Citizen Science and Game with a Purpose to Foster Biodiversity Awareness and Bioacoustic Data Validation

Pedro Loureiro
Instituto Superior Técnico
Lisboa, Portugal
pedro.loureiro@tecnico.ulisboa.pt

Nuno Nunes
Instituto Superior Técnico
Lisboa, Portugal
nunojnunes@tecnico.ulisboa.pt

Catia Prandi
University of Bologna, Italy
Bologna, Italy
catia.prandi@m-iti.org

ABSTRACT

Over the years, humans have heavily contributed to environmental degradation. Because of that, nature conservation efforts and environmental monitoring have become increasingly important, especially efforts that engage the general public to contribute to the cause. These initiatives fall under citizen science projects and are heavily dependent on the public's support. During our research, we found that the combination of game design with citizen science projects boosts user participation, resulting in the project's overall success.

With this in mind, we devised a project that would combine citizen science, a gameful experience via bioacoustic sensing, creating a game with the purpose (GWAP) of helping classify animal sounds in a given area. Through different game modes and game mechanics we support the analysis of old and recent data, by the players, in the close proximity of the location or from a distance, through an enjoyable and educational game experience.

In the end we created a game with an interesting concept that provided the players an intriguing and fun experience. Although we couldn't acquire enough data to completely validate the audio classification portion of the game, participants enjoyed the game, with a majority stating that they would like to play a version of this game in the future.

KEYWORDS

Biodiversity Awareness, Biodiversity Monitoring, Citizen Science, Game with a Purpose, Bioacoustic Sensors.

1 INTRODUCTION

Over the years, human activities like hunting, collection and poaching have affected the abundance of species that exist in the environment, leading to the their extinction. Habitat destruction and modification is big contributor to this phenomenon and human beings are the main harbingers of environmental destruction[15]. These types of activities originated a mass extinction event, the sixth in roughly 540 million years. Many life forms have succumbed to this event, and studies show that many more could be extinct by the end of the century[20].

In this scenario, monitoring species has become increasingly important. Not only to protect the species but simultaneously protect the environment. The fluctuation of a species population can indicate the general health of the habitat. This type of monitoring is referred as biodiversity monitoring[15] and through it, we can detect possible problems in an habitat, allowing us to find the source of the problem and mitigate the repercussions on the environment.

However, biodiversity monitoring is a complex task, both in collecting and analyzing biological data. Involving a group of scientists to manually collect and analyze data is not scalable. If we consider the size of certain habitats and the number of species that inhabit it, we soon come to that conclusion. The use of technology in collecting and analyzing biological data seemed an apparent and scalable solution. Collecting data with cameras, microphones and other sensors became a widespread practice in the field of biodiversity monitoring, as well as the use of the Internet to share and upload information to specific databases or websites.

However, even if technology enables scientists to collect and analyze data more efficiently, the goal is to globally monitor the environment, creating teams of scientist to do so is still not a scalable solution considering the size of the ecosystems and the number of species that inhabit them the problem becomes overly complex.

To solve this problem, scientists turned to the general public for help, crowdsourcing the collection and analysis of data. Scientists issue detailed instructions for these "amateur-scientist" to follow, enabling them to collect and analyze data without the need for specific knowledge. These instructions also help improve the reliability of the obtained data. These type of projects are referred to as citizen science projects[2]. Citizen science, as the name suggests, is dependent on the citizens' participation. The number of people who volunteer correlates with the amount of data of data that can be acquired which equates with a successfull project. Therefore captivating and engaging the general public becomes a key factor in the project's overall success.

Participants in these types of projects are often motivated by contributing to actual scientific research and by social interactions that result from them. There is an underlying enjoyment in this type of participation. When thinking of enjoyment, one might immediately think of games and gaming. In 2010, Jane McGonigal gave a TED talk entitled "Gaming can make a better World"[17]. McGonigal states that her goal for the next decade is to make saving the world in real life as easy as saving the world in an online game. Collectively we spend more than 3 billion hours weekly playing online games[17]. Like many other serious games applications, biodiversity and environmental monitoring is a promising field of research and development. McGonigal also states that "(...)gamers are willing to work hard all the time, if they're given the right work"[17] this suggests that carefully designing a game that directly contributes to scientific research can be the solution to boosting participation numbers in citizen science projects. Adding game elements to these types of projects can prove to be a great way to motivate the general public to participate, boosting the number of citizen scientists associated with the project[19].

This project leverages on this idea by combining acoustic sensing with gamification and citizen science to help monitor biodiversity in a given area, educating the playerbase about the environment that surrounds them.

We developed a mobile gameful experience that uses audio data from remote sensors and player input to help monitor biodiversity in a given area. Through different game modes and game mechanics we support the analyzes of old and recent data, by the players, in the close proximity of the location or from far away, through an enjoyable and educational game experience.

Motivation

A very efficient and non-evasive way to monitor animal species in a given area is through the use of bioacoustic sensors.

These sensors are placed in various areas, collecting audio data from animal calls, for researchers later to analyze and use in their research. The main problem is that this type of sensor network has the possibility of producing a very large number of unclassified data that the researchers will have to classify. We want to implement a Citizen Science project where users will classify this type of data for the researchers. Projects like bat detective[6] prove that people are very efficient in this type of classification.

We also wanted our project to engage users in a deeper level. Most participants in Citizen Science projects help the effort because they have a great interest in the topic being researched[1]. We wanted to make an enjoyable experience for users, independently of their interest in biodiversity monitoring, while still catering to the enthusiasts and educating the participants, as most Citizen Projects do[1]. So our solution was to build a mobile gameful experience around the classification of animal sounds.

