Low-Delay Video Control in a Personal Area Network for Augmented Reality

R. Razavi, M. Fleury, and M. Ghanbari

University of Essex, United Kingdom,
{rrazav,fleum,ghan}@essex.ac.uk

Abstract

A Personal Area Network (PAN) is a feature of an Augmented Reality system, transmitting modified video for real-time display. In the paper, low-delay communication of encoded video over a Bluetooth wireless PAN is achieved by a combination of dynamic packetisation of video slices together with centralised and predictive rate control. The result is minimized packet delay (below 0.05 s) and high-quality 40 dB video, with packet loss limited to 4% from RF noise.

1 Introduction

Video transmission is an essential feature of augmented reality (AR) systems and this paper considers how video communication latency can be reduced across a Personal Area Network (PAN) supporting AR. As an augmented scene is viewed in real-time, it is important to give the illusion of an instantaneous modification of the view. In particular, the paper applies low-delay control to an IEEE 802.15.1, Bluetooth [8], real-time video stream in the face of sensor and external data cross traffic. Bluetooth is a low-power system [5] with typical currents of 1-35 mA, compared to IEEE 802.11's (Wi-Fi) 100-350 mA, though with shorter range of 10 m for class 1 devices as opposed to 100 m. In the AR system modelled, a video camcorder captures the surrounding scene, which is subsequently played back to the wearer with the addition of superimposed information. Such a system has applications in technician tasks such as airplane maintenance, and generally in the medical and military fields [1].

AR is an adjunct of wearable computers. Wireless PANs are now preferred rather than the pioneering but cumbersome wired networks on early wearable computers [15]. Though a near-field intra-body network [22] may be ideal for low-power sensor/actuator devices, Radio Frequency (RF) networks such as Bluetooth are immediately deployable. Bluetooth is similar in star topology, spread spectrum, and modified Time Domain Medium Access (TDMA) to BBN's BodyLAN [12], also designed as a PAN. Bluetooth version 2 has another advantage, as its Enhanced Data Rate (EDR) [18] makes it more capable of supporting low bitrate encoded video transmission , with a maximum (shared) payload of 2.1 Mb/s. Bluetooth calls its PAN a piconet, consisting of one master node and up to seven active slaves. An extension mechanism, allows the construction of scatternets, with more nodes when necessary.

Fig. 1 shows an archetypal AR system with a PAN loosely based on Sony's NaviCam [17]. In the original NaviCam system, apparently raw video was transferred from the camera through direct memory access, whereas the proposed system would employ an H.264 hardware codec, with current ICs listed in [14]. A central, wearable computer processes incoming encoded video, adding additional information from an internal source and/or transferred from a wireless access point. Hardware chroma-keying for this purpose allows video-rate display to be approached. The video is then re-transmitted in encoded form to the display device at a rate of 20 frame/s (updated from the 10 frame/s NaviCam rate). At least Common Intermediate Format (CIF) resolution (366×288 pixel) is advisable to ensure that text fonts are readable. In NaviCam, the user views either a PDA or a see-through head-up display. In addition to constructing the AR view, the central microprocessor is also responsible for identifying visual targets placed in the scene, which serve to match augmented information with video sequence. A Bluetooth beacon (rather than NaviCam's IR beacon) additionally serves to identify the scene location[1]. Once identified, the orientation of the camcorder to the visual target is calculated. Clearly, the various processing tasks are likely to cause delay between the capture of the external scene and its display in AR form. In fact, research in [10] went as far as a parallel computing system for AR visual target processing for accurate registration in a virtual TV studio.

However, the focus of this paper is the need to reduce delay to a minimum across the PAN itself. In Bluetooth, separate send and receive buffers are utilized for each master to slave link. (There is no direct slave-to-slave communication in Bluetooth.) Packet delay occurs in the presence of cross traffic (i.e. from sources other than the primary AR video source), because the master node must spend time servicing those sources, resulting in delayed access by the video source to the shared channel. When there is no cross traffic, delay may still occur by virtue of inefficient utilization of the channel, such as through partially-filled packets or greater packet header overheads. This can result in effective bandwidths well below the nominal capacity. And, if rate control is not in place, packets wait in buffers until the channel becomes available.

