Exploring the Extent and Impact of Playout Adjustments within VoIP Applications on the MOS Scale

Hugh Melvin, Liam Murphy

***Abstract*— In coping with best-effort service, many VoIP applications employ adaptive playout strategies. Objective methods of speech quality assessment such as the ITU-T Recommendation P.862 (also known as Perceptual Evaluation of Speech Quality PESQ) typically do not capture distortion due to playout adjustments as they match up short segments prior to analysis. Similarly, the ITU-T E-Model does not capture the effect of delay variation and uses an average delay figure in its calculations. In this paper we explore in some detail, the extent of playout adjustments within VoIP applications and assess the likely impact on Mean Opinion Score MOS. We review the impact of various factors such as Voice Activity Detection (VAD) settings and hangover thresholds on talkspurt/silence period distribution. In this context we examine the distribution of playout adjustments resulting from various playout algorithms and assess the likely impact on MOS. We show that our hybrid playout strategy which utilises synchronised time to implement an informed fixed delay playout strategy wherever possible will significantly reduce playout adjustments and any consequent MOS degradation.**

***Keywords:* Talkspurt/Silence Period Distribution, Playout Adjustments, MOS, Synchronised Time.**

I. Introduction

THE non-deterministic nature of the default best-effort Internet presents a significant challenge for delay-sensitive applications such as VoIP. Significant research has thus focused in recent years on developing receiver playout strategies that adapt to network conditions. Such strategies can be categorised as either per-talkspurt or per packet. The former take advantage of silence periods within natural speech and adjust such silences to track network variations, thus preserving the integrity of talkspurts. Examples of this approach include [1] and [2]. The distribution of talkspurts and silence periods within speech is influenced greatly by the Voice Activity Detection (VAD) and hangover settings. This in turn will impact on the performance of receiver playout strategies. More recent research proposes an adaptive approach that makes adjustments both during silence periods and during talkspurts by *scaling* of packets. This approach is more responsive to short network delay changes in that the per-talkspurt approach can only adapt during silences even though the timescale of many delay spikes may be less than that of a talkspurt. The main disadvantage of this approach is the degradation caused by the *scaling* of speech packets. Examples of the latter approach include [3] and [4]. Finally, recent research by the authors has proposed a hybrid playout strategy that utilises synchronised time in order to implement an informed fixed delay playout whenever possible thus minimising playout adjustments. It reverts to an adaptive approach when delays become excessive. Details of this approach can be found in [5].

Little research to date has specifically examined the precise impact of silence period adjustments within speech. Although both Ramjee et al. [1] and Moon et al. [2] cite Montgomery [6] in claiming that such distortion does not have a noticeable effect, the latter which was published in 1983 does not provide any evidence in this regard. All three simply qualify their assertion regarding the impact of silence period distortion by stating that *small* adjustments are not noticeable.

The ITU-T Recommendation P.862 (also known as Perceptual Evaluation of Speech Quality PESQ) outlines an objective method of speech quality evaluation. PESQ reacts to variations in playout delay by firstly matching up short segments of input and output speech before comparing these speech segments and as such does not capture distortion due to playout adjustments [7]. In [8], Voran argues that the distortion caused by such variations should be quantified and added to the overall impairment determined by the PESQ algorithm. Similarly, the ITU-T E-Model [9] which was intended as a planning tool to predict how the average user will rate the voice quality of a phone call (but which is increasingly being used as an analysis tool) does not capture the effect of delay variation and uses an average delay figure in its calculations. It is however recognised by the ITU-T that the E-Model requires further work to incorporate the impact of time-varying delay.

The remainder of this paper is organised as follows. Section II assesses the impact of Voice Activity Detection (VAD) and hangover settings on talkspurt/silence period distributions and the consequent impact on the performance of adaptive playout strategies. Section III examines the extent and frequency of playout adjustments resulting from the implementation of adaptive playout algorithms over both real and simulated networks and compares these with results achieved using the authors hybrid playout strategy. Section IV attempts to quantify the impact of such adjustments on MOS. Although very little work has been done in this area, the recent work by Voran provides some guidance in this regard. Section V concludes the paper.

H. Melvin is a lecturer at the Dept. of Information Technology, NUI,Galway, Ireland and is contactable at hugh.melvin@nuigalway.ie. Liam Murphy is a lecturer at the Dept. of Computer Science, UCD, Ireland and the director of UCD's Performance Engineering Laboratory.