Objectives

Create an enjoyable experience for players in order to motivate participation and consequent success of this Citizen Science project.

We have defined three main objectives for this thesis: i) Increase people's awareness about environment biodiversity; ii) Involve and engage users in data analysis through a gameful experience; iii) Validate collected data so we can obtain an accurate classification of animal sounds.

Giving players biodiversity awareness through a gameful experience will involve and engage users in data classification allowing us to validate sensor collected data, obtaining a data set of classified animal calls.

2 RELATED WORK

This project is heavily inspired in the fields of Bioacoustics (the research field that combines biology and acoustics, studying the production, dispersion and reception of sounds[23]), Gameful Experiences (experiences that incite delight, playfulness, and deep engagement through careful game thinking[16]) and Citizen Science (a collaborative process in which volunteers work with professional scientists to study real-world problems or help further research in a given field[2]). With that in mind, we conducted research in those areas.

The collection and analysis of biological data have been key challenges for any project focused on biodiversity monitoring. To solve this issue, some projects rely on the citizen science framework[2]. Projects like Bat Detective, iBats, and the New Forest Cicada fall into this category, exploiting acoustic sensors to engage citizens in collecting data for biodiversity monitoring. In particular, Bat Detective[6] is an online citizen science project launched in 2012, that relies on volunteers to identify bat calls in audio samples. Those audio samples were recorded during surveys, which were done by teams of volunteers with specific equipment for the effect. The ultimate goal of the project is the creation of a program that automatically extracts the relevant information out recordings to be used by researchers all over the world, simplifying the tracking of bat populations. Another project focusing on bats is the iBats[10] program, established in 2006 through a collaborative effort between the Zoological Society of London (ZSL) and the Bat Conservation Trust (BCT). The aim of this program is to carry out coordinated volunteer-led bat population monitoring on a global scale. This project is powered by citizen scientist. The iBats development team released an iPhone application which can be directly attached to an ultrasonic detector to record grid references, sound files, and other survey data along the route.

The New Forest Cicada Project[22] equipped the millions of visitors of the New Forest, a national park on the south coast of England, with a smartphone application that can detect and recognize the song of the cicada. This project aimed to discover if the cicada is now extinct in the UK or simply migrated to a yet undiscovered site[22]. Although they gathered data that suggested that the cicada was still in the forest, because there were no sightings of it, only possible detections, although the application is extremely accurate in detecting cicadas, in such a way that it is used in the slovenian New Forest to track and find cicadas.

There are some examples of mixing both citizen science and gamification. A prominent example is Foldit. Foldit[8] is a revolutionary citizen science computer game in which the public is invited to help researchers predict the structures of proteins by using their puzzle-solving intuitions, playing competitively to fold the best proteins[5]. To attract the widest possible audience for the game and encourage prolonged engagement, the game has a diverse range of motivations and rewards, including short-term rewards (game score), long-term rewards (player status and rank), social praise (chats and forums), the ability to work individually or in a team, and the connection between the game and scientific outcomes[5]. The obtained results were notable. Foldit players helped decipher the crystal structure of the Mason-Pfizer monkey virus (M-PMV) retroviral protease (a scientific problem that had been unsolved for 15 years, which was solved in ten days[13]), outperformed algorithmically computed solutions[5] and achieved the first crowdsourced redesign of a protein[7].

Another source of inspiration was Safari Central, a game concept with the purpose of making wild animals a part of daily life for millions of people while generating a new revenue stream for conservation[4]. They accomplish this by using GPS data from real animals, as well as their appearance, to create an augmented reality experience, where players can interact and observe 3D model representation of wild animals in a suburban environment. Safari Central aims to be an awareness application that makes a significant effort to educate the public on wildlife conservation. We were inspired by the use of augmented reality to show 3D models of real-life animals. We then to incorporate this characteristic in our project, to boost the player's experience, making the game more enjoyable and educational.

Building on these inspiring examples, our approach takes advantage of acoustic sensing and Citizen Science to categorize, validate and map local species and biodiversity, exploiting a gameful experience to motivate users in actively be part of and contribute to the community. In fact, different research works have proved that gameful experiences are powerful tools to engage users to participate in crowdsourcing activities, including both digital task and real tasks in the urban environment, exploited as a digital playground.

3 GAME DESIGN

Aware of the challenges surrounding the collection and analysis of data for project focused on biodiversity monitoring, including cost and problematic scalability in using scientists and professionals, we embraced the GWAP framework and the involvement of players as citizen scientists, in categorizing and analyzing data, aiming at providing fun and knowledge in exchange of time and efficiency.

Initial Concept

We developed an initial concept where players could play the game both remotely and locally, this enabled players to immerse themselves in nature and classify the sounds near their place of origin. It also enabled enabled people to perform the same tasks remotely (e.g. at home). As a consequence we developed several game mechanics and elements supported and validated by a fictional context, in order to encourage gamer audience to contribute to a project that they might otherwise ignore[2]. In this process we created to scenarios that exploited the same core mechanics.

