for some participants. In a question asking for comments on the most frustrating part of the game, the speed at which instructions could be given were mentioned.

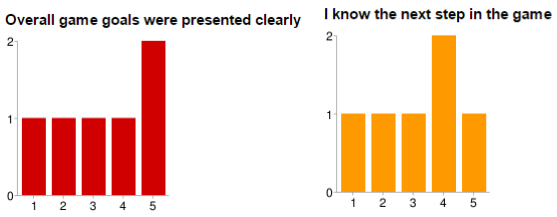


Figure 6.3: Answers from question 2 & 7 of the survey

"The speed at which I could communicate was quite slow. When I was giving directions, if I saw a mistake made by the other player it could take quite a while for me to stop the other player from making the mistake." and "When they do things before you tell them what to do". Depending on the quality of the audio that is returned, the speed of the recognition message reaching a player can vary greatly, but generally takes several seconds. In a fast-paced game, that is a very long time. A solution to this problem is difficult, as faster speech recognition might not be possible. Having the recognition performed on the device might return faster results. However, this would depend on the performance of the device playing the game and the quality of the results would be reduced. One part solution would be to inform the other player when their partner is attempting to send a message. An indicator that the speech recognition system is being used, and a message will be received soon. This might solve the issue of players moving ahead of the instructions they have been sent. Also asked in the survey was how often players used speech recognition in everyday life and if so how accurate they found it to be. They were also asked to rate the accuracy of the in game speech recognition system. One verbal comment which was recorded was that the participant found the speech recognition very inaccurate, until they altered their speech. It is an interesting topic to study, as to how people naturally change their speech to be better understood by a recogniser.

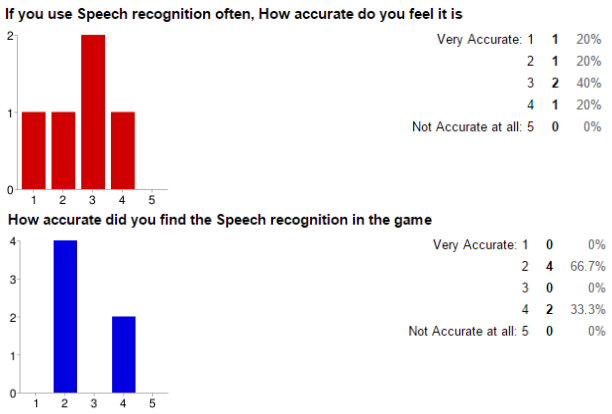


Figure 6.4: Answers from question 17 & 15 of the survey

Graphics and Visuals

There were a number of comments suggesting upgrades of many of the graphical elements of the game.

6.3 Limitations of The Evaluation

The evaluation of this game focused entirely on its merits as an entertainment tool and not at all on its possible help to speech recognition technology. Still to be determined is whether the audio collected from SpeechIsHard is of sufficient quality to be used in a speech corpus, or as an evaluation tool. To do this, more rigorous testing would be required.

Chapter 7

Conclusions

Throughout this project, the game SpeechIsHard has been researched designed and built. The report details the research undertaken, in particular identifying possible areas to aid speech recognition technology, a review of some of the techniques currently used in the fast growing mobile game market, a look at what serious games are and what they are capable of. The report then goes through the main ethical questions that this research brings about, mainly the ethics of producing willfully addictive games and data protection. Next an early design is outlined, that of a simple transcription game and why its small screen and difficult scoring system makes it unsuitable. The Final design is then described, a game based on the map task experiments used in speech recognition. Implementation follows the technical challenges that this design produced and the solutions that were implemented to over come them. These were the P2P network solution, using Google Speech recognition for recognition purposes and the Unity game engine which produced the 3d game. The Final design is then evaluated for its enjoyability. In short SpeechIsHard is found to be an enjoyable game concept, with major criticism of the instructions and the speed of the recognition system. It

remains to be evaluated on the ability of the game to produce usable results in speech recognition research. The Goal of this project was to design and build a prototype that could aid in speech recognition research. The results of the evaluation would suggest that SpeechIsHard has potential to be a successful gamification of the traditional Map tasks experiment, with further development of the game.

7.1 Further Works

After evaluation of the design, I have identified some new possible areas of work, moving forward from this project.