As an AR system is centralized upon its CPU, it is possible to regulate the video stream bit-rate across the shared wireless channel according to a prediction of traffic conditions. In this paper a *P*-order linear prediction filter (LPF) [9] tracks the anticipated bit-rate. As Bluetooth automatically replies to each data packet with an acknowledgement (ack) packet, this prediction is easily piggy-backed upon a one time-slot ack. For more complex systems than Fig. 1, global knowledge of the traffic conditions might be disseminated by Bluetooth broadcast, which can be made reliable [4].

---

[1] As a Bluetooth beacon is omindirectional, this would require the addition of Bluetooth's optional Received Signal Strength Indication. Of course, an IR beacon to IR sensor with Bluetooth transceiver is also possible.

Apart from efficient rate control, this paper also proposes dynamic Bluetooth packetisation to effectively utilize the available bandwidth. In [16], fully-filled five time-slot Bluetooth packets were formed from arriving one slice IP packets, regardless of slice boundaries, which we call dynamic packetization (as it requires cross-layer intervention). While this results in some loss in error resilience, as each slice contains a decoder synchronization marker, in [16] it is shown that the overall video performance is superior to partially-filled packets that preserve slice boundaries. Dynamic packetisation has direct benefits in terms of minimizing end-to-end delay through reduction of buffer queuing time. In fact, the research herein demonstrates that rate control is only effective when combined with dynamic packetisation.

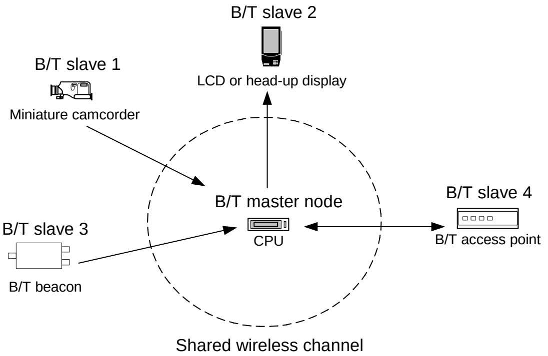

Figure 1: AR Bluetooth-based PAN system.

2 Related work

A variety of strategies have been adopted to reduce real-time video delay across wireless networks. Work in [3] addresses the speed mismatch between an encoding video-server, as part of a surveillance system, and decoding PDAs (with low processing power and restricted buffer memory) connected over a wireless LAN. Feedback messages to the server control its encoding rate. Selective packet drop of predictive picture over intra-coded picture packets also contributes to reducing buffering delay at the receiver. Additionally, a low-delay decoding library, MS GAPI intended for video games on pocket PCs, is employed. In the MPEG-2 to MPEG-4 frequency domain transcoder of [19], transcoding complexity is reduced, as is the bit-rate to reduce buffer overflow. Predictive picture skipping is also employed in this approach. In [7] for lossy channels, unequal error protection through H.264's data partitioning is combined with low bit-rate communication. A stable output rate was achieved by combining channel with source coding. In [20], the need for high complexity rate-distortion calculations was avoided by means of dynamically updating a rate table. Macroblock Sum of Absolute Differences (SAD) value and the allocated number of bits index into the rate table.

3 Methodology

3.1 Bluetooth packetisation

Because of packet quantization effects, the Bluetooth Asynchronous Connection-Less (ACL) packet sizes become significant and their effect on user payload are summarized in Table 1 for a single master-slave ACL link. In Table 1, only the highest gross air rate is represented, as for simplicity only this rate was employed in the simulations of Section 4 (lower rates are applicable in noisier channel conditions).

The assumed Bluetooth controller behaviour is that, given a maximal Bluetooth packetization scheme, for example 3DH5 or 3DH3, packets up to the maximum user payload will be formed. However, if the arriving packets do not justify the pre-set maximal scheme, then a reduced scheme is applied. For example, the controller swaps from 3DH5 down to 3DH3 or even 3DH1. In Figure 2, it is apparent that selection of a small packet size results in relatively low throughput. A question arises in the case of partially-filled packets whether it is preferable to utilize two single 3DH1 packets or one 3DH3 packet. In terms of throughput, link utilization, and end-to-end delay, it is almost certainly the case [6] that a single 3DH3 packet should be preferred. Unfortunately, in [21] due recognition of the uneven step size in user payload between one, three, and five time-slot packets (Table 1) did not occur, resulting in the inefficient one-slot packets being utilised more than was necessary.