Focus Group

Upon creating the scenarios of the two game concepts, we conducted a focus group session engaging eight researchers, experts in game design, multimedia entertainment, digital media, locative storytelling, HCI (Human-Computer Interactions) and sound design. The goal of the focus group was to criticize and refine the two gameplay models. The session lasted one hour, including: the introduction of the context (biodiversity, sensors stations, Citizen Science); motivation behind the creation of a purposeful/persuasive game framework (issues we would like to address); presentation of the two game ideas (concept, goal of the players, gameplay, and so on); discussion/brainstorming on the two game ideas.

During the focus group, based on the experts' feedback, some game mechanics were refined, aimed at better meeting our objectives. After the focused group, the two games were developed into Experience Prototype sessions in order to be tested and further refined. Moreover, we gained more stability in our game concept because we started focusing on making the game elements serve the main objectives and not make the game objectives serve the game elements, which was a very strong criticism in the focus group session. In this way, we were able to attain a much-needed focus on our game's design, making the transition between the conception phase to the implementation phase, much smoother.

Experience Prototype

We evaluated our concept ideas through an experience prototyping (EP) session, which is a form of prototyping that enables developers and the players themselves to gain a first-hand appreciation of existing or future conditions through active engagement with prototypes. These gathered conditions can afterward provoke changes in the game design or validate certain choices[3].

We used a mixed methods approach to collect data during and after the experience, performing observations, personal interviews (qualitative data) and asking participants to answer a questionnaire (quantitative and quantitative data). The testing would revolve around the exploration mode, leaving out the remote game mode, which was still being designed.

We engaged 14 young-adults ranging from 21 to 30 (7 females and 7 males), 7 participants for each game scenario. We organized the EP session in the University of Madeira's Biodiversity garden, rich in plants and animals, representing a suitable location for our game experience. In particular, we focused on the animals that inhabit that environment, such as a red-footed falcon (for sound classification), a Selvagens gecko and a rock grayling butterfly, all of these species exist in Madeira (see Fig. 1).

Figure 1: The red-footed falcon (falco vespertinus)[14], the Selvagens gecko (tarentola bischoffi) and the rock grayling butterfly (hipparchia semele)[21].

The experience prototype was conducted with the use of cardboard prototypes of the visual interfaces, with which the participant could interact. We built two prototypes, one for each scenario, and compiled a set of scenario-specific tasks for the participants to complete. The researcher conducting the session acted as a storytelling voice (i.e. the scientist) assisting the user while playing, changing out the cardboard prototypes panels according to the participant's interactions with the interface and playing the audio samples through a smartphone, according to the player's selection, in the audio classification phase. Two further researchers assisted in documenting the session and conducting interviews: one was taking pictures while the other was taking observation notes and performed the interviews, after each user's session. Each session lasted an average time of 15 minutes.

We randomly assigned one scenario for the experience prototype to each participant, and, in the end, we explained the mechanics of the other scenario.

Figure 2: Depiction of the first task of the experience prototype.

At the end of each session, we interviewed the participants. We asked them seven questions, regarding the scope, the value, the mechanics, the touchpoints, the visual elements, the language and terminology used in the game experience, and suggestions to improve it. After the experience, we sent participants an online questionnaire. The questionnaire included a few general questions (name, gender, age, education); two questions about their game behaviour (how often they play and motivation for playing); five items related to interest in nature and hiking; three items related to the experience prototype session.

Through real-time observations and photographs, the brief interviews and the questionnaires that the users filled after the experiment, we obtained a significant amount of quantitative and qualitative data.

From the analysis of all the data collected we can conclude that all the participants visibly enjoyed the experience prototyping session. This is documented in the pictures taken during the trial, as users' smiles, laughter and having fun expressions (see Fig. 3). Moreover, the written accounts of the observations also support this data, for example, one observer noted: *The player looks amazed and she looks around with enthusiasm ... When she found the animal in the point of interest, she laughed and exclaimed "Awesome!".* Users found the game interesting, easy to play and with a clear scope, independently from the scenario played. This was documented through the interviews, and questionnaires. The was also a clear receptiveness to have remote based play in this game, with a majority of players wanting a remote experience instead of a location-based one.

Figure 3: Photos taken during the experience prototype sessions.

Game Design Choices

Since the initial game concept, we found it was important to contextualize our game with a story, validating certain game mechanics.

Through the Experience Prototype we found that the context we gave the participants contributed a lot to the overall experience, as well as novelty mechanics like the mutant creation. So we decided to add some detail to the story we initially created, contextualizing the data classification in a different way and giving room to proven game elements, like creating mutants, and new elements, in hopes of engaging the users even further.

We placed our game's story in the distant future. A future molded by devastating climate changes and nuclear wars, with completely warped habitats and world-wide animal extinction. The player is part of a very advanced society with access to very advanced technology, like time-travel and very little knowledge about the past. An effort is made to repopulate the environments of the present by going back to the past, to repopulate the present's habitats, by collecting and combining animal DNA and capture animals in the past, and sending them to the present. So the player is sent to the past with a very high-tech device, that he uses to perform his tasks and a team of drones to scout areas for animals and collect samples from them.

The time-travel mechanic is a very relevant mechanic in this game. If a park has bioacoustic sensors capturing data for a year, that's a year's worth of data that we still would like to get analyzed. The time travel mechanic justifies the analyzes of older data, still allowing the player to gain awareness of that time's biodiversity and enabling us go get older data classified. In the exploration mode of the game we found relevant that the player had data of the current day, it's more interesting and relevant to the player to know that two hours earlier there was a group of birds passing by this point. It gives the player more recent data to analyze and act upon. For example, if it's 4 p.m. and the player gets near a sensor and hears that at 3 p.m. there were some birds near the sensor, he might try to look for the birds or if the player notices that through consecutive visits to a specific park that there's always bat activity at 6 p.m. near a given sensor, he might just go there to see the bats. This kind of engagement is very beneficial for our project but also to enriches the player's game experience.