7.1.1 Further Game Development

New features of the game could be added to improve the game play. Based on original hopes for the game and on the feedback received from the survey, I have described some possible additional features that could improve the game.

- Tutorial round - As mentioned in the discussion, a tutorial level might properly explain the game and the player goals, something that it was criticized for in the evaluation.
- Locked Doors - Some of the interior doors in the maze could be locked, impeding the player to the key and forcing further interaction between the two players.
- Increasing Difficulty - This could be achieved by new features such as locked doors, a larger map or a shorter time limit. The addition of

increased difficulty would add further game play. As players improve, the game would continue to be challenging.

- Enemy Agents & New Dangers - An AI agent which would chase the player through the map, as well as new dangers such as rooms which decrease the time limit when entered. These features would add excitement to the game play and provide an interesting element in the recorded speech. The dangers would be a measure of the stress that player was experiencing at the time. The performance of the recogniser when dealing with hurried and stressed speech might be of interest.

- Speech correction round - A major limitation of SpeechIsHard, when compared to the game Treadris, is it has a less practical addition to speech recognition research. This might be improved by a bonus round after a match when a player would be given the chance to correct the messages that they sent throughout the last match. These corrections might be useful for improving the speech recognition system, or make speech corpus collection easier.

7.1.2 New Fields of Research

SpeechIsHard might also be of use in other fields that make use of map tasks, such as automatic language translation. The Game could be altered to make use of translation instead of speech recognition. The game pairing function could also be altered to ensure the two players were speaking different languages.

Handwriting

The game could be altered to use a handwriting tool for when the user does not want to use the microphone. A large amount of mobile gaming is done in public places such as on commutes. In these environments it is likely that players would prefer a silent option for playing. The Handwriting tool, a tool which takes touch input as writing and converts it to text would have many of the errors which come with speech recognition, keeping the game mechanics similar. The process of handwriting is also quite similar to that of speech recognition, also making use of Hidden Markov Models.

Speech Control

A common misunderstanding about SpeechIsHard was that the player controlled the character by speaking commands to it. This and some other feedback suggests another application for the game. The games focus could be changed to that of tutoring the player in the pronunciation of words. The character could be controlled by way of speaking key words into the phone, the instructions being carried out to the degree that the recogniser matched the speech to the key word. The key words could change for each level, and the aim of the game might be to learn correct pronunciation for those learning a new language, or possibly as a form of treatment for those with speech impediments. It is unclear from the research done for this project whether a speech recognition system could deliver the necessary accuracy in pronunciation. That said it remains an interesting area, which would certainly function as a game for a purpose.

Bibliography

[1] Hayakawa Akira, Nick Campbell, and Saturnino Luz. Interlingual map task corpus collection. In *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[2] Anne H Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. The hcrc map task corpus. *Language and speech*, 34(4):351–366, 1991.

[3] Ian Bogost. Cow clicker about. `http://cowclicker.com/`, 2015.

[4] Susanne Burger, Zachary A Sloane, and Jie Yang. Competitive evaluation of commercially available speech recognizers in multiple languages. In *Proc. of Fifth International Conference on Language Resources and Evaluation (LREC), Genoa, Italy*, 2006.

[5] CMUSphinx. Basic concepts of speech. `http://cmusphinx.sourceforge.net/wiki/tutorialconcepts`, 2015.

[6] Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, and Dan Dixon. Gamification. using game-design elements in non-gaming contexts. In *CHI '11 Extended Abstracts on Human Factors in Comput-*

ing Systems, CHI EA '11, pages 2425–2428, New York, NY, USA, 2011. ACM.

[7] S. Folkman. *The Oxford Handbook of Stress, Health, and Coping*, chapter 11. Oxford Library of Psychology. Oxford University Press, USA, 2010.

[8] Bryan Ford, Pyda Srisuresh, and Dan Kegel. Peer-to-peer communication across network address translators. In *USENIX Annual Technical Conference, General Track*, pages 179–192, 2005.

[9] Fong-Ling Fu, Rong-Chang Su, and Sheng-Chin Yu. Egameflow: A scale to measure learners enjoyment of e-learning games. *Computers & Education*, 52(1):101–112, 2009.

