Benefits of immersive spatial audio on athlete reaction times in e-sports

Alex Brandmeyer
alex.brandmeyer@dolby.com
Dolby Laboratories
San Francisco, CA, USA

Mark Hertensteiner*
Dolby Laboratories
San Francisco, CA, USA

Nathan Swedlow
Dolby Laboratories
San Francisco, CA, USA

Poppy Crum
Dolby Laboratories
San Francisco, CA, USA

ABSTRACT

E-sports athletes compete in a wide variety of environments where the technologies used during gameplay directly influence the speed of player reactions and, in turn, the results of the competition. This is well-understood in the context of video display components (e.g. monitors, graphics cards). The impact of audio technologies on player reaction times has received less attention, but is highly relevant, particularly in the context of first-person 3D gaming environments. Here, audio has the ability to inform athletes of important events occurring outside the current field of view. The present study assessed player performance metrics including reaction time in a simple first-person gaming environment. Standard stereo audio presentation was used alongside two commercially available immersive spatial audio solutions. Athletes completed a training round plus three additional rounds of gameplay. It was observed that the immersive spatial audio solutions led to improved reaction times relative to stereo presentation, and that there were differences in the relative benefits of the two immersive playback solutions. These results are discussed in the context of their implications for competitive e-sports technology considerations, and in particular with respect to the use of standardized versus personalized head-related transfer functions (HRTFs).

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; • **Hardware** → **Sound-based input / output**.

KEYWORDS

e-sports, spatial audio, HRTF, reaction time

ACM Reference Format:

Alex Brandmeyer, Nathan Swedlow, Mark Hertensteiner, and Poppy Crum. 2021. Benefits of immersive spatial audio on athlete reaction times in e-sports. In *ACM CHI Workshop. EHPHCI: Esports and High Performance HCI*, May 08–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

Recreational gaming and competitive e-sports represent a highly significant and growing sector of the home entertainment market. As the underlying technologies (consoles, PCs) that support gaming have scaled in capacity (e.g. CPU, GPU, RAM, DSP, controllers/interfaces, display technologies, etc.), game developers have been able to leverage this increased capacity to create richer worlds for athletes to engage with and different forms of innovative gameplay. Additionally, generational improvements in individual technology components such as displays and graphics processors have enhanced and contributed to the development of competitive e-sports, where reduced latencies and increased field of view can enhance individual reaction times and improve overall gameplay.

While much of the focus in the past decades has centered on visual components of the athlete's experience, the past decade has seen an increasing emphasis placed on the role of sound in informing and enhancing gameplay. As with real-world environments, game audio provides crucial information about events taking place both on- and off-screen and guides the athlete's actions and visual focus to the most relevant parts of the gaming environment. Moreover, spatial audio has become an integral part of the contemporary gaming landscape. Current PC and console gaming support spatial audio formats ranging from stereo through 5/7.1 surround and immersive formats that support rendering of individual 3D object positions. These additions allow for game design that informs athletes of key events, such as 'approaching footsteps behind and above me to the left', and facilitate optimal athlete reactions to enable a richer, more impactful gaming experience.

Importantly, headphones represent the primary means by which athletes experience gameplay audio. In turn, the delivery of surround and immersive-type spatial audio depends on the use of binaural rendering ('headphone virtualization') techniques to simulate the acoustic cues that inform our real-world perception of sound sources in the 3D environment [8]. Such techniques typically rely on measurements of the head-related transfer function (HRTF) to capture the acoustic interactions of sound sources with

*Current affiliation is Magic Leap, Sunnyvale, CA, USA

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Yokohama '21, May 08–13, 2021, Yokohama, Japan

© 2021 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/1122445.1122456>

Figure 1: Screen shots of active gameplay plus athlete overlay.

the listener’s individual anatomy (i.e. the ears, head and torso) as sound travels from different positions in the environment and stimulates the sensory periphery at the right and left ears. Crucially, every individual has a unique HRTF, due to individual physiological differences (i.e. shape and size of the ears, head and body) [6]. This fact has important consequences for the effectiveness of individual headphone virtualizer technologies and their benefits for competitive gameplay.

In the remainder of this paper we present a study evaluating the relative benefits for e-sports athlete reaction times of two immersive headphone technologies alongside stereo headphone playback. We also go on to discuss the implications of differences in performance of the rendering technologies for the use of spatial headphone audio technologies in competitive e-sports settings.