The drone mechanic exists for a different number of reasons. We wanted our players to gain biodiversity awareness not only in their area but also in other parts of the world, we also wanted our application to not depend on park visitants and be available to play independently of the player's location. This way there is a possibility to engage players that normally don't visit parks and hike, to do so, engaging users even further. The drones also allows us to give players older audio data to analyze, reserving the more recent audio samples to the more suitable exploration mode.

Another major factor in citizen science projects is how the data is collected and the tasks are crafted for the volunteer to complete them. The audio analysis task needed to be clear and attainable to fulfill[18], but we also needed to preemptively combat player behaviour, where the players randomly complete these tasks just to progress in the game. So we devised a way to validate user input and discouraging random these types of random behaviour, by presenting the player with pre-classified sounds and the actual sound captured by the sensor. Players that fail on classifying these sounds are penalized in their rewards, and their classification is deemed unusable or untrustworthy.

Collecting different types of creatures through capture or creation is a concept we came up with to reward the player for his actions. In essence the goal of the game is to collect every creature. Correct actions are rewarded with new creatures. Classifying sounds rewards the player with the materials necessary to create a new mutant, repopulating the environments gives the player access to unique creatures and visiting the parks allows player to capture animals in interest points.

In terms of the mutant design there were a lot of options. We could have created a specific game functionality that would splice together different types of animals according to player's characteristics, but this looked like a very troublesome process and we didn't have much time to look into this. Another alternative was use a photo editing tool and splice different types of animals together, encompassing all of the possible combinations the player could do, creating a realistic mutant. However we didn't have sufficient knowledge with any photo editing tool to achieve this. There was a possibility to limit the user's combinations, which would allow to create fewer creatures, but we weren't very inclined to limit the player in this regard. So we decided that drawing the mutants would be the best course of action. We also decided to draw the creatures in a very cartoony fashion, which simplified the process and also gave the game a more charming quality.

It's important to note that the network of bioacoustic sensors placed in different areas is completely simulated, as well as the data obtained from them. Our purpose was only to build the game in such a way that in the future it could be easily applied to a network of sensors. Incorporating real sensors in this project would be too much work so we were advised to focus on establishing a network of bioacoustic sensors in a park or focus on building the game. So we chose to simulate the bioacoustic sensors and focus on the game.

To narrow the scope of the project, we also decided to limit the types of sensors we are simulating, focusing only on sensors capable of detecting very high frequencies. We applied the knowledge obtained from the bat detective project[6] and compiled a list of possible animal species that these sensors could detect. The type of sensors typically detect bat calls, insects flying close to the sensor, really high pitch insect noises (e.g. crickets and cicadas) and artificial/mechanical noises (e.g. traffic, power tools, construction sites, etc...). This allowed us to give players a multiple choice question, comprise a list of plausible sounds for the player to analyze and narrowing the possibilities of mutant creation (with the possibility of detecting Bats, Insects and capturing Birds, the player his able to create 43 different creatures).

4 SOLUTION ARCHITECTURE

Our objective when designing the system architecture was to create a functional system that would allow us to field test the game, allowing loss of internet connection, able to support multiple players playing our game in the Android OS and download audio files from a main server.

To do this we implemented a server using a Droplet from Digital Ocean, which are a flexible Linux-based virtual machines (VMs) that run on top of virtualized hardware[11].

In terms of communication protocol, we decided that the use of HTTP would suffice, as it would allow the loss of internet connection without compromising the application side of the system.

To support the HTTP communication, we used the ok http library[12] in the application side of the system, to enable the game to send and receive HTTP messages, and implemented multiple Java Servlets supported by the Apache Tomcat Server. The Apache Tomcat Server, is an open-source Java Servlet Container developed by the Apache Software Foundation that handles incoming requests and out-going responses. The application sends a request to a specific servlet, the servlet processes the request and produces a response back to the application.

The server stores data both in a SQL Database and in folder system. Player information (username, email and password) and Sample Classifications (audio file name, username of the classifier, classification and reliability score) are stored in the SQL Database in two separate tables. The Player's Game Data and Captured Audio both stored in folder system. The Player's Game Data is a compilation of folders and JSON files with the player's progress in the game, that the application uploads to the server each the time the game closes. We did this for testing purposes, but decided to keep it in the system for it's usefulness. The Captured Audio folder system is a carefully indexed folder organized by week. In each week's folder there is an audio file with the ID of the sensor it was captured from, followed by the day it was recorded and the time (the audio file follows this format *SENSORID-yyyy_MM_DD-hh_mm.extension*). This allows us to integrate a network of audio sensors with this system, just by adding a servlet to deal with the audio uploads and take care to save the files with the appropriate name.

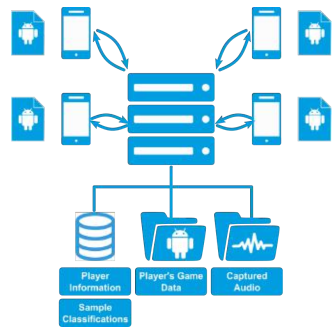

Figure 4: Basic System Architecture Diagram.

This game is very navigation based and uses gps data to track the player's current position and also to show the player different types of markers in a satellite image map. We used the Google Maps SDK for this purpose, allowing us to have these functionalities in an Android environment.

