Personalized Training in Fast-Food Restaurants Using Augmented Reality Glasses

Mehmet Selçuk Albayrak, Alper Öner, Idris Murat Atakli
Innovation and Transformation Department
ATA Technology Platforms
Istanbul, Turkey
{mehmeta, alpero, idrisa} @atp.com.tr

Hazım Kemal Ekenel
Department of Computer Engineering
Istanbul Technical University
Istanbul, Turkey
ekenel@itu.edu.tr

Abstract—Employee training in fast-food restaurants is a long, practice-based process, which is mainly done on the job. Employee performance during training directly affects service quality and customer satisfaction. In this study, we optimize the training process in fast-food restaurants with the use of augmented reality glasses. We increase the quality of the training by gamifying and personalizing the process. We also aim to help businesses lower the labor costs by shortening or possibly fully eliminating one-on-one training sessions instructed by senior employees. We implemented a new training system on a game engine with the aid of various sensors such as gyroscope, accelerometer and audio & visual inputs. We also intend to eliminate the culture of high employee turnover rates of the fast-food sector by increasing the quality of the training. The outcome of this study will be used to fully train and supervise employees in the future.

Keywords—Personalized employee training, augmented reality, gamification, marker based object detection, fast-food restaurant.

I. INTRODUCTION

Quick service restaurants - also known as fast-food restaurants - have become a sector with a market size of 668 billion dollars in 2018 worldwide. Between 2013 and 2018, this sector has grown 3.5% annually and 1% by the number of restaurants. The sector, which employs more than 12 million people worldwide, is growing by 2.4% per year in the means of human resource [1]. In an industry, where the number of employees is so high, increasing the training and work qualities of employees is of significant importance for businesses.

Today, it takes 296 hours of training for a fast-food employee to be fully equipped for the back of the house [2]. It is observed that the new recruits in this sector are both slow and make a lot of mistakes during the training period. High amount of errors directly affect sales and customer satisfaction. In particular, the speed of the staff, the completion of the work with the minimum rate of error in the product, and the consistency in product quality are important factors in terms of sales and customer loyalty. Therefore, the importance of correct personnel training planning is inevitable.

The size of the fast-food market in Turkey is 2.4 billion USD [3]. Biggest fast-food chain operating in Turkey trains about 10,000 employees every year and most of this training is done on the job while serving the customers. When the employee’s efficiency and speed are not at the desired level, customer satisfaction is negatively impacted. This negative impact causes high stress on employees forcing them to quit work prematurely, resulting in high employee turnover rates.

Considering high number of turnover rates and poor training processes in fast-food sector, the importance of benefitting from technology in employee training drastically increases. In this study, augmented reality glasses are used to train the fast-food kitchen employees. This approach aims to speed up the learning process as well as making it more efficient. By increasing the efficiency, it is intended to boost employee satisfaction, service quality, and to lower labor costs for businesses by decreasing the need for a senior employee to be used as an instructor.

This study aims to pave the way in employee training process in fast-food sector using an augmented reality based solution. While some fast-food chains utilize PCs or tablets to train new recruits via educational videos and to track their improvements; they still fall behind when it comes to the quality of the training and the engagement of the employees. We provide a concrete solution for employee training using augmented reality glasses with a gamification tool, as well as a selection method for the current hardware and a detailed description of compatible software.

In the following section, we go through the concept of augmented reality, how similar technologies are used in the restaurant sector and the importance of gamification in training. In section III, we explain the components of the proposed training system. Moreover, we give detailed information about possible complexities of selected frameworks and suggest potential alternatives. In section IV, we describe the developed training, personalization, and gamification processes, step by step. Finally, in section V, we convey the benefits of the study, chosen methods, and tools.

II. RELATED WORK

A. Augmented Reality

Augmented Reality (AR) is the reality, in which virtual additions are made on images taken from the real world. The pioneering work on head-mounted displays was initiated by Ivan Sutherland in the 1960s [4]. In 1968, with his augmented reality glasses, called the Sword of Damocles, Ivan Sutherland
design the first head-mounted display. He is also seen as the founder of the augmented reality with the name passing through the literature or in his own words: ultimate display. Later on, this research and the likes have seen much interest in defense industry, education, entertainment industry, and beyond. In 1998, the first AR game was developed. Since the early 2000s, the US armed forces have begun taking advantage of a simulation environment -namely MR MOUT- in which, a combination of real and virtual elements are used in urban combat training [5].

