An Immersive Voice Over IP Service to Wireless Gaming: User Study and Impact of Virtual World Mobility

Ying Peng Que
Smart Internet CRC
Telecommunications and Information Technology Research Institute
University of Wollongong, Australia
ypq01@uow.edu.au

Farzad Safaei
Smart Internet CRC
Telecommunications and Information Technology Research Institute
University of Wollongong, Australia
farzad@uow.edu.au

Paul Boustead
Smart Internet CRC
Telecommunications and Information Technology Research Institute
University of Wollongong, Australia
paul@titr.uow.edu.au

ABSTRACT
The current generation of VoIP services offer only single channel (mono) party-mix of voices. Thus when added to a network game engine, mono VoIP service can not really contribute to the gamers' sense of virtual world immersion. As a future development, the Immersive VoIP service delivers to each gamer an Auditory Scene, mixing the live voices of surrounding gamers which are all directionally placed and distance attenuated according to the appropriate virtual world positions. In previous work, we have proposed the Wireless Immersive Communication Environment (WICE) which is special type of Immersive VoIP service tailored to the scarce resources of wireless gaming clients. Also in previous work, we have already addressed the processing scalability of WICE with the conjecture of distance-governed relaxation of *acceptable angular errors* which allows multiple gamers to share the voice localised along the same direction. In this work, we verify the conjecture of *acceptable angular errors* in a series of subjective listening tests. An important test finding is the apparent user sensitivity to angular shifts in voice localisation when such auditory movements correspond to little or no visual movements. This finding suggests that the re-establishment of Auditory Scenes across time can not be memoryless in the face of gamer mobility in the virtual world. A mechanism has thus been developed to address this issue and analytic results are obtained on the impact of virtual world mobility on the required execution frequency of such mechanism.

Categories and Subject Descriptors
C2.1 [**Network Architecture and Design**] Wireless Communications; G. 1.6 [**Optimisation**]: Integer and linear programming

General Terms
Algorithms, Measurement, Performance

Keywords
Wireless Immersive Communication Environment (WICE), Immersive Voice Over IP Service, Voice Processing Reduction, Auditory Scene Creation.

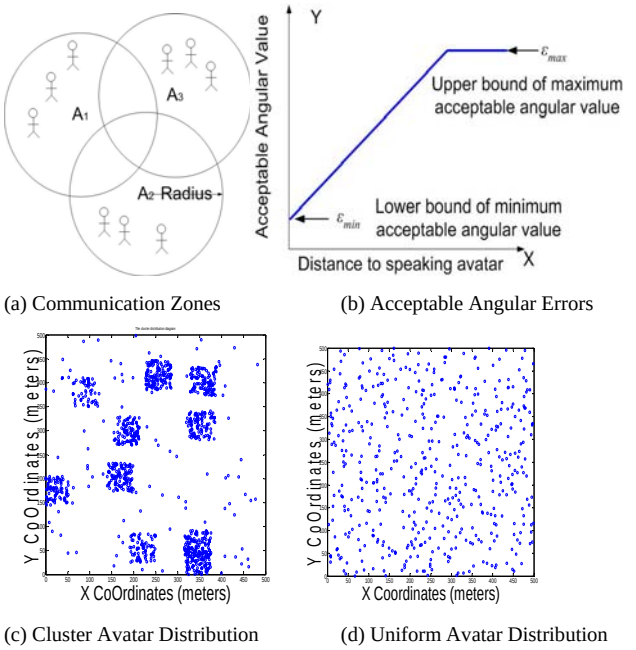
NetGames'07, September 19-20, 2007, Melbourne, Australia.

1. INTRODUCTION
Recently, there has been an increasing trend in adding Voice Over IP (VoIP) service to support the networked game engines, as epitomised by the deployment of Xbox Live [1]. In a virtual world such as that of a Massively Multi-player Online Games (MMOG) [2], each user is represented by an *avatar* in the virtual world. In our work, we assume each avatar has a one to one correspondence with a particular gaming client (wired or wireless) in the physical world. As shown in Fig. 1a, a given avatar has a *communication zone* which encloses all the neighbouring avatars that particular avatar communicates with. One of the main attractions of online games is the ability to enhance the social user interactions with a sense of immersion in the virtual world [3]. As shown in [4], the current generation of single channel (mono) VoIP services such as Xbox LIVE, seems disconnected from the visual game scenes and can not contribute to the gamers' sense of immersion. In our prior work [5, 6], we have proposed to deliver the Immersive VoIP service to complement the visual game scenes and further enhance the gamers' sense of immersion. In Immersive VoIP service, a personalised *Auditory Scene* is created for each avatar which is the mixture of all the voices heard by that avatar as defined by its *communication zone*. In an Auditory Scene, each voice stream is localised (directional placement) and distance-attenuated with respect to the appropriate virtual world avatar positions, thus creating a virtual *"cocktail-party"* effect [7].