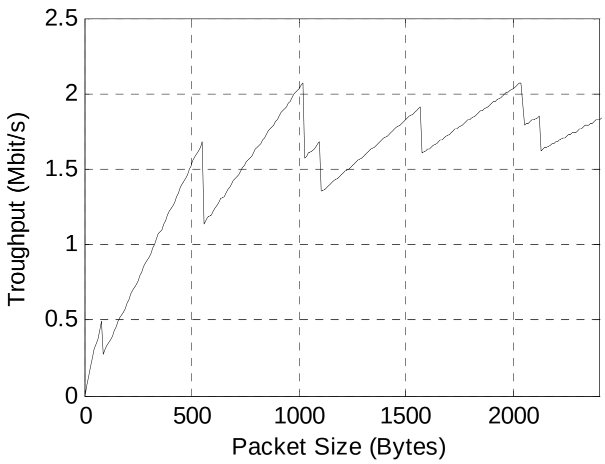

Figure 2: Variation in throughput with packet size.

Packet type	User payload in bytes	Asymmetric max. rate in kb/s
3DH1	0-83	531.2
3DH3	0-552	1776.4
3DH5	0-1021	2178.1

DH = Data High Rate, *e.g.* 3DH3 is 3.0 Mb/s datarate with a three time-slot packet.

Table 1: Length and master to slave bitrates, for a single ACL master-slave link at gross air rate of 3.0 Mb/s.

Over and above decisions on packet time-slot selection, a further restriction arises in video due to the presence of the slicing structure introduced for error resilience purposes. For static packetisation, a single slice per Bluetooth packet is allocated. Maximum slice extent in the MPEG2 codec is limited to a single macro-block row-sized slice, but in the H.264/AVC codec, even with more flexible slice formats, the same problem occurs, as it is difficult to predict in advance the optimal slice size, because of the presence of cross traffic on the PAN. Therefore, the results herein are equally applicable to the H.264/AVC codec.

In Section 4, a comparison between video rate control with static and dynamic packetisation occurs. For a single Bluetooth link, Fig. 3 shows that for static packetisation of a typical MPEG2 video sequence (as described in Section 3.3) at low normalized rates, one-slot packets predominate, leading to inefficient transmission, owing to the header overheads. Fig. 4 demonstrates that, as a result, across a single Bluetooth link, the bandwidth gain from employing dynamic packetisation rather than static packetisation increases at low rates, when the same MPEG2 video sequence is transferred.

Therefore, at the highest rate, dynamic packetisation is, from Fig. 4, approximately three times more efficient than static packetisation. *Vice versa*, static packetisation results in an approximately threefold reduction in throughput over the input encoded rate. In these circumstances, many packets are dropped through transmit buffer overflow.

In the simulations of Section 4, a transmit buffer of 50 packets per link was assumed at each node, with an equal-sized receive buffer. This size buffer was also employed in [13] for video transmission over an IEEE 802.11a network, though with layered video. Larger buffer sizes may lead to further queuing delay.

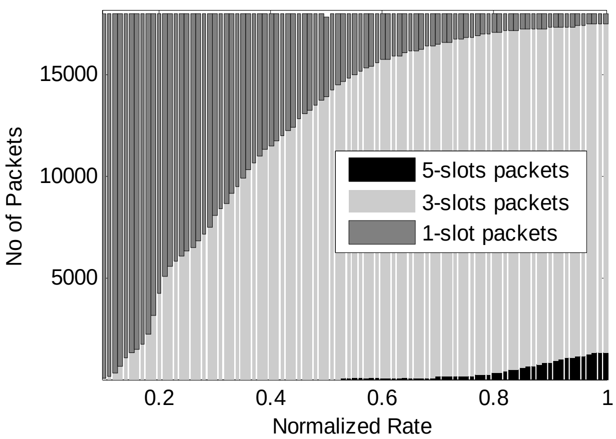

Figure 3: Effect of static packetization of MPEG2 video slices on packet type distribution with varying rates.

