Entertainment Computing 5 (2014) 83–90

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: ees.elsevier.com/entcom

Real-time sensory substitution to enable players who are blind to play video games using whole body gestures ☆

Tony Morelli [a,*], Eelke Folmer [b,1]

[a] Central Michigan University, Pearce Hall 413, Mt. Pleasant, MI 48859, United States
[b] University of Nevada, Reno/0171, Reno, NV 89577-0171, United States

ARTICLE INFO

Article history:
Received 17 October 2011
Revised 3 July 2013
Accepted 23 August 2013
Available online 3 September 2013

Keywords:
Accessibility
Natural interfaces
Gestures
Exergames
Visual impairments
Blind
Haptics
3D interaction

ABSTRACT

Gesture-based interaction adds a new level of immersion to video games, but players who are blind are unable to play them as these games use visual cues to indicate what gesture to provide and when. Though visual cues can be substituted with audio or haptic cues, this often requires access to the source code, which is not attainable for commercial games. We present a solution that uses real-time video analysis to detect the presence of a particular visual cue, which is then substituted with a vibrotactile cue that is provided with an external controller. A user study with 28 sighted participants with a popular commercial gesture based game showed no significant difference in performance between visual and vibrotactile feedback. A follow up study with seven visually impaired participants revealed no significant difference in performance between both groups of users. Both studies demonstrate the feasibility of real-time sensory substitution as a cost-effective approach for making gesture-based video games accessible to players who are blind.

1. Introduction

The way we interact with software is increasingly modeled after how we interact with the real world; as such interaction is most natural to us. The emergence of more immersive and healthier forms of interaction – through the use of whole-body gestures – has propelled video gaming to the cutting edge of human computer interaction design. All console manufacturers offer gesture-based interaction that use computer vision (Sony EyeToy, Microsoft Kinect) or computer vision combined with inertial sensing using a handheld controller (Nintendo Wii, Playstation Move) (Fig. 1).

Because gesture based games are intuitive to play, they have successfully attracted the elderly [1] and they also facilitate more social forms of gaming [2]. Despite their appeal to non-traditional gamers, gesture-based video games are inaccessible to players who are blind as they require the player to perceive visual cues that indicate what gesture to provide and when. Access to gesture based games, exergames in particular, could create new exercise [3,4] and socialization opportunities for users who are blind, which is important as: (1) they suffer from higher levels of obesity due to fewer opportunities to be physically active [5]; and (2) users who are blind are often isolated and lonely [6].

Video games can be made accessible to players who are blind using sensory substitution [7], e.g., replacing visual cues with non-visual, i.e, audio or tactile cues [8]. Previous research found that gesture based exergames can be made accessible using vibrotactile cues [3,4], however, computer vision based gesture recognition systems are controller-less and to implement sensory substitution access to the source code is required, which is not attainable for commercial games. This paper presents real-time sensory substitution, a technique for sensory substitution without modifying the source code of a game. This paper is organized into seven sections. First, background and related work will be discussed, Section 2 discusses real-time sensory substitution. Section 3 presents two different user studies. Section 4 presents the results of the user studies. Section 5 discusses these results, Section 6 outlines future work and the paper is concluded in Section 7.

Gesture-based video games typically simulate real physical activities, because they use whole body gestures. A physical activity, such as tennis, typically involves spatial (where to hit the ball) and temporal challenges (when to hit the ball). Performing spatial and temporal challenges relies upon a combination of sensorimotor (senses and motor coordination) and cerebellar (muscle) control [9,10]. Sensorimotor-based physical activities are difficult to perform for players with visual impairments, as they are mostly

☆ This paper has been recommended for acceptance by Matthias Rauterberg.

* Corresponding author. Tel.: +1 989 774 3005; fax: +1 989 774 3728.
E-mail addresses: tony.morelli@cmich.edu (T. Morelli), efolmer@cse.unr.edu (E. Folmer).
[1] Tel.: +1 775 784 6974; fax: +1 775 784 1877.