In many cases, an online game and hence the add-on Immersive VoIP service needs to cater for the concurrent access of a large number of avatars and the distribution of avatars can be quite dense. More importantly, the avatars can be very close in the virtual space but yet spread over a large geographical scale in the physical world. It is therefore important for an Immersive VoIP service to achieve a good balance between the system scalability and the voice quality delivered. To this end, in our prior work [5, 6], we have identified and addressed two key scalability constraints, i.e., the access bandwidth and local processing constraints on the gaming clients. In comparison to the wired gaming clients, these two scalability constraints are even more stringent for the wireless gaming clients, e.g., the SONY PSP [8] for the following reasons:

- The access bandwidth of the wireless clients is limited by transmission spectrum and propagation path loss.
- Despite the rapid progress in processor speed and power efficiency, the processors of the wireless devices are still limited by battery life.
- The requirement of portable headsets instead of cumbersome loud speakers for voice playback.

- The unlikelihood of the gaming clients to commit all their resources to support immersive VoIP due to other simultaneously running applications such as visual rendering of game scenes.



(a) Communication Zones (b) Acceptable Angular Errors

(c) Cluster Avatar Distribution (d) Uniform Avatar Distribution

Figure 1. The virtual world model applied to the Wireless Immersive Communication Environment.

In view of these scalability challenges, we have proposed in [5, 6] the Wireless Immersive Communication Environment (WICE) to deliver Immersive VoIP service to the wireless gaming clients. In order to minimise the client-side processing load, the WICE service uses dedicated servers to complete the *Auditory Scene Creation* (ASC) tasks on behalf of the clients [5]. We applied the *Head Related Transfer Function (HRTF)* for the directional placement of voices due to the playback ability of HRTF streams over the portable headsets [9]. Each client in WICE only has to download two streams of Left and Right HRTF channels from the ASC server for final playback. The *HRTF localisation* [5] accounts for the dominant cost in Auditory Scene Creation, when compared to linear voice mixing and amplitude attenuation. As shown in [5], the *"brute force"* approach of Auditory Scene Creation performs exact voice localisation for each pair of communicating avatars. The *brute force* approach leads to a prohibitively high server-side voice processing cost which increases linearly with avatar density, defined to be the average number of avatars inside each communication zone. To improve server processing scalability, we introduced the conjecture of *acceptable angular errors* [5] which relaxes the accuracy requirements in voice localisation for the avatars further away in comparison to the nearby avatars. We developed the *Voice Processing Minimisation Algorithm* [5] which optimises the sharing of *localised voices* in each communication zone, utilising the concept of *acceptable angular errors*. Our results showed that for an avatar distribution of density 25, the *Voice Processing Minimisation Algorithm* reduced the processing cost of WICE by 84% from the *brute force* approach.

In this paper, the concept of *acceptable angular errors* is verified through the subjective listening results. In the first section of the listening test, the subjects gave satisfactory ratings on the *acceptable angular errors* introduced to the Auditory Scene established at one time instant. However, in the second part, the users were annoyed by the angular shifts in the *localised voices* used by the same avatar pair at adjacent time instants. This finding brings about the important issue of ensuring smooth transition between Auditory Scenes create at successive time instants in support of avatar mobility in the virtual world. To address this issue, we have developed a mechanism to minimise the undesired angular shifts between Auditory Scenes. Analysis was carried out to ascertain the necessary execution frequency of such mechanism in support of avatar mobility. The rest of the paper is organised as follows: Section 2 describes the model of the game virtual world and the two types of avatar distributions applied. Section 3 discusses the user acceptance results of the concept of *acceptable angular error*. In Section 4, we identify and address the key issue in the continuous maintenance of Auditory Scenes in support of virtual world mobility. The final conclusion is given in Section 5.