3.2 Traffic rate prediction

Bluetooth is a master-driven Time-Division Duplex (TDD) system in which the master polls the slave before the slave replies with data. A Bluetooth frame, consisting of the master's polling message and the slave's reply consists of an even number of time slots, implying that if duplex data is not exchanged then a single time slot should be used for some other purpose. In this paper, that slot is used to supply rate control information to the video source, slave 1 in Fig. 1. Several master-driven scheduling schemes, e.g. [2][11][21], have been devised that seek to optimize the performance of the system as a whole, in terms of energy reduction, fairness, bandwidth satisfaction, as the default Bluetooth scheduling system (round-robin) actually does not perform well in master-driven TDD. In an AR system, the video stream is given priority and the rate predicted based on sampling the available bandwidth over time.

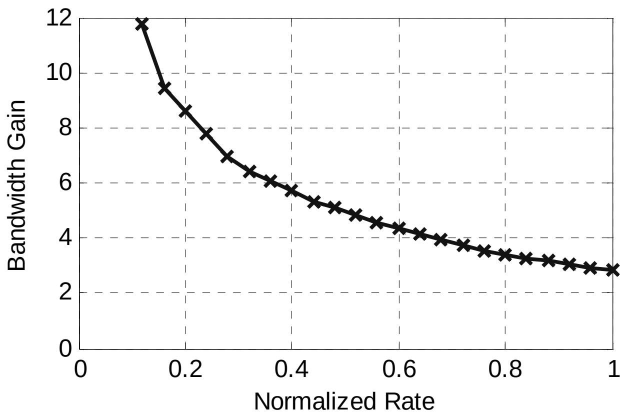

Figure 4: Bandwidth gain from employing dynamic rather than static packetisation of MPEG2 video slices with varying rates.

The predicted rate was arrived at by the P-order LPF mentioned in Section 1. The *P*-order LPF prediction filter is represented by

$$X(m+1) = \sum_{k=1}^{P} w_k . X(m-k+1)\,, \qquad (1)$$

where $X(m+1)$ is a predicted value estimated from P previous samples, while the w_k are the adaptive filter weights indexed by k. The weights are estimated [9] through (2)

$$\mathbf{w}(m+1) = \mathbf{w}(m) + \frac{e(m).\mathbf{X}(m)}{\left\|\mathbf{X}(m)\right\|^2}, \qquad (2)$$

where $\mathbf{w}(m)$ is m^{th} sample of the length P column vector of weights and $\mathbf{X}$ is a length P column vector of measurements over time, as in (3).

$$\mathbf{X}(m) = \left[X(m), X(m-1), \ldots, X(m-P+1)\right]^T, \quad (3)$$

when T represents the vector transpose. The variable $e(m)$ is the error between the measured and the predicted value.

3.3 Traffic model

Table 2 summarizes the traffic rates modelled in Section 4. The video source rate (Slave S1 to Master M) is subject to rate control and, therefore, its rate is the mean encoded rate before control.

A 40 s MPEG2 CIF-sized 20 frame/s video sequence was input with moderate motion and with a Group of Picture (GOP) structure of N=12 M=3 (M is the number of pictures from the I-picture to the first P-picture, i.e. including two B-pictures). It is unlikely that AR would be regularly used with fast motion occurring due to the need to read the augmented material.

In the tests of Section 4, an Additive White Gaussian Noise (AWGN) channel is modelled, with a uniformly distributed Bit Error Rate (BER) of 10^{-5} corresponding at 3.0 Mb/s gross air rate to an E_b/N_0 of 16 dB. To avoid complicating the analysis and to reduce delay, Bluetooth's ARQ mechanism was turned off.

Communication	**Mean bitrate (kb/s)**	**Type (packet size)**
S1 to M	800	VBR
M to S2	800	VBR
S4 to M	250	CBR (800 B)
S3 to M	50	CBR (50 B)

Table 2: PAN traffic conditions.