Since the very beginning, AR is seen as a promising medium for industrial training. Even, the coiners of the name augmented reality, Caudell and Mizell developed an AR training program for an aircraft factory providing virtual diagrams to be overlapped and fixed on a specific position of a real-world object. The inventors aim was to reduce costs and improve efficiency by decreasing the need for customary items like paper manual and computer-aided instructions [7]. Further examples can be found in the case studies of Wang et al.[8].

In another study, Baird and Barfield compared four types of medium -paper manual, computer-aided instructions, opaque augmented reality display, and see-through augmented reality display- to instruct the subjects to perform a manual assembly task. Their study shows a performance increase with use of an augmented reality display. Both the assembly times and the error rates decrease with use of such displays compared to paper manual and computer-aided instructions [7]. Further examples can be found in the case studies of Wang et al.[8].

B. Similar Technologies in Restaurants

Augmented Reality is a subtopic of Mixed Reality (MR), which is the name given to all levels of reality-virtuality transition between the real world and the virtual world. As MR technologies became more widespread on mobile devices, it became more applicable to restaurant industry. In a study which was published in one of the most famous restaurant industry magazines of America, the design of the restaurant has been made in a way that can be examined with virtual reality glasses. With this approach, an extra 300 man-hours were saved, compared to traditional design methodologies [9].

A start-up has developed an application that shows the restaurant menus in 3D with AR technology using customers’ mobile devices. In this application, customers first identify the markers on the plates of the restaurant. Then they can see a 3D model of the desired product from the restaurant menu on the dish through their phones or tablets [10].

In another study, augmented reality technology was employed to help restaurant chefs determine serving sizes in a more accurate manner [11]. Participating chefs for this study were divided into 3 groups. While no information about the correct food portions was given to one of the groups, the second group was spoken orally about the portion sizes and the third group was asked to measure the portions using AR application. The third group has emerged as the group with least amount of errors in all error margins.

C. Gamification in Training

Gamification is the use of game elements in non-game contexts. It became a popular technique to improve outcome of both organizational and educational contexts. In [15], authors test a model in the gamification context by assigning trainees to read scenarios describing gamified instruction or traditional PowerPoint instruction in a random order and assessing their training valence. It is concluded that among potential learners with high experience, gamification produces better outcomes than PowerPoint instructions.

Good training programs have also been linked to improved levels of job satisfaction [12], in the hospitality and restaurant industries [13], giving restaurants the best evidence yet that investments such as MR assisted training stands as a way to better nourishing a wider group of talent, using different gamification methodologies.

In restaurant industry, where training is dominated by a sink-or-swim mentality, MR offers nontraditional training opportunities without risking the businesses as all the training can be done in a virtual environment. For instance, an MR-based training program with 360-degree screening capability can help trainees see all around them, with audio and/or visual cues giving detail to important sections of their work environment.

According to a study published on Future Workforce [14], 77% of millennials were willing to use MR technologies in their professional life. Considering the increase of millennials in the workforce, these technologies will potentially be widespread in the near future.

III. System Architecture

Our flow of the training process through AR glasses is shown in Figure 1. This flow is designed to obtain a burger as in Figure 1(e), by using the AR glasses with given features in Figure 1(a). The triggers shown in Figure 1(b) are used for invoking the functions in Figure 1(d) to play or repeat the videos in Figure 1(c). The used components of the glasses include WiFi module, motion sensor, camera, and a headphone input, which provides to access voice of the video and transmits speech commands via microphone. Alternatively the employee may give commands by head gestures or marker based object detection methods. Then, the user controls the training video by these commands. Finally, all preparation process of the burger is presented by the AR glasses.