2. THE GAME VIRTUAL WORLD MODEL

In our simulations, the game virtual world is modelled as a square area of certain size. As illustrated in Fig. 1a, we assume all the avatars to be actively communicating (no muted avatars) and the radii of all the *communication zones* are fixed at 30 m. We define three types of *virtual world grouping behaviour* characterised by varying level of avatar density. A sparse group of avatars represents *loners* which are separated by large distance and have very limited interactions. A dense group of avatars represents either a *clan* (small to medium cluster, e.g., size < 50) or a *crowd* (large cluster, e.g., size > 200) of users. The clans and crowds capture the popularity of some locations present in the real game traces [10] where avatars converge due to common interest, e.g., watching a virtual live show. The real game traces reveal that the distribution of avatars is likely to be non-uniform and a hybrid of loners, clans and crowds. In order to capture such grouping behaviour, we devised a Cluster Avatar Distribution model as illustrated in Fig. 1c. At the first step, we randomly spread a small percentage of avatars (e.g. 10%) across the virtual world to simulate the loners roaming between clans and crowds. At the second step, the rest of the avatar population are placed around $R \geq 1$ clusters. To form the clusters, we randomly position R avatars around the virtual world as the cluster centres. We define a uniform radius as the boundary of each cluster. The number of avatars placed in each cluster is varied according to the normal distribution such that a hybrid of small clusters (clans) and large clusters (crowds) are spread across the virtual world. Within the boundary of each cluster, the assigned avatars are uniformly positioned around the cluster centre. We also compared the cluster distribution to the simple uniform distribution as illustrated in Fig. 1d, where all avatars are randomly placed across the virtual world.

3. SUBJETIVE LISENTING RESULTS

We conducted a series of subjective listening tests with 27 subjects to verify the conjecture of distance-governed relaxation of *acceptable angular error*. We studied the user acceptance of a linear model (see Fig. 1b) of relaxing *acceptable angular errors* over increasing virtual world distance between communicating avatars [5]. We present both visual cues (video record of moving avatars) and auditory cues (accompanied voice recording) to each subject. The *acceptable angular error* was assessed as the perceived discrepancies between the visual and auditory cues. Table 1 shows the Mean Opinion Score (MOS) [11] scale used by the subjects to assess each combination of video and audio clips in terms of the degrading impact of the *acceptable angular error*

introduced. The test subjects with network game experience commented that the genre of the game played has significant influence over the voice localisation accuracy required in Immersive VoIP. For instance, a First-Person-Shooter (FPS) game player is likely to tolerate less *angular errors* than a Role Playing Game (RPG) player. For the test results presented herein, we define the rating of 3 (fair) as the threshold rating on *acceptable angular error* for a typical RPG game scenario. The quality of voice communication between a particular avatar pair is only impaired when the MOS rating falls below this threshold. This threshold is expected to be higher for a FPS game scenario, e.g. at 4 (Good).

Table 1.
Mean Opinion Score Chart for the subjective listening test.

MOS	Localisation Accuracy	Perceived Discrepancy
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	**Fair**	**threshold rating**
2	Poor	Annoying
1	Bad	Very annoying

In the first part of listening test, the subject assessed the impact of the *acceptable angular error* introduced for a *secondary* speaker when there were two simultaneous speaking avatars. The primary speaker remained at close distance to the centre of the view of the subject while the secondary speaker gradually moved away. This simulates an auditory scene with conversations at both close and far distances. In a purely visual sense, the same *acceptable angular error* accounts for a larger geometric spread at larger distance. Therefore increasing the *acceptable angular error* with distance leads to even greater visual positional error at larger distance. However, Fig. 2 shows that within the 95% confidence intervals, when coupling together the visual and the auditory cues, none of the MOS ratings fell below the threshold of 3 and the lowest MOS rating is 3.2 which occurred for the far distance segment with the *acceptable angular errors* ranging from 3.5 to 35 degrees. The Wireless Immersive Communication Environment (WICE) participants are less sensitive to the *acceptable angular errors* introduced, especially at a large distance, because they can also see the speaking avatar. This coupling of visual and auditory cues is always present because WICE is an add-on service to a network game engine. Moreover, the WICE participants are less sensitive to the *acceptable angular errors* introduced because they need to perceptually separate and process simultaneous conversations in an auditory scene [12]. As shown in Fig. 2, the WICE participants were desensitised towards a distant speaker when the other near by speaker sounded more dominant due to distance attenuation. On the basis of Fig. 2, we have chosen the range of *acceptable angular errors* ranging between 3.5 and 35 degrees for our simulations which should suffice for a RPG game. However, for the more accuracy-sensitive FPS games, the maximum *acceptable angular errors* should not exceed 15 degrees to preserve a threshold MOS rating of 4 (Good).