We wanted to give a new dimension to our game and establish one of the mechanics we developed in our initial concept, which revolved around searching and capturing animals in the wild. For this we created a Augmented Reality mode using Vuforia (Augmented Reality SDK) and Unity (Game engine). This enabled us to render 3D scenes through a portable smart device's camera unto a given image target. In this case we used QR codes as the image targets and used two 3D models, one butterfly and one sparrow.

The application also has the list of parks, with the latitude and longitude of the sensors and also the latitude and longitude of the interest points where there are QR codes for players to use in a specific game mode. This is the information the game uses to place the different markers in the maps.

5 IMPLEMENTATION

The game we created is a game with a purpose (GWAP) that aims to monitor biodiversity of certain species in a given area, by compiling data obtained via bioacoustic sensors and sharing that data with the players through the game. The players will then analyze and classify the data through the game's interface.

One of our goals when building this game, was to provide the players with an enjoyable and didactic experience, engaging them further in the project. We created a story for our game and developed several game mechanics and elements to achieve this.

We implemented a register and login interface allowing us to support multiple different players, keeping track of their game progress (in terms of in-game rewards but also in terms of data classification).

Figure 5: Diagram of the main game modes.

One of the main modes of play is referred to as the "Drone Dispatcher Mode". This mode is our remote-based mode. In the dispatcher screen the player can see information regarding the research points, the week in time he's currently accessing, the number of collected samples and located animals, the park he is currently searching and the state of his drones. The player can use his drones to trigger the sample classification mechanic.

The location-based mode of play is referred to as the "Exploration Mode". This mode is triggered when the player is near an area where there are bioacoustic sensors. This is know by keeping track of the player's current position relative to the different marker's location. In exploration mode the interface shows the time and place the player currently his in, as well as the corresponding data. The map also shows the player's current position. The player can always explore the map by dragging, zooming in and zooming out, and by pressing the recenter button, will be able to recenter the map according to his position. In this mode players can analyze samples, similarly to the drone dispatcher mode. The main difference is that players need to use the map and travel to the research point's location.

There's a second type of markers in Exploration Mode which we call interest points, where a player can capture animals and add them to their collection. When a player gets close to one of these markers he triggers a search mode that gives him clues to find a QR Code. When the player finds the QR code, he can point the phone's camera to it which will display a 3D Model of an animal in Augmented Reality and then capture it, adding it to the their collection.

The main mechanic of the game focuses on audio classification of animal sounds. For this we developed an interface where players can cycle through the sounds and use a cross-reference library to classify the sound that is presented to them. The cross-reference library is divided into general categories. Each category has an expandable list of sounds related to that category. The answer we collect from the player is just the category from which he selected the item, because we are aware that a taxonomic identification of a species

Figure 6: Capturing an animal through the exploration game mode.

by sound alone would not work[9]. The player is given 4 sounds to classify, but only 1 of them was the sound captured by the sensor, the other 3 are pre-classified sounds. This is a reliability test we execute to discourage random answers and assert the player's sound classification skills. From these answers, we calculate a percentage that we call a reliability score. This percentage corresponds to the percentage of correctly classified pre-classified sounds. Any score below 50% is deems the player's sample classification as unreliable.

Each time players classify sounds and/or capture animals„ they receive mutagen points, which are used to create creatures. This is a way to reward the players by properly executing their audio classifications and also encouraging players to experience nature for themselves. We gave players a large variety of animals that they can create with mutagen points as well as the possibility of being rewarded with unique creatures attainable only through the completion of certain tasks.

6 EVALUATION

In our project we had a few concerns regarding player experience and the classification portion of our game.

We wanted to prove that through a gameful experience we can engage the players to classify animal sounds, furthering biodiversity monitoring research. So we have to assess if our game provides a good experience to the player. We decided to perform preliminary usability tests with the users, to evaluate our interface's performance and our game's mechanics. We applied a thinking aloud method to perform these tests.

In a thinking aloud usability test, participants use the system while continuously verbalizing their thoughts as they move through the user interface. Through this we can assess the participants honest thoughts about the game and also check for misconceptions or misunderstandings of interface elements, which in most cases leads to a redesign of the misinterpret element. This will allow us to validate our game's mechanics, elements and overall design.

Through the natural course of playing the game, the players will evaluate some audio samples. In the end of all the sessions we can combine all player's classifications and analyze the resulting data set.

7 SESSION DESCRIPTION

The usability sessions took place in Instituto Superior Técnico (IST), in Lisbon, Alameda. We engaged 7 young-adults ranging from 20 to 26 (2 females and 5 males) to test our game. We created a game scenario where the players would go to a week in time, spanning from the 24th to the 30th of September, 2018. In this week there was only research in the Instituto Superior Técnico area with a total of 16 samples for players to classify. In exploration mode players would time travel to the current date (October 11th, 2018) where there were 2 research points, with data to collect and classify, and 2 interest points, where they could capture animals (via the augmented reality mode).

Each session started by filling out a form with the participants personal information (name, age, gender, occupation), gaming habits (how frequently do you play games?), hiking habits (how often do you hike in natural parks or similar areas?), frequency of phone use (how often do you use the phone while hiking and what do you use it for?) and what kind of navigation does the user prefer (physical maps, digital maps or both).