Duration (msec)	*Percentage*
< 10	17
< 20	30
< 30	37
< 40	47
< 50	53

TABLE I
ITU P.59 SILENCE PERIODS (W/OUT HANGOVER)

Duration (msec)	*Percentage*
< 100	0
< 300	7
< 500	20
< 1000	43
< 1500	60

TABLE II
ITU P.59 SILENCE PERIODS (WITH HANGOVER)

II. TALKSPURTS/SILENCE PERIOD DISTRIBUTION

Voice Activity Detection (VAD) or Silence Detection mechanisms within codecs take advantage of the existence of silence periods in speech. This saves bandwidth (enabling multiplexing) as well as facilitates per-talkspurt adaptive playout strategies. VAD schemes vary significantly between codecs and often have configurable thresholds resulting in varying distributions of talkspurts/silence periods. Furthermore, codecs often employ *hangover* techniques whereby the codec avoids clipping the end of talkspurts and bridges over very short silences. The ITU-T P.59 [10] recommendation specifies that a model of conversational human speech should have mean talkspurt/silence periods of 227 / 596 msec respectively without hangover and 1.004 / 1.587 seconds with hangover. Other work [11] report a range of 200-400 / 500-700 msec for mean talkspurt/silence periods without hangover and approximately 1.2 / 1.8 seconds with hangover.

Table I presents data from P.59 regarding the distribution of silences without hangover.

As evident from Table I, in the absence of hangover, 30% of silences are less than 20 msec and over 50% are less than 50 msec. With hangover, the ITU model is very different as Table II illustrates, resulting in no silence periods less than 150 msec.

The use of hangover also increases the overall % of time in talkspurt mode, eg. from 27.6% to 38.7% for the ITU model.

In [12], Jiang et al. examine the impact of hangover and VAD settings on talkspurt and silence period distributions within actual speech using the NeVoT application [13] and the G.729B codec. Their findings can be summarised as follows:

- With a hangover of 20 msec, mean talkspurt and silence periods were 267 and 272 msec respectively. Approximately 40% of silence periods were less than 50 msec and 25% were greater than 200 msec.
- With a hangover time of 0 msec, 25% of talkspurts were greater than 200 msec. By increasing the hangover setting to 60 msec, 25% of talkspurts were greater than 800 msec; a further hangover increase to 280 msec resulted in 25% of talkspurts being greater than 2500 msec.
- With a VAD threshold of -50dB, approximately 1% of silence periods were greater than 6600 msec. By altering this threshold to -45 and -35 dB, the resulting silence period duration (to meet the 1% threshold) increased to 7900 and 8200 msec respectively.

A. Impact of Silence/T'Spurt Distribution on Adaptive Playout Strategies

In the absence of significant hangover and with a medium VAD silence/speech threshold, processed speech will thus contain a large number of very short silence periods i.e. less than 50msec. In such a situation, relatively small silence period adjustments of 10-20 msec will represent a large fraction of many of the silence periods in which they occur and thus may have a much greater impact on MOS than on larger silence periods. By introducing significant hangover, many short silences are bridged over resulting in much larger average silences. However, such an approach greatly restricts the sensitivity of per-talkspurt adaptive approaches in that there are fewer talkspurts. For a given packet delay distribution, this may result in a larger average adjustment as the extent of the required adjustment may build up during each talkspurt. More critically, this will result in increased late losses as the capacity of the adaptive algorithm to react to short delay spikes is much reduced. This particular point has led to the development of per-packet adaptive approaches that respond both within and between talkspurts. The degradation caused by per-packet approaches has been examined to some extent in both [3] and [4]. The former [3] reported that scaling of packets resulted in little degradation of audio quality. A $0.3 - 0.5$ score using the DMOS (Degradation Mean Opinion Score) was reported using the DCR (Degradation Category Rating) Method [14]) though they qualify this by noting that scaling occurred infrequently during the reported tests (17-24% of packets). In [4], the impact on speech quality was very much dependent on the extent of scaling and specifically they outline that an expansion of over 150% or a contraction of more than 50% will noticeably affect quality. Although such results are useful the degradation is caused by both silence period distortion *and* the impact of packet scaling. The latter comprises both temporal distortion and degradation caused by the interpolation mechanism. As such these results tell us nothing of the impact of silence period distortion alone.

III. A SURVEY OF PLAYOUT ADJUSTMENTS

In order to gauge the extent and frequency of silence period adjustments, a series of tests were carried out. These used a number of baseline per-talkspurt adaptive strategies (namely Alg. 1 and 4 from [1] and referred to hereafter as Alg. A and B respectively) as well as the authors hybrid playout strategy. The latter utilises synchronised time between sender and receiver so that each receiver has precise knowledge of end-to-end delays and can implement a fixed delay playout approach whenever network conditions allow. It reverts to a conventional adaptive approach under certain conditions and

