

LAHAJET: A GAME FOR CLASSIFYING DIALECTAL ARABIC SPEECH

Waed Hakouz, Abdurrahman Ghanem, Samantha Wray, Ahmed Ali

Qatar Computing Research Institute, HBKU
Doha, Qatar

waedhakouz@hotmail.com, swray@email.arizona.edu, {abghanem, amali}@qf.org.qa

ABSTRACT

We present Lahajet, a game with purpose (GWAP) for crowdsourcing classifications of different varieties of Dialectal Arabic in multi-dialectal audio. Players listen to short audio clips and select a character representing the dialect they have heard. Character movement, point rewards, and obstacles are implemented to ensure player engagement and interest. Lahajet tracks multiple player submissions to obtain multiple judgments for each audio file, while utilizing a single-player implementation that allows players to participate even in the absence of additional online players. Classifications gathered from the GWAP contribute to the building of a multi-dialectal speech corpus of Dialectal Arabic.

Index Terms— Game with a purpose, crowdsourcing, dialect classification, Arabic

1. INTRODUCTION

Arabic is a language with great dialectal variety, with Modern Standard Arabic (MSA) as the dialect of limited formal domains. The colloquial varieties differentiated by region are collectively known as Dialectal Arabic (DA). Despite the widespread usage of DA, computational resources for these dialects are scarce, especially in the speech domain.

In this paper we present our game **Lahajet** as a crowdsourcing solution for building DA speech resources by having players classify segments of multi-dialectal broadcast audio streams into DA varieties.

Lahajet is designed as a game with purpose (GWAP), an entertaining activity for players that serves a computational function[1]. Players engage with rapidly moving avatars of characters who represent the four major regional DA groups (Egyptian, North African, Gulf, Levantine) while short audio clips automatically play. Players make classifications by tapping characters who represent the dialect they have heard in each clip, while avoiding obstacles which enter the playscreen.

1.1. Data Pre-processing

Audio for player classification is pulled from QCRI Advanced Transcription System (QATS)[2] which is deployed at Al-Jazeera Arabic website¹. Initially, Voice Activation Detection (VAD) is applied to remove non-speech segments. Then, speaker diarization, and speaker linking are performed to determine when a new speaker begins, and the audio is cut using silence and speaker linking information. More details about data preparation can be found here[3]. Finally, each audio clip is initially assigned a dialect as explained in [4] to create a baseline from which to provide in-game feedback.



Fig. 1. Lahajet Gameplay

We gratefully acknowledge Al Jazeera for their contributions to this research. Thanks are also due to Felix Stahlberg and Yasser Makram for their valuable assistance, and to Tasneem Toghoj for game artwork and graphics.

¹aljazeera.net

2. IMPLEMENTATION

2.1. Backend Design

The server receives initial dialect classifications from the client. Each time a player classifies an audio clip, the classification is sent back to the server and the tally for the selected dialect is incremented for that audio clip. If the clip is a golden standard segment, the client determines if the classification matches the initial dialect assessment, and provides the player with feedback regarding a match or mismatch. Note that for any non-gold standard segment, the classification done by the player is always considered correct by the client as the initial dialect is still unknown. The server aggregates the classifications for each audio segment, and marks it done if it matches one of the following two conditions whichever comes first:

- Minimum of three different players with inter-player agreement of 70% is reached.
- Maximum of nine different players have classified the same file irrespective of the inter-player agreement.

Same approach was used in [3] in their crowdsourcing experiment.

2.2. Game Design

Lahajet aims to give players an enjoyable fast-paced game where they only need a short amount of time to identify a certain dialect, and to be the type of game where one could simply play for amusement. Therefore, it was built to be very easy to understand and play, fun to play and engage in, and amusing to look at. Lahajet was built using Unity Game Engine "Fruit Ninja" template². This enabled the fast creation of a fun playing environment with internet connection, menus, character movements, backgrounds, scoring, life losing and regeneration, and sound playing and streaming.

The first time a player begins to play Lahajet, s/he can register on the server to store their points. As the game begins the player is given a set of three lives (represented by hearts), as well as a counter which tallies points, and characters and obstacles start flying across the screen. The four characters representing the four dialect groups as well as obstacles in play are shown in Figure (1). When an audio clip plays, the player has three options: to identify the dialect and tap the character representing it, wait until the clip finishes and moves to next clip, or tap the "skip" button to move to the next clip immediately in case the audio clip is unintelligible. When the player taps on a character, scoring is done based on the initial dialect assessment of the clip that is played. If it matches, the player is rewarded with three points, otherwise the player will lose a life. If the player clicks on an obstacle instead of a character, the player will lose a life. Once all

lives are consumed, the collected points are added to the main player score and the player is taken to a menu where s/he can choose to play again or review the rules. Gameplay states are shown in Figure 2.

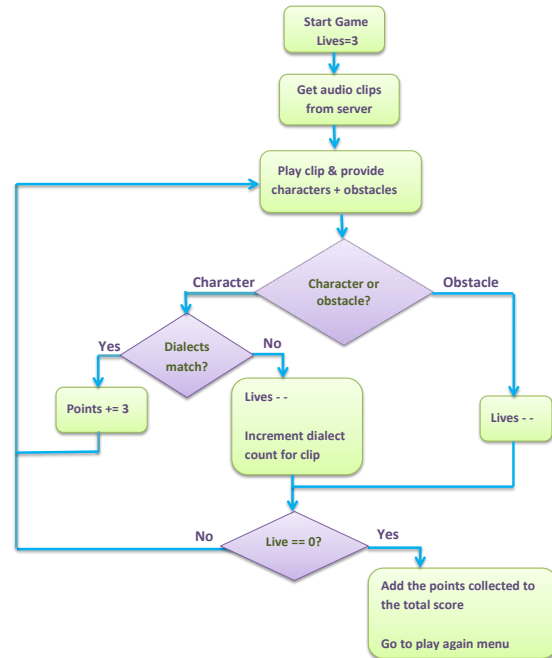


Fig. 2. Overview of gameplay architecture

3. LAHAJET DEMO

The Lahajet demo showcases the game in its entirety. The demo will consist of a hands-on opportunity to try out the game, and will serve as a platform for researchers interested in novel GWAP-based approaches to solving data sparseness problems in speech, as well as a chance to explore innovative possibilities of non-standard dialect speech usage in games. The demo will be featured on iPhone and iPad, which are the devices for which the game is currently available. Support for other platforms is future work for this project.

4. REFERENCES

- [1] L. Von Ahn, "Games with a purpose," *Computer*, vol. 39, no. 6, pp. 92–94, 2006.
- [2] Ahmed Ali, Yifan Zhang, and Stephan Vogel, "Qcri advanced transcription system (qats)," in *SLT*, 2014.
- [3] S. Wray and A. Ali, "Crowdsource a little to label a lot: Labeling a speech corpus of Dialectal Arabic," in *Inter-speech*, 2015.
- [4] Ahmed Ali, Peter Bell, and Steve Renals, "Automatic dialect detection in arabic broadcast speech," *arXiv preprint arXiv:1509.06928*, 2015.

²unity3d.com