An Audio File Tagging Mobile Game, mTagATune

Francisco Javier Díaz, Claudia Alejandra Queiruga, Alejandro Ferraresso, José Ignacio Larghi

LINTI - Facultad de Informática – Universidad Nacional de La Plata – 50 y 120 La Plata, Buenos Aires, Argentina
jdiaz@unlp.edu.ar,claudiaq@info.unlp.edu.ar,
aferraresso@cespi.unlp.edu.ar,jlarghi@cespi.unlp.edu.ar

Abstract. mTagATune[1] is a mobile game based on TagATune[2]. mTagATune implements the concept of GWAP[3] and seizes the capabilities and wide acceptance of current smartphones[4]. GWAP promotes the creation of computer games that encourage people to do voluntary work. mTagATune implements a game that collects information on audio files to facilitate future searches on them. By means of a collaborative game, mTagATune enables an ubiquitous collection of information on audio files that can later be used in search results.

Keywords: TagATune, GWAP, Ericsson, Mobile Applications, JAVA,

1 Introduction

Despite technological advances, computers still do not have the creativity or perception human beings have by nature. Due to this fact, computers today cannot subjectively classify certain sets of elements such as audio files, image files, etc.

Currently, data bases exist that contain thousands of audio files, although searching these repositories with subjective criteria is not possible due to the fact that each audio file would have to be tagged first with words that necessarily convey subjective meanings.

One solution to this problem is using the technique known as Human Computation [5]. This technique views the human brain as a processor inside a distributed system, where each can process a small part of a much larger computation.

Currently, there are millions of people around the world who use digital games as a form of entertainment. Many of these games can be accessed through the Internet. The massive expansion of mobile telephones allows users to play games where they could not play before: during trips or while queuing at the bank. The first games launched for mobile telephones were very simple due to physical limitations, but with the new generation of cellphones, the so-called smartphones, games are becoming more and more complex, with better performance, even using resources such as global positioning systems and the Internet.

One branch of Human Computation, called Games with a Purpose (GWAP) promotes the idea of creating games in which the activity people engage in forms part of a processing that produces information, which can later be used in other successive processings.

GWAP encourages people to do voluntary work, but not with the intention of obtaining income, as is the case of employment. If we think about the task of tagging music fragments, the amount of elements requiring processing is enormous which would require a tremendous workforce to complete it, yielding the task impractical due to cost.

Smartphones, with their many advantages, allow for the implementation of GWAP on mobile telephones, providing permanent access to the games and increasing the amount of players (and, as a result, the amount of hours dedicated to each game as well). This increases the data processed, which gives better use to the time players spend on each game.

2 Bases

As we have mentioned before, Human Computation posits the theory that the brain can be seen as a small processor inside a distributed system, where each brain can process a small part of a much larger computation. Human Computation is a technique in which a computation executes its function by delegating certain steps to humans. In traditional computation, humans use a computer to solve a problem: the human provides the computer with a formalized description of the problem and receives a solution they must interpret. Human Computation reverts the roles; the computer asks a human or group of humans to solve a problem and collects, interprets and integrates their solutions.

GWAP is a combination of the Human Computation technique and the billions of people around the world willing to invest time playing online. The concept of GWAP could be defined as games in which each participant processes a part of a larger computation, which is solved by combining the processings contributed by each player. In this context, "processing" makes reference to the mental exercise the player engages in to solve the part of the computation they are assigned. Some examples of GWAP can be found in work by Luis von Ahn[6]; the Google engine has an experimental version of a GWAP for the classification of images, called Google Image Labeler[7]. The goal of the Google Image Labeler game is to generate tags associated to images that can be used to improve image search results. The mechanism of the game consists of showing two users an image, and for every match, both users get points. These points motivate users to input a large amount of tags. Afterwards, when a word has been entered many times for the same image, it is assumed that it describes the image correctly and fit to be used in the search engine. Note that in this mechanism, players have no knowledge of who is their team mate during the game and have no way to communicate, thus it is impossible for users to cheat by agreeing on the words they will use.

3 Description of mTagATune

mTagATune[1], mobile TagATune[2], is a GWAP application for mobile devices, smartphones specifically, based on TagATune.

The growing trend in the use of mobile devices and the advantages they offer encourages an environment adequate for the application of GWAP in mobile phones, making it possible for people to play in more places, thus increasing the amount of data published.