This research employed the University of Cincinatti Bluetooth (UCBT) extension (download is available from `http://www.ececs.uc.edu/~cdmc/ucbt`) to the well-known ns-2 network simulator (v. 2.28 used). The UCBT extension supports Bluetooth EDR but is also built on the air models of previous Bluetooth extensions such as BlueHoc from IBM and Blueware.

4 Results

4.1 Calibration experiment

In Fig. 5, the ability of the P-order LPF to track available bandwidth is demonstrated. Three slaves send CBR traffic across the Bluetooth PAN at varying time instances, as plotted in the three upper plots. To reflect this cross traffic, the prediction filter calculates the available bandwidth based on the cross traffic and the video source from the camcorder in Fig. 1. In the lowest plot of Fig. 5, the prediction is presented as the predicted throughput achievable against the actual throughput achieved. The prediction is capped to avoid suggesting an available bandwidth that exceeds the maximum rate of the source (800 kb/s).

4.2 Rate control with packetization experiments

By way of comparison, Fig. 6 shows the end-to-end delay experienced by Bluetooth packets forming the MPEG2 video sequence of Section 3.3. Static packetisation was applied. Delay was recorded over time at the receiver (slave 2 in Fig. 1), after transmission from slave 1 in Fig. 1. All processing time at the master node is neglected. In Fig. 6, at time 13 s, the other CBR sources of Table 2 commenced transmission. A sharp increase in delay is observed.

In Fig. 7, rate control is introduced with the result that delay is halved even though static packetisation is applied to the input video sequence. When no cross traffic is present, rate control rapidly reduces delay. Fig. 8 shows the resulting video quality (luminance PSNR), taking into account any packet losses and missed display deadlines. The quality is already reasonable, despite the effect of the AWGN channel.

If dynamic packetisation is applied without rate control, then delay remains high, Fig. 9. Video quality, Fig. 10, once the cross traffic starts, is similar to that in Fig. 8. When dynamic packetisation is combined with rate control, there is a dramatic decrease in delay, Fig. 11, and the received video quality is very good, Fig. 12.

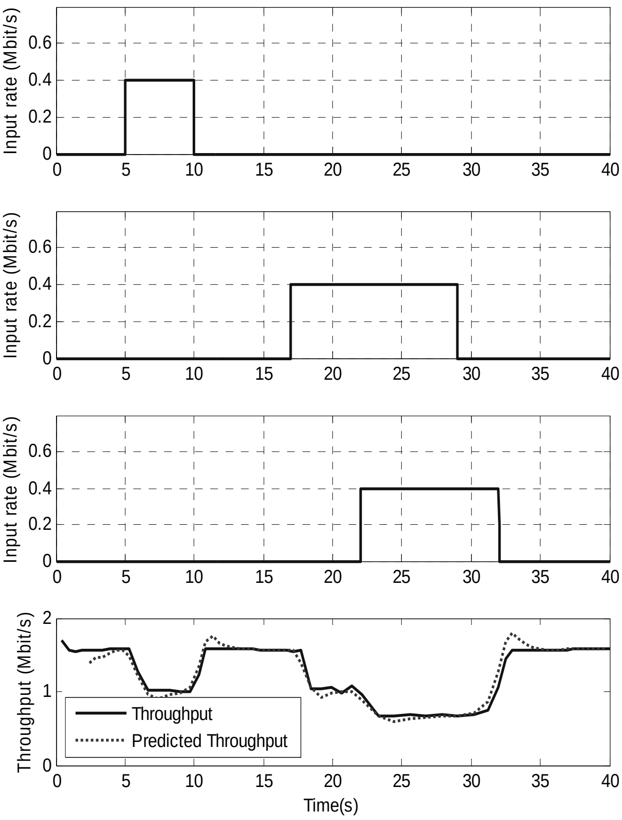

Figure 5: Predicting performance by an LPF with CBR cross traffic, showing three CBR inputs and the impact on video throughput, together with the predicted video throughput.