The results in Fig. 2 are concerned with verifying the conjecture of *acceptable angular error* for establishing an Auditory Scene at a particular time instant. In the second part of listening test, we assessed the impact of discrepancies between the visual cue movements and the auditory cue movements in relation to the continuous re-establishments of Auditory Scenes at successive time instants in support of avatar mobility. The subjects listened to two copies of a particular speech localised along different directions to simulate auditory movements. One the other hand, in the accompanying visual cue, the avatar position remained the same. The subjects gave mean MOS ratings of 1.88 (below fair) and 0.88 (below poor) respectively, for the 10 and 20 degrees of angular shifts in voice localisation. This suggests that there must be smooth transitions between Auditory Scenes create at successive time instants in the face of avatar mobility in the virtual world. This is especially true when there is no corresponding visual movement. In [13], 5 degrees was found to be the threshold of just noticeable angular shifts in voice localisation. In Section 4, we devise a mechanism to adhere to the 5 degrees of minimum threshold in transitional angular shifts between successive Auditory Scenes in support of avatar mobility.

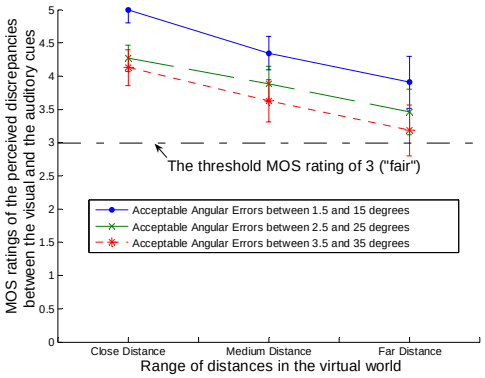


Figure 2. The MOS ratings of the angular error with two speakers

4. CONTINUOUS MAINTENACNE OF AUDITORY SCENES IN SUPPORT OF AVATAR MOBILITY

Our prior work in [5] is only concerned with establishing *"static"* Auditory Scenes in Wireless Immersive Communication Environment (WICE) at a particular time instant. In this section, we address the issue of continuous maintenance of Auditory Scenes in support of avatar mobility in the virtual world.

4.1 The Need to Minimise Transitional Deviation between Auditory Scenes

In light of the listening results in Section 3, we recognise the need to minimise the time shifts between successive Auditory Scenes in the face of avatar mobility. We define such time shifts in voice localisation as the *Transitional Deviation*. Such need to maintain hysteresis in 3-D audio localisation is also recognised in a related work [15] which reduces audio rendering cost reduction using different perceptual criteria. We illustrate the occurrence of *Transitional Deviation* in Fig. 3, using the scenario of avatars A_1 and A_2 listening to avatar B_1. Initially at t_1, A_1 and A_2 share the *localised voice* v_1^{220} of B_1 along 220 degrees for voice processing reduction, with the position of B_1 as the origin. At instant t_2, due to the movement of A_1, the *localised voice* v_1^{230} is computed for sharing between A_1 and A_2, amounting to 10 degrees of *Transitional Deviation*. In this scenario, avatar A_1 can be annoyed because A_1 can perceive the extent of *Transitional Deviation* (auditory movement) to be in excess of the corresponding visual movement relative to the position of B_1. On the other hand, avatar

A_2 can be even more annoyed than A_1 because A_2 did not move from t_1 to t_2 but A_2 has to experience the same *Transitional Deviation* as A_1 due to the sharing of *localised voice* between the two avatars. In theory, there is another less common cause to *Transitional Deviation* when the *Voice Processing Minimisation Algorithm* [5] finds slightly different *localised voices* of avatar B_1 (all capable of meeting the appropriate *acceptable angular error* constraints) for the stationary avatar A_2 at different time instants. In practise, however, an optimisation package such as CPLEX [14] controls this randomness and we observed negligible differences between different runs on the same input.