AR glasses will contribute a significant improvement on employee training in near future but today there are some issues for real time implementations. AR glasses usually have low computational power and RAM capacity due to their size limitations. On the other hand, AR glasses, which provide high computational power, cost a lot of money. Therefore, selection of a suitable AR glasses gains importance on various business cases. Furthermore, usage of proper frameworks for an architecture is the most crucial part of AR glasses. This section gives the architecture of a sample solution for an employee training with AR glasses.
A. Selection of the Hardware

We examined 6 AR glasses, which provide desired performance specifications. These are: Microsoft HoloLens\(^1\), Vuzix M300\(^2\), ODG R-7\(^3\), DAQRI\(^4\), RealWear HMT-1\(^5\) and Epson Moverio BT-300\(^6\). We have made a performance comparison for prices and specifications such as CPU & RAM according to a selected computer and mobile phone. This comparison was made based on the performance of a specific framework, which ran on the specified computer and mobile phone. The framework uses the same amount of computing resource needed when employee training occurs in real time on AR glasses. We observed the CPU and RAM usage ratios while the framework was performing similar operations. Performance evaluation values were calculated for CPU & RAM of AR glasses by a Bayesian Inference method, which utilizes usage ratios of the framework. Here "success" means that the program should run smoothly. Furthermore, \(P(CPU)\) points out the performance ratio of the CPU among other devices according to AI Benchmark software [16]. If we take into consideration these values, then \(P(T_C)\) could be obtained as a normalization parameter. Finally, \(P(RAM/T_R)\), \(P(T_R/RAM)\), \(P(RAM)\), \(P(T_R)\) indicate same meanings for the RAM respectively. Equations (1) & (2) show the formulas used for the Bayesian Inference.

\[
P(CPU/F_C) = \frac{P(F_C/CPU)P(CPU)}{P(F_C)} \quad (1)
\]

\[
P(RAM/F_R) = \frac{P(F_R/RAM)P(RAM)}{P(F_R)} \quad (2)
\]

We examined various AR glasses, which could contribute to the employee training operation in restaurant kitchens. As a result of this Epson Moverio BT-300 is selected by the Bayesian Inference algorithm. We have considered computational power and price performance with the purpose of determining the right AR glasses.

B. Programming the Training

The training process on AR glasses is like a food preparation game. There are many steps to create the final product, e.g. a burger. Hence, we used a game engine to program the flow. The engine we used is Unity3D [17] one of the most powerful game engines, which is compatible with almost any platform. That way, we eliminated any future hardship that may be encountered in case of a hardware change. Unity3D also has a very wide range of assets available through its asset store and thanks to its reputation various frameworks support the engine. These features enables users to develop high-quality 2D, 3D, and VR/AR games. We used it to script our flow, control the hardware sensors, display instructions, play the training videos, and of course, to guide user to the ingredients that will be used. Figure 2 shows the flow of the training procedure.

C. Software Frameworks on the AR Glasses

While preparing a burger, there are more than a few steps to the final product. In this study, we segmented each step of adding a new ingredient to the burger. Since the trainee’s hands will be occupied while preparing the burger, we considered three trigger alternatives to move to the next part of the process. These alternative methods are:

- Speech to Text
- Head Gestures
- Marker Based Object Detection

These methods can be implemented with various frameworks. We have chosen the best ones by the smoothness and compatibility with the selected AR glasses.

1) Speech to Text: First method we have implemented is speech to text, the trainee is instructed to start the process with a "Start" command, then at the end of each video played, the trainee has two options to continue to next part or to repeat the last seen video. At the end of the last video, trainee can restart the application with a voice command. There are two obstacles of this method, first is the used natural language processing cannot be done on the device but via a cloud

\(^1\)https://www.microsoft.com/en-us/hololens
\(^2\)https://www.vuzix.com/Products/m300-smart-glasses
\(^3\)https://www.osterhoutgroup.com/r-7-smartglasses
\(^4\)https://www.daqri.com/products/smart-glasses/
\(^5\)https://www.realwear.com/products
\(^6\)https://tech.moverio.epson.com/en/bt-300/
Fig. 2: Flow of the training

service, in our case IBM Watson speech to text service [18]. Cloud services obviously require a stable internet connection. Secondly, fast-food restaurants are loud environments. Stable internet connection and noise cancelling are not the strongest features of today’s AR glasses.

2) Head Gestures: Second method that we have implemented is head gestures. In this method, the trainee is instructed to make precise head gestures to continue to the next part of the training or to repeat the last viewed part. The chosen AR glasses have enough sensors to detect such movements natively. The trainee is explicitly reminded the gestures between every step, a ”YES” nod as seen in Figure 3(a) is used to start the next video, a ”NO” nod as seen in Figure 3(b) is used to replay the last video and at the end of the process a nod to the sides as seen in Figure 3(c) would restart the process from the beginning.