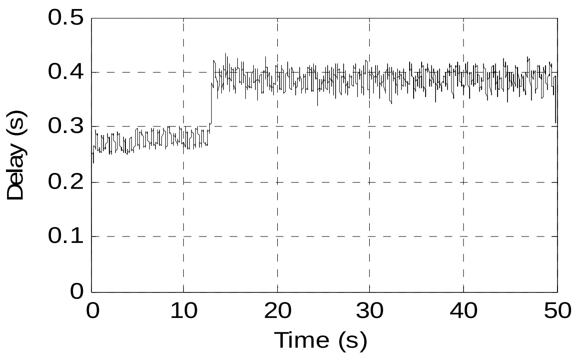

Figure 6: Video packet delay across the PAN with static packetisation but no rate control.

Table 3 presents summary statistics of the results. In the case of, no dynamic packetisation and no rate control, the effect of poor utilization results in severe packet loss principally through transmit buffer overflow. Packet delay is also relatively high and at a mean of 386 ms would be reflected in a perceptible delay at the display. Introducing rate control reduces delay and packet loss. Introducing dynamic packetisation improves the received video quality, but without rate control delay remains high. Finally, combining rate control with dynamic packetisation solves all delivery problems on the PAN. Any residual packet loss is due to channel noise and, without reliable packet transport, is unavoidable.

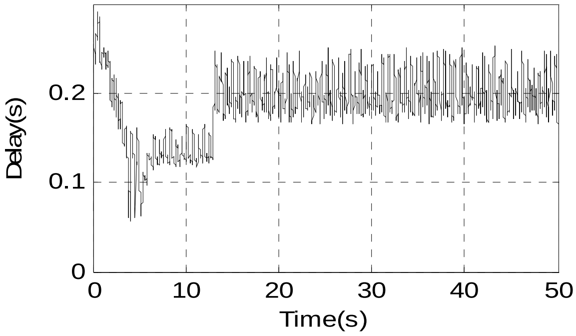

Figure 7: Video packet delay across the PAN with static packetisation and rate control.

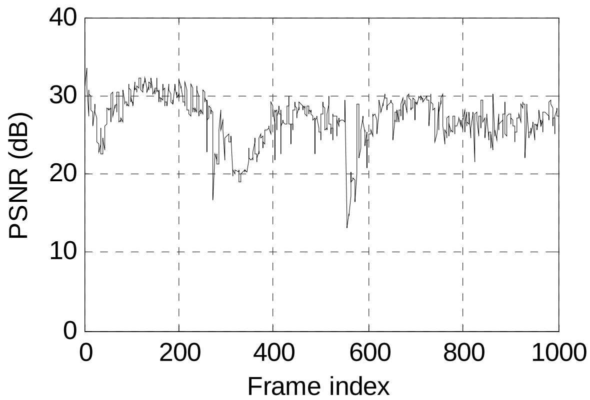

Figure 8: Video quality across the PAN with static packetisation and rate control.

5 Conclusions

AR systems rely on real-time video communication in which delay between scene capture and display should be minimized. In respect to packet delay across an underlying Bluetooth PAN, this paper has shown that without careful consideration of packetisation issues and/or video rate control, packet delay will result in a perceptible lag. A feature of Bluetooth and AR is centralised control and processing, which allows monitoring of available bandwidth. By measuring the past available bandwidth in the face of cross traffic, it is possible to predict the future achievable video rate. Bandwidth utilization of the shared Bluetooth channel is vitally dependent on efficient packetisation of incoming video slices and this is the more so if a lower rate is chosen. Rate control and efficient packetisation are shown in the paper to jointly solve the problem of achieving low-delay and high-quality video. Ultra WideBand (UWB) impulse radio in Bluetooth v. 3.0 will strengthen the attraction of Bluetooth for AR applications with higher bandwidths for video transmission, if its low power advantage is maintained.

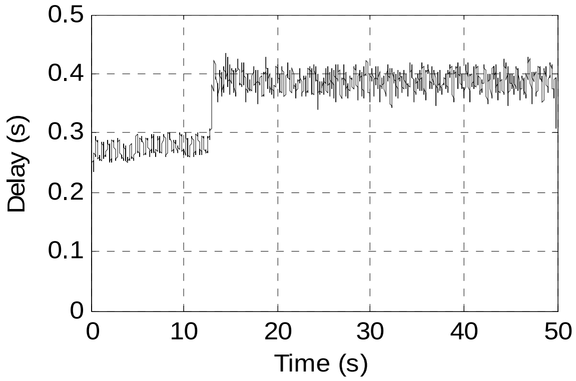

Figure 9: Video packet delay across the PAN with dynamic packetization but no rate control.