To simulate the mobility of avatars in the virtual world, we applied the Networked Game Mobility Model (NGMM) [10] which modifies the definition of stationary states in the Random Waypoint Motion (RWM) model. NGMM is characterised by non-straight line motion and is thus capable of preserving the popularity location characteristic of our cluster avatar distribution. We simulated the speed range of 0.1 to 5 meters per second. We set 0.5 second of pause time between the stationary and mobile states to simulate a fast paced First-Person-Shooter (FPS) game, which has more stringent workload requirements than a slower paced Role Playing Game (RPG). We applied the NGMM mobility model over 2000 avatars distributed according to the cluster model in Fig. 1c. However, when studying the impact of avatar mobility, we observed very small difference between the cluster avatar distribution and the uniform avatar distribution, provided that the comparison is made at same avatar density. This is due to the fact that we define avatar density as the average number of avatars per communication zone rather than per square unit of area. Consequently, we have focused our analysis on the effect of varying avatar densities as shown in Fig. 4 to 7. We also found the variation of total number of avatars to have no influence over our simulation results due to the algorithm decomposability as elaborated in Subsection 4.2.

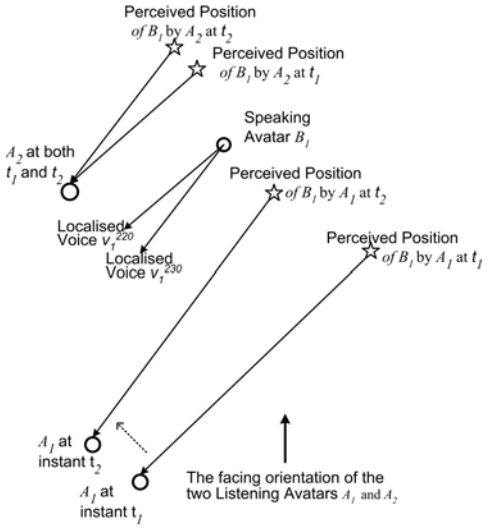


Figure 3. Transitional Deviation between two listening avatars

Avatar mobility leads to significant changes in the relative distance separating avatars in the virtual world and consequently contributes to changes in the *Acceptable Angular Error* constraints between pairs of communicating avatars. Our simulation results found that after 800 milliseconds of avatar movements, at avatar densities ranging between 5 and 25, between 30% and 45% of avatar pairs experience changes in their *acceptable angular error* constraints. The extent of change in *acceptable angular error* was between 0 and 4.32 degrees. On the other hand, our results also reveal that for the same range of avatar densities, between 5% and 15% of the avatar pairs which were initially in communication have gone out of communication after 800 milliseconds of avatar movements. Hence for the mobility model studied, we need to execute the *Voice Processing Minimisation Algorithm* once per 800 milliseconds in order to avoid significant adverse impact of avatar mobility on the voice localisation accuracy and processing scalability of the WICE service. As stated in Subsection 4.2, this requirement can be easily met due to decomposability of *Voice Processing Minimisation Algorithm*. The real issue thus lies with minimising the *Transitional Deviation* between the *Auditory Scenes* created at successive time instants in support of avatar mobility as illustrated in Fig. 3. In the ensuing section, we devised a mechanism to address this important issue.

4.2 The Mechanism to Minimise Transitional Deviation between Auditory Scenes

A simple approach to minimise *Transitional Deviation* would be to reuse the results of *Voice Processing Minimisation Algorithm* from a prior time instant. Nonetheless, this simple approach does not guarantee the minimisation of voice processing cost while meeting all the appropriate *Acceptable Angular Error* constraints. Instead, we propose a mechanism of sequential execution of the *Voice Processing Minimisation Algorithm* and the Linear Programming (LP) based *Transitional Deviation Minimisation Algorithm*. The *Transitional Deviation Minimisation Algorithm* minimises the *Transitional Deviation* incurred between the sets of *localised voices* found by the *Voice Processing Minimisation Algorithm* at two successive time instants, provided that the two sets of *localised voices* are shared by the same avatar pairs at both time instants. The *localised voices* found by the *Transitional Deviation Minimisation Algorithm* are still subject to the minimum processing limit obtained by the *Voice Processing Minimisation Algorithm* and the appropriate *Acceptable Angular Error* constraints. The mechanism to sequentially execute the two algorithms is generally as follows:

0. Execute the *Voice Processing Minimisation Algorithm (Initial)*
1. begin (avatar starts moving)
2. Verify if there is sufficient level of avatar movements;
3. Check whether the *localised voices* found previously can be reused to meet the new set of *acceptable angular error* constraints;
4. if false
5. Rerun the *Voice Processing Minimisation Algorithm*;
6. Examine if the mean *Transitional Deviation* incurred by rerunning the *Voice Processing Minimisation Algorithm* is greater than 5 degrees
7. if true
8. Execute the *Transitional Error Minimisation Algorithm*;
9. end
10. end
11. end

The details of the *Transitional Deviation Minimisation Algorithm* are now given:

Known variables:

The *localised voice* v_i^k is the voice of i^{th} avatar localised along the k^{th} angle with respect to the i^{th} avatar position as the origin (0, 0).