Fig. 3: Head gestures

The main difficulty in this method is that it is unsustainable since certain muscle groups are repetitively used. This repetitions would cause trainees develop severe neck pain, which is another drawback of this method.

3) Marker-based Object Detection: A third method has been implemented, since the first two methods had some drawbacks. We considered an object detection trigger or a hand gesture recognition but since today’s AR glasses usually have low computational power and RAM capacity, a more simplistic method is used: Marker-based Object Detection. We benefited from a popular marker based AR framework, Vuforia [19], to detect two markers placed on the trainee’s gloves. One that says ”PLAY” is placed on the right hand of the trainee, it is used to play the next video, while another marker placed on the left hand saying ”REPEAT” is used to replay the last video.

IV. EMPLOYEE TRAINING WITH AR GLASSES

In this section, we elaborate on the training procedure for burger preparation through AR glasses. It is important to keep in mind that training procedure is assumed to be done on the job, in a restaurant environment, by a new employee (trainee), without assistance of an experienced staff, with AR glasses worn by the trainee connected to the system.

A. Flow of the Training

The process starts with the trainee wearing the AR glasses. When a burger order comes, the application notifies the user to begin the burger preparation procedure. Trainee starts the procedure by using one of the triggers from Section III-C:

- Speech to Text
- Head Gestures
- Marker Based Object Detection

Fig. 4: In-app images

Start of the training with marker-based object detection is shown on Figure 4(a). When the trainee uses the selected trigger, system guides the trainee towards the place where the next ingredient will be found, an example is shown in Figure 4(b) for lettuce. Then, after giving the trainee couple of seconds to find or acquire the ingredient, the instructive video starts to play. In the video, the trainee is shown how much or how many of each ingredient should be added and
how it should be added to the product. A screenshot of the mayonnaise spread video can be seen in Figure 4(c). When the video ends, trainee is alerted with two options: to replay the instruction video again or to continue to the next step, with the chosen trigger. If the video is replayed, the trainee returns to the same screen. If he/she chooses to continue, he/she is guided through the next ingredient. This process is repeated until the product is completed. At the end of the video stream, user has the option to repeat the last video or return to the start screen as it can be seen in Figure 4(d).

B. Gamification and Personalization of the Training

We also designed a gamification module that goes along with our personalized training system to help increase productivity and employee motivation in work environment. Our gamification module is based on different metrics such as food preparation speed, average number of delivered burgers, quality of burgers -based on customer feedback, hygiene, and other workspace metrics -based on supervisor feedback.

In our gamification system, employees earn points as they reach different goals. Every goal an employee reaches earns him/her a different badge. Some possible badges: two months without a complaint (kindness badge), reaching certain number of successfully delivered burgers (seniority badge), decreasing the average time of burger delivery to a certain interval (speed badge) and beyond.

We designed our module to incorporate different levels such as trainee, beginner, junior, senior, and expert. As employees receive enough points and badges, they progress to higher levels. Each employee is able to follow their progress and earned points on the game screen and they can convert their earned points to giftcards.

There is also a leader board for each and every level of expertise. These boards are open to all employees see the rankings across the organization to increase engagement and competition.

The main goal of our gamification tool is to engage and inspire employees to interact and to enhance their engagement and their learning process.

V. Conclusion and Future Work

In this study, we developed a training program using augmented reality glasses for fast-food restaurant employees. By gamifying, personalizing, and shortening the training process, we aimed to make the on-boarding of new employees to workspace easier and more pleasant.

As a result, we expect an increase in employee satisfaction and service quality. Also, businesses are anticipated to benefit financially both by quickly adapting the new employees to the workforce and by decreasing the need for a senior employee to instruct trainees.

We are continuing on this work with improvements on key performance indicator tracking, employee monitoring, and a full system integration. We are also planning to use deep learning as a trigger or a guidance tool in the future versions.

The AR glasses and the game engine that we preferred are advantageous as they are vastly compatible with other software and hardware options, respectively. As new and more advanced AR glasses are made available in the market, it will be much easier for us to develop more comprehensive solutions without the constraints of today’s AR glasses.

ACKNOWLEDGMENT

We would like to thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) for supporting this project within the scope of project #3170763 "Fast Food Preparation Training for Personnel with Augmented Reality and Image Processing".

REFERENCES