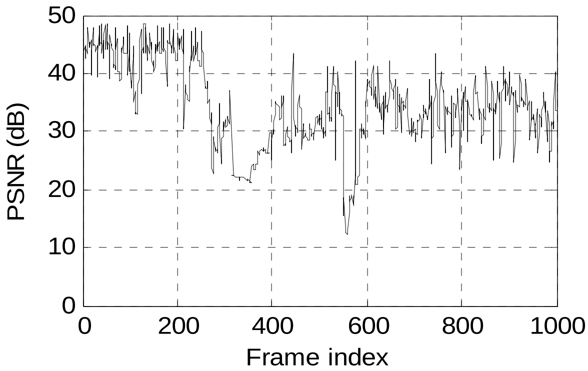

Figure 10: Video quality across the PAN with dynamic packetisation but no rate control.

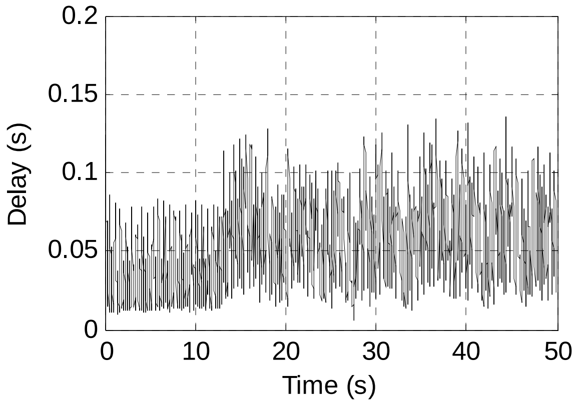

Figure 11: Video packet delay across the PAN with dynamic packetisation and rate control.

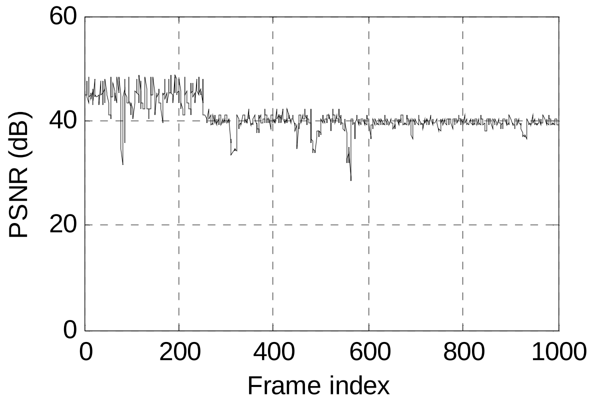

Figure 12: Video quality across the PAN with dynamic packetisation and rate control.

Acknowledgements

This work was supported by the EPSRC, UK under grant no. EP/C538692/1.

References

[1] W. Barfield and T. Caudell (eds.), *Fundamentals of Wearable Computers and Augmented Reality*, Lawrence Erlbaum, New Jersey, (2001).

[2] I. Chakraborty et al., "MAC Scheduling Policies with Reduced Power Consumption and Bounded Packet Delays for Centrally Controlled TDD Wireless Networks". *IEEE Int. Conf. on Communications, pp. 1980-1984*, (2001).

[3] K.-W. Cheuk, S.-H. Chan, K.-W. Mong, C.-M. Lee, S.-S. Sy, "Developing PDA for Low-Bitrate Low-Delay Video Delivery", *IEEE Int. Conf. on Mobile and Wireless Communications Networks*, (2003).

[4] K. Dogan, G. Gurel, A. K. Kamci, and I. Korpeoglu, "Bluetooth Broadcasting Performance: Reliability and Throughput", *Int. Conf. on Comp. Science*, LNCS 3992, pp. 996-999, (2006).