Let θ_{ij}^k denote the precomputed angular displacement between the position of j^{th} listening avatar and the *localised voice* v_i^k .

Let $\beta_{ij}^{kl} = \left|\theta_{ij}^{k} - \theta_{ij}^{l}\right|$ denote the *Transitional Deviation* between two *localised voices* v_i^k and v_i^l computed at two successive time instants for the same pair of j^{th} avatar and i^{th} avatar.

Q: The maximum number of potential *localised voices* for each avatar voice, $1 \le k \le Q$ and $Q \le 360^o$.

N: The total number of avatars in WICE, $1 \le i, j \le N$.

$$a_{ij} = \begin{cases} 1 \text{ if the } j^{th} \text{ listening avatar can hear} \\ \text{the } i^{th} \text{ speaking avatar} \\ 0 \text{ otherwise} \end{cases}$$

Decision Variables:

$$x_i^k = \begin{cases} 1 \text{ if localised voice } v_i^k \text{ is created for} \\ \text{the } i^{th} \text{ speaking avatar} \\ 0 \text{ otherwise} \end{cases} \quad (1)$$

$$z_{ij}^k = \begin{cases} 1 \text{ if the localied voice } v_i^k \text{ is used} \\ \text{by the } j^{th} \text{ listening avatar} \\ 0 \text{ otherwise} \end{cases} \quad (2)$$

Objective Function:

$$\text{Minimise: } \sum_{j=1}^{A}\sum_{k=1}^{N} \beta_{ij}^{kl} z_{ij}^k \ \forall i \quad (3)$$

Similar to the case of *Voice Processing Minimisation Algorithm* [5], the *Transitional Deviation Minimisation Algorithm* is decomposable and can be executed independently for each communication zone. The objective function of the *Transitional Deviation Minimisation Algorithm* minimises the sum of *Transitional Deviation* across two successive time instants for all the avatar pairs in each communication zone.

$$\text{Subject to: } \sum_{k=1}^{N} z_{ij}^k = a_{ij} \ \forall i \quad (4)$$

$$z_{ij}^k \le x_i^k \ \forall i, j, k \quad (5)$$

$$\sum_{j=1}^{A} z_{ij}^k \ge x_i^k \ \forall i, k \quad (6)$$

$$\sum_{k=1}^{N} x_i^k \le L_i \ \forall i \quad (7)$$

$$x_i^k, z_{ij}^k \in \{0,1\} \forall i, j, k$$

As discussed in [5], the *Voice Processing Minimisation Algorithm* can be executed independently for each communication zone. The *Transitional Deviation Minimisation Algorithm* is also decomposable to each communication zone but with greater complexity (not elaborated due to space limit) than the *Voice Processing Minimisation Algorithm*. Hence the scalability of both algorithms is affected by the variation of avatar density rather than the variation of avatar number. We implemented the *Voice Processing Minimisation Algorithm* using the CPLEX software package [14] on a PC running Linux with AMD Athlon64 Dual-Core 2.0 GHz processor and 4.0 GB of RAM. For the *Voice Processing Minimisation Algorithm*, the mean execution time was measured to be between 0.68 and 1.96 milliseconds per communication zone, for avatar densities between 5 and 25. Over the same range of avatar densities, the mean execution time of the *Transitional Deviation Minimisation Algorithm* was measured to be between 8.62 and 58.78 milliseconds per communication zone. Hence, over a range of avatar densities, our two algorithms can scale to support a large number of avatars, e.g. millions.

4.3 The Impact of Virtual World Mobility

We use two metrics to represent the actual level of *Transitional Deviation* incurred. The first metric is the percentage of avatar pairs with above zero *Transitional Deviation* across two time instants (referred to as the "*affected avatar pairs*"). The second metric is the *mean Transitional Deviation* among the *affected avatar pairs*. Figures 4 to 7 show the comparison in the two metrics of *Transitional Deviation* between the case of applying the *Transitional Deviation Minimisation Algorithm* and the case of not applying.