[10] Mark Gales and Steve Young. The application of hidden markov models in speech recognition. *Foundations and Trends in Signal Processing*, 1(3):195–304, 2008.

[11] Think Gaming. Candy crush saga revenue details. `https://thinkgaming.com/app-sales-data/2/candy-crush-saga/`, 2015.

[12] HabitRPG. How it works-habitrpg. `https://habitrpg.com/static/features`, 2015.

[13] John Hopson. Behavioral game design. *Gamasutra, April*, 27, 2001.

[14] Google Inc. Top grossing games - google play store. `https://play.google.com/store/apps/category/GAME/collection/topgrossing`, 2015.

[15] Unity Inc. Unity public relations documentation. `http://unity3d.com/public-relations`, 2015.

[16] Benjamin Jackson. Hard fun, the zynga abyss. In *Distance 01*, pages 13–21. Distance, 2013.

[17] K. Koffka. *Principles Of Gestalt Psychology*, pages 334–337.

[18] Kevin Lenzo. Cmusphinx documentation. `http://www.speech.cs.cmu.edu/sphinx/doc/Sphinx.html`, 2015.

[19] Max M Louwerse, Patrick Jeuniaux, Mohammed E Hoque, Jie Wu, and Gwineth Lewis. Multimodal communication in computer-mediated map task scenarios. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pages 1717–1722, 2006.

[20] Saturnino Luz, Masood Masoodian, and Bill Rogers. Supporting collaborative transcription of recorded speech with a 3d game interface. In *Knowledge-Based and Intelligent Information and Engineering Systems*, pages 394–401. Springer, 2010.

[21] Andrew Cameron Morris, Viktoria Maier, and Phil Green. From wer and ril to mer and wil: improved evaluation measures for connected speech recognition. In *INTERSPEECH*, 2004.

[22] Leif D Nelson, Tom Meyvis, and Jeff Galak. Enhancing the television-viewing experience through commercial interruptions. *Journal of Consumer Research*, 36(2):160–172, 2009.

[23] Jordi Quoidbach and Elizabeth W Dunn. Give it up a strategy for combating hedonic adaptation. *Social Psychological and Personality Science*, 4(5):563–568, 2013.

[24] Lawrence Rabiner. A tutorial on hidden markov models and selected

applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[25] Anne H Schneider, Johannes Hellrich, and Saturnino Luz. Word, syllable and phoneme based metrics do not correlate with human performance in asr-mediated tasks. In *Advances in Natural Language Processing*, pages 392–399. Springer, 2014.

[26] Gabriel Skantze. Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication*, 45(3):325–341, 2005.

[27] Javvin Technologies. *Network Protocols Handbook*, page 27. Javvin Technologies, 2005.

Appendix A

Evaluation Survey

SpeechIsHard Evaluation Form

Each question is optional. Feel free to omit a response to any question; however the researcher would be grateful if all questions are responded to

Please do not name third parties in any open text field of the questionnaire. Any such replies will be anonymised

The game grabs my attention
How closely does this statement match your opinion of the game

1 2 3 4 5

Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

Overall game goals were presented clearly
How closely does this statement match your opinion of the game

1 2 3 4 5

Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

I received feedback on my progress through the game
How closely does this statement match your opinion of the game

1 2 3 4 5

Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

I enjoyed the game without feeling bored or anxious
How closely does this statement match your opinion of the game

1 2 3 4 5

The challenge is adequate, neither too difficult nor too easy

How closely does this statement match your opinion of the game

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

I feel a sense of control and impact over the game

How closely does this statement match your opinion of the game

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

I know the next step in the game

How closely does this statement match your opinion of the game

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

I feel emotionally involved in the game

How closely does this statement match your opinion of the game

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

I forget about time passing while playing the game

How closely does this statement match your opinion of the game

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

I forget about time passing while playing the game

How closely does this statement match your opinion of the game

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

I feel cooperative toward other players while playing

How closely does this statement match your opinion of the game

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

Are there any Features you felt were missing from the game?

What aspect of the game did you feel was most frustrating ?

Any comments on the game?