1875-9521/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.entcom.2013.08.003

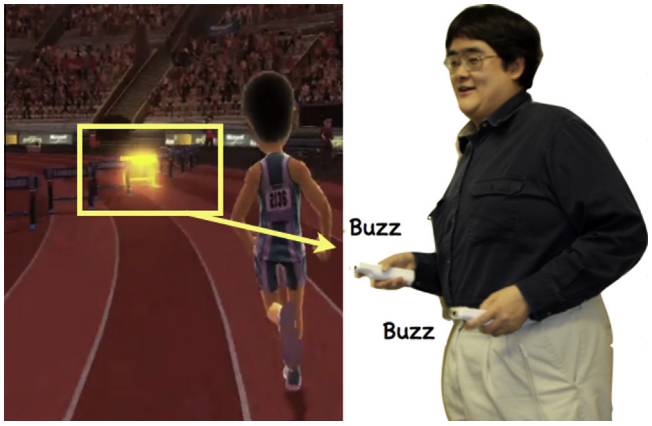

Fig. 1. A legally blind player (right) playing Kinect Hurdles game (left) where visual cues that indicate when to jump are detected using real time video analysis and substituted with vibrotactile cues that are provided with a handheld controller.

visio-spatial [11]. For example, in tennis to successfully hit the ball, which is a combination of spatial and temporal challenges, the location of the ball is predominantly acquired visually, which – when you are visually impaired – may be difficult or impossible to achieve.

Gesture-based games typically involve similar temporal and spatial challenges as the physical activity that they simulate. Some gesture-based games only involve a temporal challenge, for example Wii Sports Tennis [12], only involves swinging the Wii remote at the right time but the direction in which the Wii remote is swung is not taken into account, to keep the game simple to play. Other exergames such as Eye Toy Kinetic [13] involve spatio-temporal challenges as this game superimposes virtual objects at random locations to be punched and kicked over a video image of the player using an external camera. Targets are defined in space in front of the player and player has to provide directed gestures aimed at these targets. Because video games and gesture based video games primarily use visual cues to indicate what input and what gesture to provide they are inaccessible to players with visual impairments. This limitation affects players who are legally blind and totally blind more than players who are partially sighted or those with low vision, as these individuals have been identified to be able to play existing exergames through small modifications, such as increasing the contrast or volume of the game [14].

Gesture-based interaction using non-visual modalities has been explored in the following approaches. Finger gestures have been defined that allow users with visual impairments to interact with mobile touch screens [15,16], where primarily audio feedback is provided, but these approaches don't use whole body gestures or involve games where fast responses are required. AudiOdyssey [17] is a music game in which players receive audio instructions that indicate what gestures to provide using a motion-sensing controller (Wii Remote), as to create and record musical beats. The player can then layer these recordings to create complex musical tracks. Providing the right gesture as indicated using an audio cue is primarily a temporal challenge.

In previous work we explored how to make gesture based games, specifically exergames, playable using non-visual feedback. VI Tennis [3] (see Fig. 2:left) implements the gameplay of a popular upper-body tennis exergame (Wii Sports Tennis) that is played with a Wii remote. Wii Sports Tennis only involves performing a temporal challenge, e.g., when the ball approaches, the player has a few seconds to provide the upper-body gesture, e.g., swing their Wii remote from back to forward like a tennis racket to return the ball. VI Tennis implements the same audio feedback as Wii Sports Tennis, though vibrotactile cues provided with a Wii remote are used to convey the location of the ball. Because this game requires fast responses, simple vibrotactile patterns were used to increase their correct identification. A 250 ms vibrotactile cue, provided with a fixed frequency of 250 Hz indicates the bouncing of the ball to indicate to the player to prepare for returning the ball, and a 2000 ms cue indicates the time frame in which the player must provide their gesture.

Fig. 2. Individuals who are blind playing VI Tennis (left); and VI Bowling (right).