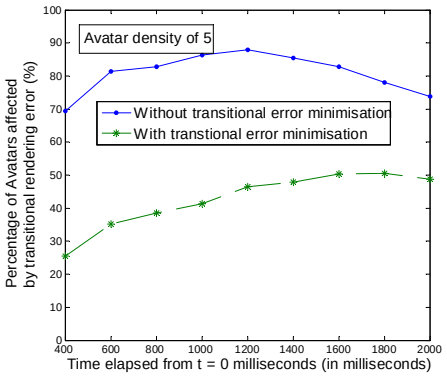


Figure 4. Percentage of avatar pairs affected by *Transitional Deviation* at low avatar density of 5 avatars per communication zone.

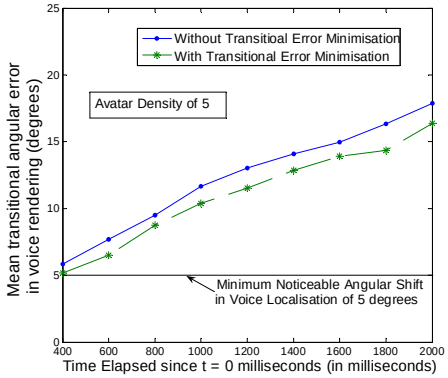


Figure 5. *Mean Transitional Deviations* at low avatar density of 5 avatars per communication zone.

To evaluate the effect of varying avatar densities, we show the two examples of avatar densities 5 (low) and 25 (high). As shown in Fig. 4 and 6, the first metric of *affected avatar pairs* does not increase consistently with the elapsing of time. A contributing factor to such inconsistent trend is the fact that some avatar pairs initially in communication can move out of communication range later on as discussed in Subsection 4.1. Despite the inconsistent trend with the first metric, the sum of *Transitional Deviations* for each communication zone still increases (worsens) consistently with the elapse of time. This is because the sum of *Transitional Deviations* is the product of the two metrics and there is consistent and significant increase in the second metric of *mean Transitional Deviation* as revealed in Fig. 5 and 7. Similarly, in order to minimise the sum of *Transitional Deviation*, the *Transitional*

Deviation Minimisation Algorithm can choose to minimise either or both of the two individual metrics. By comparing Fig. 4 and 5, we can see that at low avatar density of 5, the *Transitional Deviation Minimisation Algorithm* achieves greater improvements in the first metric (between 25% and 43% of reductions) than it does in the second metric (between 0.6 and 1.5 degrees of reductions). On the contrary, at high avatar density of 25, as shown in Fig. 6 and 7, the *Transitional Deviation Minimisation Algorithm* achieves greater improvements in the second metric (between 3.8 and 6.5 degrees of reductions) than it does in the first metric (between 0 and 1.7% reductions). In Section 3, our subjective listening results recommend the level of *Transitional Deviation* to be kept at or below 5 degrees. Consequently, as shown in Fig. 5 and 7, for the mobility model simulated, at avatar densities of 5 and 25, the sequential execution of the *Voice Processing Minimisation Algorithm* and the *Transitional Deviation Minimisation Algorithm* needs to occur once per 400 milliseconds. The 400 milliseconds frequency is also sufficient to avoid significant adverse impact on the localisation accuracy and scalability of the WICE service, because it is faster than the 800 milliseconds frequency previously recommended in Subsection 4.1. Moreover, as stated in Subsections 4.2, for avatar densities between 5 and 25, the sequential execution time of the two algorithms ranges between 9.3 and 60.74 milliseconds per communication zone. Hence, 400 milliseconds is ample time for the sequential execution of the two algorithms to continuously maintain *Auditory Scenes* in support of avatar mobility.

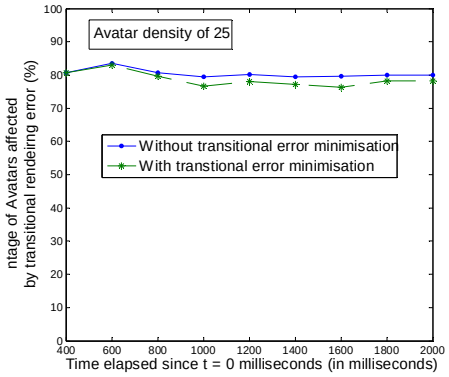


Figure 6. Percentage of avatar pairs affected by *Transitional Deviation* at high avatar density of 25 avatars per communication zone.