When playing as player 1, how helpful did you find the instructions that you received from player 2

	1	2	3	4	5	
Very Helpful	○	○	○	○	○	Not Helpful at all

How accurate did you find the Speech recognition in the game

	1	2	3	4	5	
Very Accurate	○	○	○	○	○	Not Accurate at all

How often do you use Speech recognition In normal life

- ○ Everyday
- ○ Every week
- ○ Every Month
- ○ Never

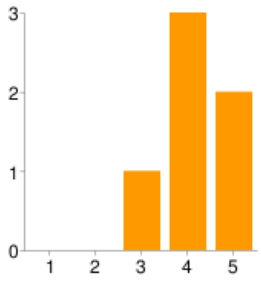
If you use Speech recognition often, How accurate do you feel it is

	1	2	3	4	5	
Very Accurate	○	○	○	○	○	Not Accurate at all

Appendix B

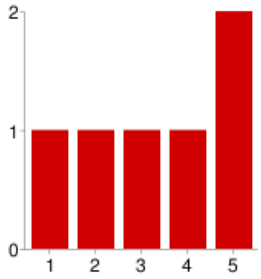
Survey Results

The game grabs my attention



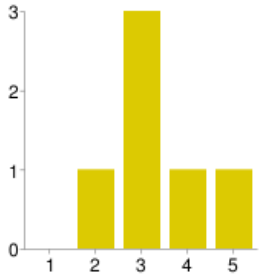
Strongly Disagree:	1	0	0%
	2	0	0%
	3	1	16.7%
	4	3	50%
Strongly Agree:	5	2	33.3%

Overall game goals were presented clearly



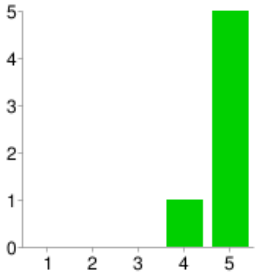
Strongly Disagree:	1	1	16.7%
	2	1	16.7%
	3	1	16.7%
	4	1	16.7%
Strongly Agree:	5	2	33.3%

I received feedback on my progress through the game



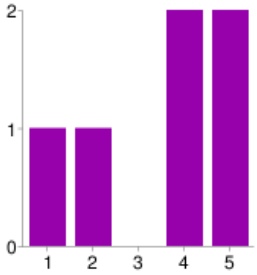
Strongly Disagree:	1	0	0%
	2	1	16.7%
	3	3	50%
	4	1	16.7%
Strongly Agree:	5	1	16.7%

I enjoyed the game without feeling bored or anxious



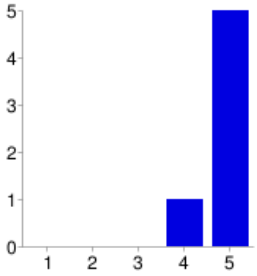
Strongly Disagree: 1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**1**	16.7%
Strongly Agree: 5	**5**	83.3%

The challenge is adequate, neither too difficult nor too easy



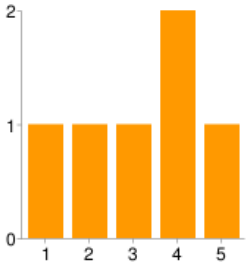
Strongly Disagree: 1	**1**	16.7%
2	**1**	16.7%
3	**0**	0%
4	**2**	33.3%
Strongly Agree: 5	**2**	33.3%

I feel a sense of control and impact over the game



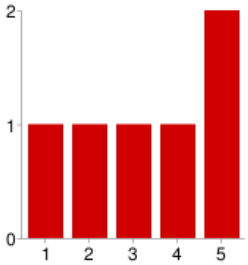
Strongly Disagree: 1	**0**	0%
2	**0**	0%
3	**0**	0%
4	**1**	16.7%
Strongly Agree: 5	**5**	83.3%

I know the next step in the game



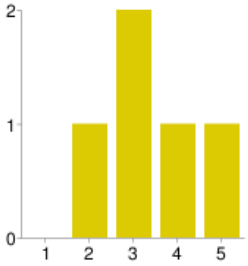
Strongly Disagree:	1	**1**	16.7%
	2	**1**	16.7%
	3	**1**	16.7%
	4	**2**	33.3%
Strongly Agree:	5	**1**	16.7%