2 METHODS

The study evaluated the impact of different headphone-based spatial audio conditions on e-sports athlete reaction-times and performance in a simplified 3D gaming scenario. 14 competitive e-sports athletes were recruited in the San Francisco Bay Area to take part in a one-day experiment lasting approximately one hour. All athletes were students at a local university and members of the university e-sports program. The gaming scenario was developed in the Unity environment using a first-person perspective and delivered via an HTC Vive VR HMD and Sennheiser HD650 headphones. The experiment was run on a Windows 10 PC in a dedicated laboratory space. Athletes made use of the HTC Vive controllers during gameplay.

Athletes were presented with a sequence of ‘planets’ which spawned at randomized positions in the 360 space (following a uniform distribution in the upper hemifield). Each planet was associated with a spatially localized audio signal consisting of white noise plus a high-harmonic content waveform generated using a saw wave. This harmonic component gradually increased in frequency as the planet moved towards the athlete. The audio stimulus was interspersed with silent intervals to create bursts. Athletes were tasked with destroying the planet by shooting it before the planet reached the athlete and caused damage and a loss of health points (see Figure 1). Gameplay followed a progressive ramp in difficulty, with planet spawn frequency and movement speed increasing with longer gameplay duration. The burst rate and speed of the frequency change of the audio signal for individual planets also

increased as the game progressed in difficulty. Individual gameplay rounds ended when the athlete’s health points reached zero. In each gameplay round, a number of metrics were tracked for each spawned ‘planet’: spawn location relative to athlete gaze direction, hit/miss and reaction time (time between spawn and hits).

Athletes completed a total of four gameplay rounds during the session. The first gameplay round made use of stereo audio playback and served as a familiarization/training round (Stereo Training). Stereo audio was also presented to users in the third gameplay round (Stereo 2). The second and fourth gameplay rounds made use of two different commercially available immersive 3D headphone technologies (‘Immersive A’ and ‘Immersive B’). Both of the headphone technologies were accessible using the Microsoft Windows Spatial Sound functionality and were capable of rendering of rendering immersive channel-based audio formats containing both surround and height channels. Game audio was rendered to a 7.1.4 channel layout using a Wwise plug-in integrated into the Unity environment, which in turn was rendered binaurally using Windows Spatial Sound. The order of the two immersive 3D headphone technologies was counterbalanced across participants (i.e. one half were presented with A followed by B, the other with B followed by A).

Analysis of the resulting data focused on the following: hits, misses, hits/miss ratio, and athlete reaction times for ‘hits’ with a focus on targets spawning outside the athlete’s visual field (i.e. off-screen). This was done by narrowing the analysis to include individual targets that spawned at absolute visual angles greater than 75 degrees relative to the athlete’s gaze direction. Statistical comparisons of individual metrics across audio conditions were carried out using pair-wise non-parametric permutation tests (permutation distribution derived using 10000 randomizations of the data) [5]. A sigmoidal curve-fitting procedure was used to model cumulative reaction time distributions across the four gameplay rounds.

3 RESULTS

Analysis results for athlete reaction times are presented in Figure 2A and B, while results for hits, misses and hit/miss ratio are shown in Figures 3A, B and C respectively. Statistically significant differences were observed across gameplay audio conditions for both reaction time data as well as hits. With respect to the latter, it was

Figure 2: Analysis of reaction time data across gameplay rounds: Panel A illustrates mean reaction time for 'hits' (i.e. for destroyed targets) across gameplay rounds with error bars indicating standard error across all observations. Brackets and asterisks indicate statistically significant pair-wise comparison (non-parametric permutation test based on 10000 randomizations) using the following convention: * $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$. Panel B presents cumulative distributions of the data for each gameplay round.

Figure 3: Gameplay performance results: boxplots illustrate distribution of results for each of the four gameplay rounds across individual subjects. Panel A shows hits, panel B shows misses and panel C shows the hit/miss ratio. Upper and lower whiskers indicate upper and lower quartiles, rectangles show 25-75% range, horizontal line shows median value. '+' indicates outliers based on 3 times the standard deviation. Brackets and asterisks indicate statistically significant pair-wise comparison (non-parametric permutation test based on 10000 randomizations) using the following convention: ** $p < .01$, *** $p < .001$

observed that the overall number of hits increased in the final three gameplay rounds (Immersive A, Immersive B and Stereo 2) from the first gameplay round ('stereo training'), and likely reflects the familiarization of the athletes with the game following training.