After the questions, we explained to the user what the thesis was about and gave a brief synopsis of the game's story, handing over the smartphone with the game in the login screen. The players registered an account and logged in to the game. We gave players a brief tutorial of the game's basic mechanics, as we intended to make an in-game tutorial and not build the game to be intuitive enough to be playable without one.

After the tutorial, we gave the player's their two goals and explained the thinking aloud method, requesting participants to continuously verbalize their thoughts and only request the researcher's help if they were stuck (due to bugs and crashes, or not knowing what to do next). After that a researcher would follow the player taking notes of what they were saying and their reactions.

When the player completed the two goals, the researcher would interview the player, asking further questions about the game (the game's story, what they enjoyed and disliked, what did they learn, suggestions for improvements or changes and if they would play the game in the future).

From these 7 sessions we collected a lot of qualitative (From both the questionnaires, the usability test and the interviews) and quantitative data (from the interviews, questionnaires and the player's sample classification).

8 RESULTS DISCUSSION

From the analysis of all the data collected (through observation and the interviews) we can conclude that parts of the game were visibly enjoyed by the users and some parts of the game were less enjoyable.

Session Observation

In the drone dispatcher mode, most of the participants didn't fully explore their options. Often staying in the same day of the week. Only 2 players time-travelled to other days.

Three of the players thought that the drone was a single use object, meaning that they used it once and thought they could not use it again. One of the players said *"Ok, so I have 0 drones, so I have to go explore"*, while another asked the researcher *"So, I can't use the drone?"*. They would see that the number of available drones was 0, but failed to realize that the drone was still in the previous spot, searching for more animals to collect samples from. This issue could be easily fixed by returning the drone to the player each time the player returns from classifying the sound. This prompted 2 of the players to go into exploration mode so they could complete their goals.

When entering the exploration mode, most of the players had no problem going to the location of the map's markers. 3 of the player's expressed difficulties in finding the markers they were looking for. The researcher wrote the following: -*The player passed by the point she was looking for and said "I'm still far away. I have to go up". Gasped. Realized she was moving away from the marker and she exclaimed "Oh no, it's further down! I'm bad at this!"*. All 3 players admitted not having a great sense of direction or following maps.

But it's also important to say that the GPS tracking of the phone was very irregular, it took a while to show the player's updated position. One player said *"I know I'm in the spot but it says I'm not."*, after a few moments his position was updated and he could resume his task.

Some of the players drew comparisons with "Pokémon GO!", as observed by the researcher: -*The player is walking to a research point. Exclaimed "This is so cool. It's like a cuter version of Pokémon GO!"*.

When players found the interest point where they could captured an animal, 6 out of 7 participants were visibly excited when seeing the 3D model of the animal. One participant talked with the animal, saying *"Hello! You're so cute! I want to catch you!"*. One participant simply exclaimed *"This is so cool!"*. Another note worthy observations was a player that already completed the goal of capturing one animal and went to capture the other one. The researcher noted:-*The player's eyes opened a lot and carefully observed the animal. He's enjoying himself. The player asked "There's another one, right? Can I go catch it?", I responded "Sure, you can do whatever you want."*.

In the research mode, 4 players failed to lock in 1 answer, reaching the end with 1 sound not classified. When this happened, the corresponding circle turned red and no other feedback was given. The players didn't understand what was going on and asked the researcher for help. The researcher told them to navigate to the red circle and lock-in an answer.

Almost all players seemed rather confused and tentative on their first sample classification task. This was noted in multiple instances by the researcher:-*The player is taking a while to lock in the answer. Looks very focused. The player asked "I can't find the matching sound. What do I do?". I explained that sometimes there is no matching sound, you just have to give your best guess.*; *The player exclaimed "There's too many options, I can't find the right one!"*; *The player asked "It's a bat! How do I answer?)"*.

It was also noted that the each time the players would do a sample classification test, they would do it faster and with less hesitation. This isn't necessarily good. Some players were actually just being more efficient and learning from past experiences: -*"I remember hearing this one! It's bat!"*. Other players seemed to just want to get it over with:-*The player is going fast. Just saying what he thinks the animal sound is and answering. Very monotone voice "Bat...Insect...Bat..."*.

We realized that the sound classification portion was not facilitating the player's task. There were too many options for the player to choose from. Part of the players got frustrated because they couldn't identify the sound being played.

All players enjoyed the process of creating new animals. One player laughed at all the animals he created as well as their names:- *Player created a mutant. Started laughing and saying "Flabeego? Because it's a bee and flamingo together, right? That's cool"*. There were different reactions to the different mutants obtained. One player obtained a mutant named "Bleedtle" and exclaimed *"Oh, he's ugly. I don't like him!"* while another participant obtained a duck exclaiming "Nice! I got a duck!. Two players after reaching their goals continued playing as they were in route to the interview spot, as they wanted to create more mutants.

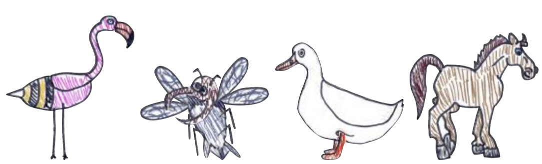

Figure 7: Some of the mutants participants obtained during the session. (Names from left to right: Flabeego, Bleedtle, Duck and Horse.)

When the participants re-populated "Parkland Forest" to it's full capacity they received a reward, which was a Horse. Players were excited with the possibility of unlocking all of the rewards. In one case the researcher noted the following:- *Player filled up the habitat. And clicked to received the reward. Player exclaimed in disbelief "I won a horse? I won a horse! I bet I can put it in the next habitat!"*. This tells us that this progression system with rewards, implemented in the habitats screen, is having it's intended effect.