[5] E. Ferro and F. Potortì, "Bluetooth and Wi-Fi Wireless Protocols: A Survey and Comparison", *IEEE Wireless Comms*., **vol. 12, no. 1**, pp. 12-26, (2005).

[6] A. Ghose, A. Razdan, H. Saran, and R. Shorey, "Enhancing Performance of Asynchronous Data Traffic over the Bluetooth Wireless Ad-hoc Network", *IEEE INFOCOM*, **vol. 1**, pp. 591-600, (2001).

[7] O. Harmanci and A. M. Tekalp, "Optimization of H.264 for Low Delay Video Communications over Lossy Channels", *Int. Conf. on Image Processing*, pp. 3209-3212, (2004).

[8] J. Haartsen, "The Bluetooth Radio System", *IEEE Personal Communications*, **vol. 7, no. 1**, pp. 28-36, (2000).

[9] M. K. Honig and D.G. Messerschmitt, *Adaptive Filters Structures, Algorithms, and Applications*, Kluwer, Boston, MA, (1990).

[10] D. J. Johnston, M. Fleury, A. C. Downton, and A. F. Clarke, "Real-time Positioning for Augmented Reality on a Custom Parallel Machine", *Image and Vision Computing*, **vol. 23, no. 3**, pp. 271-286, (2005).

[11] M. Kalia, D. Bansal, and R. Shorey, "Data Scheduling and SAR for Bluetooth MAC", *IEEE Vehicular Technology Conf.*, pp. 196-200, (2000).

[12] R. LaRowe and C. Elliott, "Computer Networks for Wearable Computing", in *Fundamentals of Wearable Computers and Augmented Reality*, Lawrence Erlbaum, New Jersey, pp. 715-745, (2001).

[13] Q. Li and M. van der Schaar, "Providing QoS to Layered Video Over Wireless Local Area Networks Through Real-Time Retry Limit Adaptation", IEEE Trans. on Multimedia, **vol. 6**, **no. 2**, pp. 278-290, (2004).

[14] D. Marpe, T. Wiegand, G. J. Sullivan, "The H.264/MPEG4 Advanced Video Coding Standard and its Applications", *IEEE Communications*, **vol. 44, no. 8**, pp. 134-143, (2006).

[15] S. Mann, "Wearable Computing A First Step towards Personal Imaging", *IEEE Computing*, **vol. 30, no. 2**, pp. 25-32, (1997).

[16] R. Razavi, M. Fleury, and M. Ghanbari "An Efficient Packetization Scheme for Bluetooth Video Transmission", *Elec. Letters*, **vol. 42, no. 20**, pp. 1143-1145, (2006).

[17] J. Rekimoto, "NaviCam: A Palmtop Device Approach", in *Fundamentals of Wearable Computers and Augmented Reality*, Lawrence Erlbaum, New Jersey, pp. 353-377, (2001).

[18] Specification of the Bluetooth System --- 2.0 + EDR, Available online at http://www.bluetooth.com, (2004).

[19] Y. Sun, X. Wei, and I. Ahmad, "Low-delay Rate Control in Video Transcoding", *IEEE Int. Symposium on Circuits and Systems*, **vol. II**, pp. 660-663, (2003).

[20] J.-C. Tsai, "Rate Control for Low-Delay Video using a Dynamic Rate Table", *IEEE Trans. On Circuits and Systems for Video Technology*, **vol. 15, no. 1**, pp. 133-137, (2005).

[21] C.-C. Yang, and C.-F. Liu, "A Flexible Bandwidth Management Scheme in Bluetooth", *IASTED Conf. on Wireless Networks and Emerging Technologies*, (2004).

[22] T. G. Zimmerman, "Personal Area Networks: Near-Field Intrabody Communications", *IBM Systems Journal*, **vol. 35, nos. 3/4**, pp. 609-617, (1996).

Scheme	**Packet Loss (%)**	**Mean PSNR (dB)**	**Average Delay (s)**
SP, No Rate Control	61.29	----	0.3498
SP and Rate control	26.08	26.93	0.1820
DP, No Rate Control	17.51	34.54	0.3865
DP and Rate Control	4.01	40.83	0.0557

Table 3: Summary statistics of video quality and delay for the experiments of Section 4.