3.1 How the Game Works

mTagATune is a mobile application that implements the concept of GWAP and allows for audio file tagging. The goal of mTagATune is to collect semantic information to be used in search results and further indexation.

When a user enters the game, after registering, they are assigned a partner. Because this type of application cannot ensure that the user is paying attention, once the partner is selected, both are asked to confirm that they are ready to begin the game. In case one of the users takes longer than stipulated to confirm presence, both users will be informed that the game has been canceled.

When the game begins, each participant is given an audio entry and both have to contribute words that describe it. Based on the descriptions entered by both participants, each must determine individually whether they are both listening to the same piece or not. If both participants choose the right answer, they obtain points. The goal of the game is to obtain as many points as possible.

3.2 Pair Selection

Participants pairs are selected at the beginning of each game without the users knowing who is paired with them. Following is an explanation of the mechanism implemented for this purpose.

Players are picked together on the basis of similar amounts of points. For this purpose, it is necessary to limit scoring differences, for example, if a user obtains 500 points and the limit for a game is 100 points, this user will be able to play with others that have between 400 and 600 points. The main disadvantage of this system is the delay in finding a suitable match for a certain player, which is greater the more reduced the amount of users. To solve this problem, it was decided that the limit increased with time, which allows for a greater range and increases the possibility of forming pairs. A waiting time limit was also introduced – if the limit is reached and there are no matches, the player is assigned a partner no matter their score. If there are no other players waiting, a game is created exclusively for the user. To avoid pairing new users, which might discourage them because both might enter a small amount of tags, it was determined that inexperienced users should be paired with players that have a long history of games. This way, the chances of winning games, and therefore gaining interest in the game, are greatly increased.

3.3 Description of a Game

First, the game waits for both users to download the audio file to their devices for two reasons: because both must discover whether they are listening to the same file at the same time, and because the length of the track determines the time the user is given to enter tags and decide whether it is the same fragment as their partner's. Once each user has their file, the round begins. During the round, the player must enter words that represent what they hear. As they do, the words are sent to their partner and shown in a fraction of the screen, so both players have real-time access to the words entered by each. When the track ends, no more words can be added.

Once the file has ended, players are given a few seconds to determine whether their fragment was the same as their partner's. Players will receive points each time both pick the right option.

Thus, if one or both get the wrong answer, non of them will receive points during that round. Figure 1 shows the screen players see during each round, which shows the tags both entered together with the file data and options.

It was decided that the player who got the right answer even though their partner did not would not receive points either because the goal of the game is to achieve cooperation and not competitiveness. This way, each player would have to concentrate on getting the right answer and in describing their file as well as possible to increase their chances of the other player getting the right answer as well.

A player might decide whether they are listening to the same file before the file ends, which may block out data entry and reduce the possibility of getting the same answer for both players. Although this option may make the player enter less words, notice that these cases add a new piece of information: the instant in which the choice was made.

In some cases, one player will enter tags that are completely opposite to the tags the other generates, in which case they will need no further proof to decide. Therefore, a later processing of the information could determine which words potentially express the opposite to the way in which the file is classified. At all times during the round, both players know whether their partner has decided.

Once both players have made their choice, the score they get this round is show on screen, as well as whether the answer each player gave was correct or incorrect. If one of the players chooses not provide an answer, the system will assume that they gave an incorrect answer and none of them will obtain points.

Fig. 1. Game in progress

mTagATune is a collaborative, non-competitive game in which players only receive points if both get the right answer. mTagATune gives a natural incentive for players to enter data that correctly describes the audio file. If it were a competitive game by, for example, granting points to the player who gets the right answer even if their partner does not, players would be motivated to win by harming their partner. This would cause them to enter wrong and malicious data to confuse their partner and make them pick the wrong answer, which would result in wrong data due to the implicit competitive nature of the game.

3.4 Scoring System

Players are given points in the following manner:

When both players answer correctly for the first time, each player gets 60 points. The second time, they get 70 points each and 80 if they answer correctly a third time. This way, two players that answer correctly three times during a game will obtain 210 points each. These correct answers do not necessarily have to be consecutive, that is, if a pair gets the first fragment right (and receives 60 points), fails to provide a correct answer for the second fragment and does so for the third one, both players will receive 70 points for the second correct answer in the game, earning a total of 130 points each.

