A Voice User Interface for football event tagging applications

Silvio Barra
DIETI, Univ. of Naples, Federico II
Naples, Italy
silvio.barra@unina.it

Alessandro Carcangiu
FootureLab
Sassari, Italy
alessandro.carcangiu@abinsula.com

Salvatore Carta
University of Cagliari
Cagliari, Italy
salvatore@unica.it

Alessandro Sebastian Podda
University of Cagliari
Cagliari, Italy
sebastianpodda@unica.it

Daniele Riboni
University of Cagliari
Cagliari, Italy
riboni@unica.it

Figure 1: Voice User Interface for football matches event tagging.

ABSTRACT

Manual event tagging may be a very long and stressful activity, due the monotonous operations involved. This is particularly true when dealing with online video tagging, as for football matches, in which the burden of events to tag can consist of many thousands of actions, according to the desired level of granularity. In this work we describe an actual solution, developed for an existing football match tagging application, in which the GUI has been enhanced and integrated with a Voice User Interface, aiming at reducing tagging time and error rate. Empirical tests have revealed the efficiency and the benefits brought by the developed solution.

A VI '20, September 28-October 2, 2020, Salerno, Italy

ACM ISBN 978-1-4503-7535-1/20/09.
https://doi.org/10.1145/3399715.3399967

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Accessibility**.

KEYWORDS

Integrated Voice User Interface, Match tagging, Online Video Tagging

ACM Reference Format:
Silvio Barra, Alessandro Carcangiu, Salvatore Carta, Alessandro Sebastian Podda, and Daniele Riboni. 2020. A Voice User Interface for football event tagging applications. In *International Conference on Advanced Visual Interfaces (AVI '20), September 28-October 2, 2020, Salerno, Italy*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3399715.3399967

1 INTRODUCTION

Graphic User Interfaces (GUI) have drastically improved the interactions between users and machines, so becoming a *must* for almost the totality of the developed software applications [4]. Albeit some of them are extremely customised on the typology of the interacting

users [3], they still present some drawbacks. Firstly, individuals with physical disabilities may find it difficult to access a classic GUI application, designed, as in the specific case, for a standard usage with mouse and keyboard[6]. Secondly, protracted and repetitive actions may be demotivating and stressful for the users, causing them to make mistakes, bouguettaya2015efficient. Many complex and focused tests have been indeed designed for evaluating their usability [2].

In such a context, the use of a Voice User Interfaces (VUI) may help to reduce the impact of the aforementioned issues. However, entirely mapping a GUI into a VUI is not always a linear operation, since there might exist specific interaction patterns which can be harder to address through the voice interface. Usually, these limitations mainly rely on the level of rigidity imposed by the VUI. In this paper, we describe a voice user interface integrated into a GUI, specifically enhanced for this purpose, aiming at fastening and improving usability in tagging events in football matches. The application has been called FooTAPP (Football event Tagging Applications) and its interface is shown in Fig. 1.

2 FOOTBALL MATCH EVENT TAGGING

Football event tagging aims at labelling all the events happening in a football match [5]. The final goal of the activity is not only the production of a detailed report about the movements of each singular player (acting in its own role as defender, midfielder, forward, winger or goalkeeper) during the match, but, in parallel, to collect team dynamics, in order to analyse forwarding and defending tactic movements. Therefore, it often happens to label a single action under two different points of view: (i) single player actions, expressed as *tag combinations* (e.g., MIDFIELDER > KILLER PASS > POSITIVE), and (ii) team actions (actions which involve more players, e.g. COUNTERATTACK). Moreover, in addition to active plays (i.e., those whose object is the ball), there are a plenty of passive plays which need to be labelled as well, in order to properly track the team dynamics, such as double teaming, defending actions, and man-to-man/zone marking. Often, this activity must be done for both teams, in order to provide a full set of tags. Considering that a football match lasts ninety minutes (excluding additional minutes and extra times), the entire tagging activities can take from 6 to 8 hours for, on average, about 2,000 events.