Overall, player's seemed to enjoy their play testing session.

Post Session Interview

Most of the players enjoyed the game's story. Regarding the theme of nuclear war and climate change, one player said *"I like it because it's kind of a possible. Climate change, nuclear war are real threats"*, another player also stated *"It reminds of a movie. And nuclear fallout and extinction are definite concerns on everybody's mind"*. Some players enjoyed the theme of time-travel and creating mutants the most.

One player said this regarding the game's story:- *"I like the story, but I don't see it in the game. It's a cool concept but the story doesn't stand out"*. This is true. Because we were focused on proving certain aspects and mechanisms, we neglected the story portion of the game.

We also asked participants what would be their main focus when playing the game. All users said that they would devote their time to collect all the mutants and unlock the rewards.

The game seemed fairly playable, players didn't show much difficulty in perform their tasks, specially the second time around. When asked about how easy it was to play the game, players deemed it fairly easy to play (2 players stated that was sort of easy and 5 player regarded the game as easy to play).

The most enjoyable part for four players revolved around catching and/or creating animals. Two players enjoyed classifying sounds the most, despite the process not being the best. One player appreciated the fact he could choose between playing from far way (through the drone dispatcher mode) and on location (through exploration mode).

We also asked what part of the game they didn't enjoy. Has expected, there were two players who did not enjoy the sound classification portion of the game. Two players did not enjoy the exploration mode, because they couldn't navigate using the map. Two players during the session were unsure of what to do next and in the interview expressed that, saying it was their least favorite part, having no clear prompt to what to do next. And one player didn't enjoy capturing the animals in the exploring mode, for him there was a lack of excitement when capturing the animal.

Most of the participants didn't learn anything from the experience. Although, some of the participants were shocked there were so many species of bats and also the sound of bat calls also was new to some. One participant said that she heard a bat call before but taught it was a bird and not bad.

One of our objectives was indeed to educate our player base enriching them with knowledge about the environment, but most of the players did not acquire any new knowledge. We hypothesize that simulating the sounds and doing the test in a college campus, diminished the game's capacity to be informative. Also, play sessions lasted an average of 20 minutes, which also might have contributed for the lack of knowledge acquirement. We believe that longer play sessions and a more adequate test environment will help to test the game's educational factor.

Players also gave us some very useful suggestions. Like enhancing the tutorial, clearer game flow, make a simpler cross-reference library and add more animals to the game

A majority of players also expressed interest in playing the game in the future.

Sample Classification Results

Seven people contributed to the classification of 11 sounds. Out of those classifications, 7 were single contributions (meaning only one person classified them), and out of those 7 classifications, 1 was an incorrect classification and 1 was correct but had a 33% reliability score.

One of the most contributed samples had 7 responses: 6 people responded with "bat", 5 of which had 100% reliability score and 1 person had 33%. 1 person responded with "insect" with a 67% reliability score.

Another with 5 contributions had an unanimous decision of "bat", which was correct. Two people had a reliability score of 100% and three people 67%.

Two people classified a sample in the Miscellaneous category (meaning it's just artificial sound or noise pollution) correctly, both had a reliability of 67%.

One sample was wrongly classified by the majority. It had a total of 5 responses, all of the participants had 100% reliability score. 2 people responded "insect" and 3 people responded "bat". The correct answer was "insect". After examining the sound, we found out that it was a sound made by a beetle rubbing itself against the microphone and making loud clicking sounds. In the cross-reference library there was no such example. So the participants were not at fault. There was nothing similar to this sound. A possible solution is to add more variety of sounds or add a "none of the above" option.

The sample classification was somewhat inconclusive. As reliability didn't play a major factor providing better classifications and there weren't enough classifications in most cases to reach a definite conclusion.

Although, we can attest to player's capability to classify animal sounds. 9 out of 11 sounds were classified correctly. One of the incorrectly classified sounds was due to lack of comparable material and the other was wrongly classified by just one person. Moreover, the tasked was not properly presented to players and the UI was not very flexible, making players frustrated and prone to mistakes. So people seem to be up to the task if presented with the proper tools, although further testing is required.

9 CONCLUSIONS

We devised a project that would combine citizen science, a gameful experience and bioacoustic sensing, creating a game with the purpose (GWAP) of helping classify animal sounds in a given area.

We developed and tested several game mechanics and contexts, not only to properly guarantee successful data analysis, but also to provide the players a enjoyable and educational experience, boosting the participants engagement in the project.

Through a series of usability tests and interviews we assessed that game elements like mutant creation, capturing animals and the game's story contributes to a good user experience. Moreover we realized that somethings needed to change or be added (more story in the game, clearer game flow, a better tutorial, among other things). More importantly, the animal classification mechanic needs to be reworked, in order to be more enjoyable, less frustrating and more manageable for users to utilize. This will lead to better classifications and less errors.

We also concluded that players have a good capacity for sound classification, provided they have the right comparison material. Although projects like bat detective already proven that humans have an innate ability to distinguish sounds, we need a larger number of participants and classified samples to reach the same conclusion.

In the end we created a game with an interesting concept that provided player's an intriguing and fun experience. Although we couldn't acquire enough data to completely validate the audio classification portion of the game, participants enjoyed the game, with a majority stating that they would like to play a version of this game in the future.