The aim of this scoring system is to stimulate the attention of the user throughout the game, as a user that answers correctly in the three rounds of a game will receive more points than a user who answers correctly in three rounds from different games. In the first case, they will receive 210 points, while in the second, they will obtain 180 points. In this way, users to maintain a good performance throughout the game obtain more benefits.

3.5 Bonus Round

Another way of playing is what is known as bonus round. The bonus round is activated during a common game, when both players get the three rounds right, thus obtaining the maximum score for a game.

This round does not generate tags on the files given, but serves to create a relationship between them. When both players get the three rounds of a common game rights, they are automatically notified that they can take part in a bonus round (they can turn down the offer).

If both players agree to participate in the bonus round, the system will select three audio files that will be played for the users. When the files end, the users have 10 seconds to decide which of the three fragments is the most dissimilar, if both coincide in their choice, they get 50 points. Figure 2 shows a screen in the bonus round, which shows the controls players can use to select answer.

Fig. 2. Screen of a bonus round

3.6 Single-Player Mode

mTagATune has a single-player mode for when a player enters the game when there are no other players to pair them with, or the total number of players is even, making it impossible for the system to assign a partner for the player.

The single-player mode allows for a single user to start a game at any time, independently from the amount of users connected at the time. This mode is transparent to the user, as the place of the other player is occupied by a bot that reproduces a series of rounds that have already been played by real users. The bot is an algorithm that reproduces a player's behavior in a previous round.

If there is no available partner for the player upon entering the game, the player selects a saved game depending on the level of experience of the user. For rounds to be reproduced, it is a fundamental prerequisite that they resulted in a positive

outcome, i.e., both players coincided in their choices for each round and that their choices were right. There has to have been a bonus round in the game as well. This is necessary for the system to be able to assume that the entered tags can be considered valid.

Once the the rounds are assigned, the user is notified that they have a partner, a bot (the user will never know that they are actually playing with a bot). During the course of each round, the bot enters the tags in the same sequence in which they were entered by the emulated user. Once the audio fragment has been reproduced in its entirety, the game evaluates the tags entered by the real user and determines whether it is the same track in both. To do this, the bot analyzes the percentage of matches between the set of tags entered by their partner and the set of tags for the audio fragment they have.

The bonus round in a single-player game is also based on a game that was stored beforehand. The game will select a saved bonus round and the bot will choose the same options the original player chose.

An important advantage is that the result of a game of this modality is just as productive as the results of ordinary games. The player will choose whether they are listening to the same fragment as their partner based on the tags entered by the bot, which is in turn based on the actions of a real user. On the other hand, the bot will take their decision based on a comparison of the tags entered by a real user and those provided earlier by another real user.

3.7 Technologies

mTagATune is a mobile application written in Java for Android phones that consists of a client and a server. The full development is based on open, free use technologies. mTagATune uses Tomcat to contain the Java servlets in charge of handling the logic and data storage. PostgreSQL was used to store data and Hibernate was used to map objects to the relational database.

For the purpose of communication between the server and its clients, mTagATune uses a Server Push mechanism called Comet [10] that allows for information to be sent asynchronously from the server to the clients. The Comet implementation was CometD, developed by the Dojo Foundation, which implements the Comet mechanism with Jetty Continuations. To handle the messages, CometD uses the Bayeux protocol, which sends the messages through named channels. These messages can be sent from the server to the client, from the client to the server or among servers, using these channels.

For the serialization of the Bayeux messages and the domain objects, mTagATune uses JSON. This is a light data exchange format, easy to read and write for humans, and easy to interpret and generate for machines.

The client was developed using Java for the Android operating system for many reasons, the first being its free license as well as that of the tools used for the development, and also because of the wide acceptance it has gained during the past year. The architecture used allows for the development of clients for other operating systems, such as iOS, and for users of different devices and operating systems to interact in a game.

4 Adapting TagATune for Mobile Devices

mTagATune is an adaptation of TagATune for mobile devices. mTagATune introduces some modifications to TagATune that allow it to improve the gaming experience and its performance in mobile devices. Following, we will describe the aspects of TagATune that were modified for mTagATune.

Given that the quality of Internet access in mobile devices can vary greatly, we decided that the audio files should be downloaded separately before each round starts, and each round should begin only after each file has been successfully downloaded. Although this characteristic introduces waiting time at the beginning of the rounds, it allows us to be sure that the file has been successfully downloaded and is available for full reproduction throughout the round. This also releases the connection for sending information about the round itself, so that, if the quality of the connection is low, it can be played correctly. Because mobile device screens are smaller and of lower resolution as compared to those of desktop computers, it was necessary to redesing the screens to accommodate all the elements of the user interface, eliminating unnecesary ones. Figure 3 shows a screen of a TagATune game.