In view of the foregoing, the aim of the integrated VUI interface is reducing the time needed for event tagging. This target is reached by applying a voice interface, so to fasten up the tagging process, and by adapting the original GUI, in order to facilitate a combined mode (voice plus touch) usage. Currently, FooTAPP GUI is organized as shown in Fig. 1. We describe its components in what follows:

- the top left block (contoured in red in Fig. 1) contains a video player, in which the match is reproduced; directional arrows ease the forward and backward skips of the video;
- the bottom left block (contoured in cyan) shows a virtual field in which the team lineups and shape are shown; the buttons with player numbers are clickable, to be inserted in a tag combination;
- the top right block (contoured in orange) contains a summary of the tag combination records already registered for the current match;

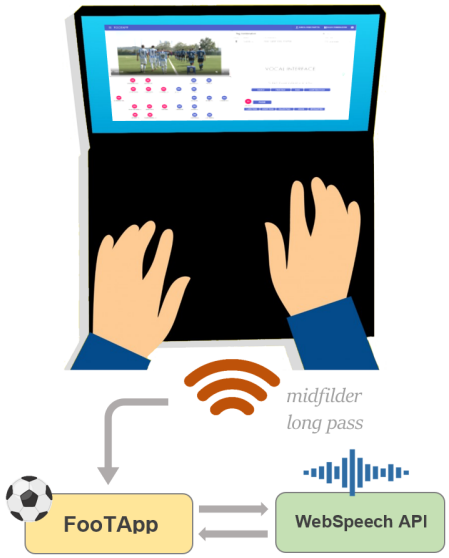

Figure 2: High-level overview of the integrated VUI.

- the middle right block (contoured in pink) contains the voice interface, which automatically activates when the user casts a command;
- the block contoured in green shows the keyboard which activates as the user selects a player from the virtual field or spells his number (first level tags);
- the block contoured in yellow contains additional details associated to the selected event. This block pops up when a main event is activated, and could potentially provide a large number of second/third level options.

3 VUI INTEGRATION AND EMPIRICAL TESTS

The voice user interface has been developed by using the Web Speech API [1], which defines a complex interface (*SpeechRecognition*), providing a set of methods for transforming speech into text.

A high-level overview of the implemented VUI structure is shown in Fig. 2. From preliminary empirical tests, we noticed that total migration to a VUI would have partially obtained the expected benefits (actually, an average reduction of the 13% of the tagging time). Therefore, we adapt the original interface to optimize the use of a combined tagging mode (voice and touch), resulting in an average reduction of the 28% of the time (almost 2 hours for each full tagged match). This happens since both: (i) the processing time of too long voice commands (which, despite the robustness of the exploited API, can last two or more seconds), could negatively impact on the global time saving, and (ii) some tagging patterns still result quicker in the manual mode (as an example, the selection of the player from the virtual field in the bottom left block of Figure 1), whereas other tagging practices (as an example, finding second/third level tags among many options) can obtain real benefits from the voice interface.

REFERENCES

[1] Julius Adorf. 2013. Web speech API. *KTH Royal Institute of Technology* (2013).

[2] Silvio Barra, Rita Francese, and Michele Risi. 2019. Automating Mockup-Based Usability Testing on the Mobile Device. In *International Conference on Green, Pervasive, and Cloud Computing*. Springer, 128–143.

[3] Gordon P Kurtenbach. 1999. Graphical user interface for defining and invoking user-customized tool shelf execution sequence. US Patent 5,867,163.

[4] B. A. Myers. 1989. User-interface tools: introduction and survey. *IEEE Software* 6, 1 (1989), 15–23.

[5] Rahul Anand Sharma, Vineet Gandhi, Visesh Chari, and CV Jawahar. 2017. Automatic analysis of broadcast football videos using contextual priors. *Signal, Image and Video Processing* 11, 1 (2017), 171–178.

[6] Amber Wagner and Jeff Gray. 2015. An empirical evaluation of a vocal user interface for programming by voice. *International Journal of Information Technologies and Systems Approach (IJITSA)* 8, 2 (2015), 47–63.