For the reaction time data, the following results were observed: similarly to the improvement in hit rate, reaction times were significantly reduced from the first gameplay round in the following

three gameplay rounds (Stereo 2, Immersive A, Immersive B). Additionally, reaction times were significantly lower in the Immersive A condition relative to the Stereo 2 (mean difference: 102 ms) and Immersive B conditions (mean difference 61 ms).

4 DISCUSSION

The present results highlight an important emerging area within e-sports and gaming: the role of audio technology in informing competitive gameplay. The data demonstrate an impact of the binaural audio rendering technology used during gameplay on in-game reaction times. This is consistent with previous results showing task performance benefits for spatial audio in gaming [4]. As game studios further integrate spatial audio into their development and leverage it in defining interactive gameplay elements, differences in headphone virtualizer solutions can potentially impact competitive aspects of gameplay. Split second differences in reaction times can mean the difference between winning and losing in the context of professional e-sports.

Both immersive headphone solutions in the study showed benefits over stereo presentation, highlighting the growing importance of spatial audio in competitive gameplay. Moreover, the differences between immersive solutions should also be considered. The underlying algorithms can potentially differ in a number of aspects that influence reaction times, including algorithm latency, spatial fidelity (e.g the number of discrete positions that can be represented in the HRTF model) and spectral fidelity (e.g the types and resolution of the individual filters employed by the HRTF model) [2, 3].

An important consideration which was highlighted in the introduction is the role of the employed HRTF. While many current headphone virtualization technologies make use of a single, 'one-size-fits-all' HRTF, it is well understood that the perceived spatial locations will vary across different individuals when using non-individualized HRTFs [7]. This is due to individual differences in HRTFs [1, 9] and has important consequences for the use of immersive audio in competitive gameplay scenarios. For instance, a given athlete may experience reduced benefits when using a headphone virtualizer technology due to the fact that the HRTF employed by the virtualizer does not provide a good match to their individual physiology, leading to inaccurate spatial cues in the rendered audio. This in turn raises important considerations for the use of immersive headphone technology in the context of integrity and excellence in competitive e-sports, where the aim is to 'level the playing field' such that individual athletes are not unfairly disadvantaged.

To conclude, the present results highlight the potential benefits immersive audio can provide to e-sports athletes with respect to reaction times in competitive gameplay. They also highlight additional considerations around the differential benefits that individual headphone virtualizer technologies can introduce into gameplay, and the impact this could have in competitive environments.

REFERENCES

- [1] Ralph V. Algazi, Richard O. Duda, Dennis M. Thompson, and Carlos Avendano. 2001. The CIPIC HRTF Database. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*. IEEE, New York, NY, 99–102.
- [2] Areti Andreopoulou and Brian FG Katz. 2018. Comparing the effect of HRTF processing techniques on perceptual quality ratings. *Proceedings of the Audio Engineering Society* 144 (May 2018).
- [3] Jeroen Breebaart, Fabian Nater, and Armin Kohlrausch. 2010. Spectral and spatial parameter resolution requirements for parametric filter-bank-based HRTF processing. *Journal of the Audio Engineering Society* 58, 3 (2010), 126–140.
- [4] Brent Cowan, Bill Kapralos, and Karen Collins. 2020. Does Improved Sound Rendering Increase Player Performance? A Graph-Based Spatial Sound Framework. *IEEE Transactions on Games* (2020).
- [5] J.J. Higgins. 2004. *An introduction to modern nonparametric statistics*. Brooks/Cole, Pacific Grove, CA.
- [6] Kazuhiro Iida. 2019. *Head-Related Transfer Function and Acoustic Virtual Reality*. Springer Singapore.
- [7] Chung-eun Kim, Veranika Lim, and Lorenzo Picinali. 2020. Investigation Into Consistency of Subjective and Objective Perceptual Selection of Non-individual Head-Related Transfer Functions. *Journal of the Audio Engineering Society* 68, 11 (2020), 819–831.
- [8] Bosun Xie. 2013. *Head-related transfer function and virtual auditory display*. J. Ross Publishing, Plantation, FL.
- [9] Harald Ziegelwanger, Piotr Majdak, and Wolfgang Kreuzer. 2015. Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization. *Journal of the Acoustical Society of America* 138, 1 (2015), 208–222.