Fig. 3. Pantalla de una partida en TagATune

If we compare the TagATune screen to that of mTagATune, we can clearly see many changes that go beyond aesthetics. The daily high scores list was deleted together with the play and volume controls. This simplification was aimed at focusing the attention of the player on producing tags only. The warning saying that the partner has already chosen was replaced with a pop-up, and in its place we put the voting buttons. Lastly, the remaining time, score and progress bar were located in a single

line to reduce the space they took up. Similar changes were introduced in the remaining screens, such as the bonus screen and the previous games screen. Another problem with mobile devices are keyboards. Firstly, the average typing speed is significantly lower than that attained with desktop computers. To mitigate this issue, we modified a feature in the original game–in TagATune, players were supposed to make up their minds during the audio reproduction, while in mTagATune, we decided to add 5 seconds of extra time before this action is allowed, so that during the game the players can focus on writing, using the extra time to analyze their partner's input and make the right choice. Secondly, many devices do not have physical keyboards, but virtual ones inside the screen itself. This is a setback for the player, which is why we have considered a redesign of the interface to integrate this kind of devices.

Because it is to be expected that players will receive incoming calls or text messages during a game, we decided to add a confirmation related to the user paying attention to the game right before each round starts. This confirmation has a timeout and, if it's reached without confirmation from the player, the game is cancelled and the partner is notified, so that they are not stalled waiting for the game to resume and can begin a new game.

The great popularity of the Android mobile operating system, its user community the proposed open development model and the agile application distribution mechanism through its online store, "Android Market", provides an optimal media for the development and validation of the results. This was the hypothesis upon which we based our decision to choose Android [8] over iOS[9], another leader in the mobile device market.

5 Conclusiones

This work shows a way to obtain information about audio files that is automatically validated by the very agreement of the players on the subject, a concept that constitutes a step towards simplicity and improvement over the manual mechanisms that are currently used for tasks such as labelling music.

With mTagATune, we demonstrate that the concept of GWAP, particularly that of TagATune, is applicable to mobile devices, so long as they have the characteristics of a smartphone, although in this particular case we have only used those with Android as their operating system.

It was also shown that this adaptation has no negative impact in quality or playability. Although changes were made in the application architecture, in some playability aspects and in the user interface, none of these changes has a negative impact on the player or represents a problem for the normal development of a game.

This adaptation for mobile devices takes the concept of GWAP to a completely accepted environment in countries with wide access to mobile technology, something that will soon come to developing countries such as our own. This will increase the amount of time users can spend on the game. The fact that these games are available for mobile devices makes it possible to play them during free time, something that was unthinkable a few years ago because they were only available for conventional computers.

We are currently planning to test the usability of mTagATune to test its performance in other mobile devices, together with an evaluation of other mobile applications such as AvatarFacedget, a widget that allows users to publish single avatars or avatar galleries in social networking sites.

References

1. Díaz J., Queiruga C., Ferraresso A., Larghi J.: "mTagATune: mobile TagATune". In Proceedings of ICMB 2011, 10th International Conference on Mobile Business, Como, Italia, June 2011. http://www.mbusiness2011.org/
2. Law E., Von Ahn L, Dannenberg R. and Crawford M: "Tagatune: a game for music and sound annotation". In Proceedings of the 8th International Conference on Music Information Retreival, Vienna, Austria, 2007.
3. Von Ahn L.: "Games with a purpose". In IEEE Computer Magazine, June 2006. Pages 96-98.
4. Smartphones: http://es.wikipedia.org/wiki/Smartphone
5. Von Ahn L.: "Human computation". In K-CAP '07, Proceedings of the 4th international conference on Knowledge capture, 2007.
6. Von Ahn L.:. http://www.cs.cmu.edu/~biglou/
7. Google Image Labeler, http://images.google.com/imagelabeler/
8. Android. http://developer.android.com
9. iPhone OS. http://es.wikipedia.org/wiki/IPhone_OS
10. Gravelle, Rob. "Comet Programming: Using Ajax to Simulate Server Push", http://www.webreference.com/programming/javascript/rg28/.