I feel emotionally involved in the game



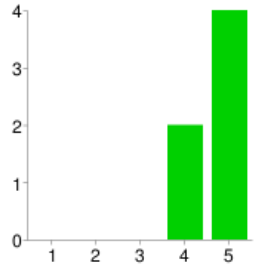
Strongly Disagree:	1	**1**	16.7%
	2	**1**	16.7%
	3	**1**	16.7%
	4	**1**	16.7%
Strongly Agree:	5	**2**	33.3%

I forget about time passing while playing the game



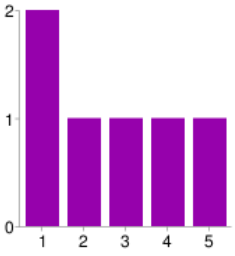
Strongly Disagree:	1	**0**	0%
	2	**1**	20%
	3	**2**	40%
	4	**1**	20%
Strongly Agree:	5	**1**	20%

I feel cooperative toward other players while playing



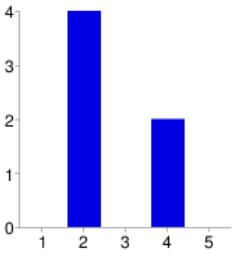
Strongly Disagree:	1	**0**	0%
	2	**0**	0%
	3	**0**	0%
	4	**2**	33.3%
Strongly Agree:	5	**4**	66.7%

When playing as player 1, how helpful did you find the instructions that you received from player 2



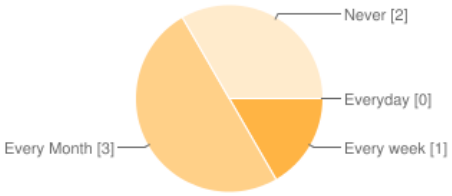
Very Helpful: 1	**2**	33.3%
2	**1**	16.7%
3	**1**	16.7%
4	**1**	16.7%
Not Helpful at all: 5	**1**	16.7%

How accurate did you find the Speech recognition in the game



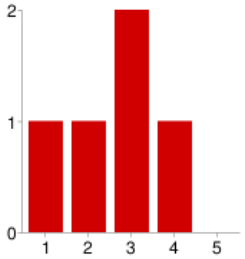
Very Accurate: 1	**0**	0%
2	**4**	66.7%
3	**0**	0%
4	**2**	33.3%
Not Accurate at all: 5	**0**	0%

How often do you use Speech recognition In normal life



Everyday	**0**	0%
Every week	**1**	16.7%
Every Month	**3**	50%
Never	**2**	33.3%

If you use Speech recognition often, How accurate do you feel it is



Very Accurate: 1	**1**	20%
2	**1**	20%
3	**2**	40%
4	**1**	20%
Not Accurate at all: 5	**0**	0%

Are there any Features you felt were missing from the game?

"Something to point to what the target is. Moving enemies in the maze maybe?"

"better mazes, better ui"

"There should be a feature that lets you know which way player 1 is facing when you above as player 2."

"I was playing with the sound off and there was no feedback that the recording had started or stopped - I would have appreciated some visual feedback of this. There was not always feedback that my message had been delivered to the other player e.g. facebook messenger or the two ticks in whatsapp. This would have been particularly useful the first time I tried to play when the game malfunctioned."

What aspect of the game did you feel was most frustrating ?

"The speed at which I could communicate was quite slow. When I was giving directions, if I saw a mistake made by the other player it could take quite a while for me to stop the other player from making the mistake. In this time there was the possibility that they may have continued on and made it worse thereby increasing the difficulty of correcting the snafu. The same applied for when I was in control of the ball. In this case I found it more convenient/less time consuming to carry out an action and wait to see if the other player told me I'd made a mistake than to actually ask them using the speech recognition."

"I was not sure what the effects of the various objects in the rooms would

be. So I did not know if rooms were to be avoided or not."

"None."

"lack of players, lack of instructions"

"When they do things before you tell them what to do"

"Talking was difficult at times."

Any comments on the game?

"Core idea is very interesting and novel. The speech recognition worked well." "most functionality is there, mainly visual upgrades are all that are need, and better instructions about the goal"

"would be better"

"Quite enjoyable to play actually"

"Very good concept."

"Unique and interesting game could be expanded upon by giving the player more maps."