Future Work

In terms of future work, we find it is important to do further testing in terms of audio classification. The results of the preliminary should also be addressed resulting in a redesign of certain elements and mechanics, most importantly, the redesign the audio classification interface to be more of a tool and less of an obstacle to the players. We find it is important to test and validate the educational factors of the game, by doing a longer play test session in a more viable environment (park, garden, natural reserve, etc...). We also envision using data from a real bioacoustic sensor network allowing for a more realistic view of the incoming data (both in volume and in characteristics), enabling a further adaption of the game's mechanics with the goal of not only facilitating the audio classification task, but also to maintain the player's enjoyment. The resulting data set, if accurate, would also be interesting and applicable to other studies.

REFERENCES

[1] Rick Bonney, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk. 2009. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* 59, 11 (2009), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

[2] A Bowser, D Hansen, J Preece, Y He, C Boston, and J Hammock. 2014. Gamifying citizen science: A study of two user groups. In *17th ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW 2014*. 137–140. https://doi.org/10.1145/2556420.2556502

[3] Marion Buchenau and Jane Fulton Suri. 2000. Experience Prototyping. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '00)*. ACM, New York, NY, USA, 424–433. https://doi.org/10.1145/347642.347802

[4] Safari Central. 2017. Safari Central. https://www.safaricentralgame.com/

[5] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and Foldit Players. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466, 7307 (2010), 756–760. https://doi.org/10.1038/nature09304 arXiv:doi:10.1038/nature09304

[6] Bat Detective. 2018. Bat Detective Website. https://www.batdetective.org/

[7] Christopher B. Eiben, Justin B. Siegel, Jacob B. Bale, Seth Cooper, Firas Khatib, Betty W. Shen, Foldit Players, Barry L. Stoddard, Zoran Popovic, and David Baker. 2012. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nature Biotechnology* 30, 2 (2012), 190–192. https://doi.org/10.1038/nbt.2109

[8] Foldit. 2018. Foldit Website. http://fold.it/

[9] Leska S. Fore, Kit Paulsen, and Kate O'Laughlin. 2001. Assessing the performance of volunteers in monitoring streams. *Freshwater Biology* 46, 1 (2001), 109–123. https://doi.org/10.1046/j.1365-2427.2001.00640.x

[10] iBats. 2018. iBats Website. http://www.bats.org.uk/pages/ibatsprogram.html

[11] Digital Ocean Inc. 2018. Digital Ocean's Droplets. https://www.digitalocean.com/products/droplets/

[12] Square Inc. 2018. OKHTTP github. https://github.com/square/okhttp

[13] Firas Khatib, Frank Dimaio, Seth Cooper, MacIej Kazmierczyk, Miroslaw Gilski, Szymon Krzywda, Helena Zabranska, Iva Pichova, James Thompson, Zoran Popović, Mariusz Jaskolski, and David Baker. 2010. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural and Molecular Biology* 18, 10 (2010), 1175–1177. https://doi.org/10.1038/nsmb.2119 arXiv:NIHMS150003

[14] Ron Knight. 2010. Red-footed Falcon (Falco vespertinus). https://www.flickr.com/photos/9919745@N03/8071356907 [Online; accessed October 8, 2018].

[15] Simon Asher Levin. 2013. *Biodiversity, definition of.* Elsevier Inc. 399–410 pages. https://doi.org/10.1016/B978-0-12-384719-5.00009-5

[16] Andrzej Marczewski. 2015. Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design. , 65–80 pages. http://www.amazon.co.uk/Even-Ninja-Monkeys-Like-Play/dp/1514745666/

[17] Jane McGonigal. 2010. TED 2010 | Gaming can make a better world | Jane McGonigal. https://www.ted.com/talks/jane_mcgonigal_gaming_can_make_a_better_world

[18] Chris Newman, Christina D. Buesching, and David W. Macdonald. 2003. Validating mammal monitoring methods and assessing the performance of volunteers in wildlife conservation - "Sed quis custodiet ipsos custodies?". *Biological Conservation* 113, 2 (2003), 189–197. https://doi.org/10.1016/S0006-3207(02)00374-9

[19] Chris Newman, Christina D. Buesching, and David W. Macdonald. 2012. The future of Citizen science: Emerging technologies and shifting paradigms. , 298–304 pages. https://doi.org/10.1890/110294

[20] William J Ripple, Christopher Wolf, Mauro Galetti, Thomas M Newsome, Mohammed Alamgir, Eileen Crist, Mahmoud I Mahmoud, and William F Laurance. 2017. World Scientists' Warning to Humanity: A Second Notice. *BioScience* XX, X (2017), 1–9. https://doi.org/10.1093/biosci/bix125/4605229 arXiv:1611.06654

[21] Katja Schäfer. 2005. Schmetterling des Jahres 2005: Rostbinde. http://www.bund-nrw-naturschutzstiftung.de/index.php?id=schmetterling2005 [Online; accessed October 8, 2018].

[22] Davide Zilli, Oliver Parson, Geoff V. Merrett, and Alex Rogers. 2014. A hidden Markov model-based acoustic cicada detector for crowdsourced smartphone biodiversity monitoring. *Journal of Artificial Intelligence Research* 51 (2014), 805–827. https://doi.org/10.1613/jair.4434<http://dx.doi.org/10.1613/jair.4434>).

[23] D Zilli, A Rogers, and G Merrett. 2015. *Smartphone-powered citizen science for bioacoustic monitoring*. Ph.D. Dissertation. University of Southampton. https://eprints.soton.ac.uk/382943/