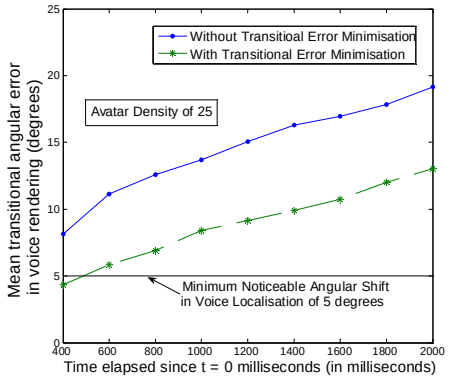


Figure 7. *Mean Transitional Deviations* at high avatar density of 25 avatars per communication zone.

5. CONCLUSION

In this work, our subjective listening results suggest that for its intended gaming scenario, the Wireless Immersive Communication Environment (WICE) can offer perceptually acceptable quality of immersive voice communication when applying the concept of distance-governed relaxation of *acceptable angular errors* for processing reduction. More importantly, our listening results identified the need to minimise time shifts (*Transitional Deviation*) between the Auditory Scenes created at successive time instants in the face of avatar mobility. To address the issue, we devise the mechanism of sequentially executing the *Voice Processing Minimisation Algorithm* and the *Transitional Deviation Minimisation Algorithm*. For the mobility model studied, we ascertained once per 400 milliseconds as the necessary frequency to execute this mechanism.

6. ACKNOWLEDGMENT

This work is supported by the Co-operative Research Centre for Smart Internet CRC (SITCRC) and the University of Wollongong (UOW), Australia.

7. REFERENCES

[1] Microsoft Corporation, Xbox LIVE, http://www.xbox.com/en-US/live, (16 May 2007).

[2] MMOGCHART.com, http://www.mmogchart.com/ (10 April 2007).

[3] Kuan-Ta Chen and Chi-Laung Lei, "Network Game Design: Hints and Implications of Player Interaction", in *Proc. of 5th ACM workshop On NetWork & System support for Games (NetGames 06)*, Singapore, 29/Oct.-2/Nov., 2006.

[4] Hew, K., Gibbs, M. R., and Wadley, G., "Usability and sociability of Xbox Live voice channel", In *Proc. of Australian Workshop on Interactive Entertainment (IE2004)*, pp. 51-58.

[5] Ying Peng Que, Paul Boustead, and Farzad Safaei, "Minimising the Computational Cost of Providing a Mobile Immersive Communication Environment (MICE)", in *Proc. of the 2nd IEEE International Workshop on Networking Issues in Multimedia Entertainment (NIME 06) at Consumer Communications and Networking Conference (CCNC)*, Las Vegas, Nevada, U.S.A, 7-10 Jan. 2006.

[6] Ying Peng Que, Paul Boustead, and Farzad Safaei, "Trading off Computation for Error in providing Immersive Voice Communications for Mobile Gaming", in *Proc. of the 5th ACM Workshop on Network and System Support for Games 2006 (NetGames 06)*, Singapore, 30-31 Oct. 2006.

[7] E. C. Cherry, "Some further experiments upon recognition of speech with one and with two ears", in Journal of Acoustics Society America, Vol. 28, pp. 889-895, 1956.

[8] Sony Computer Entertainment, http://www.yourpsp.com.au/, (9 Jan. 2007).

[9] Durand R. Begault, *3-D Sound for Virtual Reality and Multimedia*, Academic Press Professional, Cambridge, MA, USA, 1994.

[10] Tan S.W., Lau W. and Loh A., "Networked Game Mobility Model for First-Person-Shoot Games", in *Proc. of the 4th ACM workshop On Network & System Support for Games (NetGames 05)*, Hawthorn, NY, USA, 10-11 Oct. 2005.

[11] Methods for subjective determination of transmission quality, International Telecommunications Union (ITU), Geneva, Switzerland, ITU-T Recommendation P.800, August, 1996.

[12] Albert S. Bregman, *Auditory scene analysis: the perceptual organization of sound*, MIT Press, Cambridge, Massachusetts, USA, 1999.s.

[13] D. Wesley Grantham, Benjamin W. Y. Hornsby, and Eric A. Erpenbeck, "Auditory spatial resolution in horizontal, vertical, and diagonal planes", *J. Acoust. S. A.*, vol. 114, no. 2, pp. 1009-1022, May 16, 2003.

[14] Cplex optimisation software, documentation available at http://www.ilog.com, (9 Jan. 2007).

[15] Tsingos, N., Gallo, E. and Drettakis, G. (2004). Perceptual audio rendering of complex virtual environments. In *Proc. of ACM SIGGRAPH 2004*, pp. 249-258. Los Angeles, CA, USA.