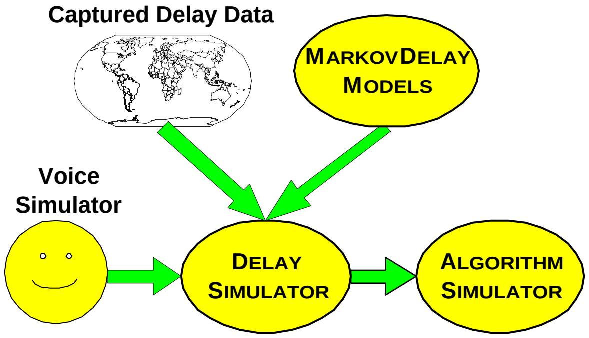

Fig. 1. Simulator

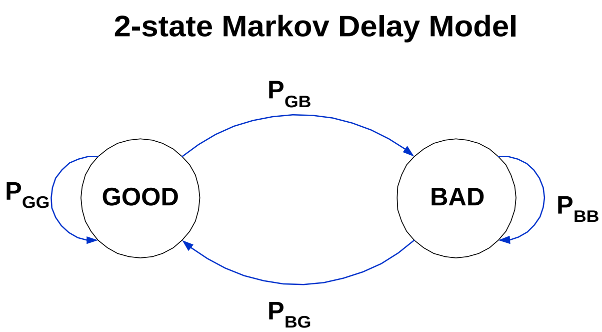

Fig. 2. 2-state Markov Delay Model

thus is termed a hybrid approach. Both Alg. A and B utilise linear recursive filters in tracking network conditions but differ in that Alg. B responds more quickly due to a lower history factor ($\alpha = 0.875$ versus 0.998 for Alg A) and also includes a spike mode that responds more rapidly to changing network conditions although it must still wait for the next silence period to do so. The performance of Alg. A, B and the hybrid were evaluated over both real and simulated network delay data. The real network delay data was captured in tests to four locations in Ireland (UCD, Dublin), the UK (UCL, London), Germany (LKN, Munich) and the US (ICIR, Berkeley). These measurements were then used within a Matlab simulator which simulated both sender voice streams and the various playout approaches. The simulated voice stream was based on actual voice traces recorded using the default settings of the openh323.org opensource VoIP application *ohphone* with a 32 msec G.711 codec frame size. The captured delay data was gathered using a modified *ping* utility configured to mimic voice data. Fig. 1 depicts the testbed design. The networks over which delay data was gathered were generally very well provisioned (and mostly limited to research institutes) and thus not representative of the public Internet. As such a series of 2-state Markov delay models were also developed to broaden the range of tests. Fig. 2 illustrates its application to delay modelling. The following summarises the most relevant characteristics of the models developed:

- BAD State Probability: This represents the percentage of packets that are affected by high delay variance.
- Average BAD State Burst Length: This determines how the BAD state packets are distributed. Much of the literature on network analysis has reported that both loss and delay have strong temporal dependency or *burstiness*. Where strong temporal dependency of delay is present, this will result in clusters of BAD state packets resulting in BAD delay bursts spanning more than one packet. Longer BAD bursts will be reflected in higher values for P_{BB} from Fig. 2.
- BAD/GOOD State Jitter Level: The delay models developed used different ranges of jitter to differentiate between GOOD and BAD states. Essentially, a GOOD state jitter metric was set (as a % of base delays) and a multiplier was applied to represent the BAD state.

An additional requirement was to ensure that out-of-order packets could not arise: in reality such events are largely due to route changes and occur infrequently and thus it was important to reproduce this. Full details of testbed design are outlined in [15].

Sum of Playout Adjustments (msec)			
	Hybrid	*Alg. A*	*Alg. B*
UCD	2.5	35	250
UCL	3	26	320
LKN	9	20	250
ICIR	9	16	160

TABLE III

DETAILS OF PLAYOUT ADJUSTMENTS

A. Results: Real Network Data

The main findings from an evaluation of the various playout strategies over real network delay data are summarised as follows:

- *Frequency of Adjustment:* The number of playout adjustments was minimal in hybrid mode relative to either Alg. A or B. Figure 3 outlines a sample delay trace to ICIR with the resulting playout strategies (In the associated legend, Alg. A is labelled Adaptive whereas Alg. B is labelled Spike) . Note that the first half of the trace was taken from the busiest network period 17:00-18:00 and shows significantly more jitter than for the second half. The remaining figures are available in [16]. There were almost 140 talkspurts in the simulated voice stream but not all these adjustments are noticeable, particularly for Alg. A. As described in [15], the number of adjustments in hybrid mode is dependent on the median delay. It thus increased from 1 for the UCD trace (geographically closest) to 8 for the ICIR trace (in the US).
- *Magnitude of Adjustment:* Table III details the sum of the absolute value of adjustments for a small subset of the measured traces.
  Although the magnitude of the majority of adjustments for Alg. A was minimal, the sum of the absolute value of such adjustments was significantly higher than for the hybrid for all traces. With Alg. B, the average adjustment was much higher than for Alg. A as was the sum of such

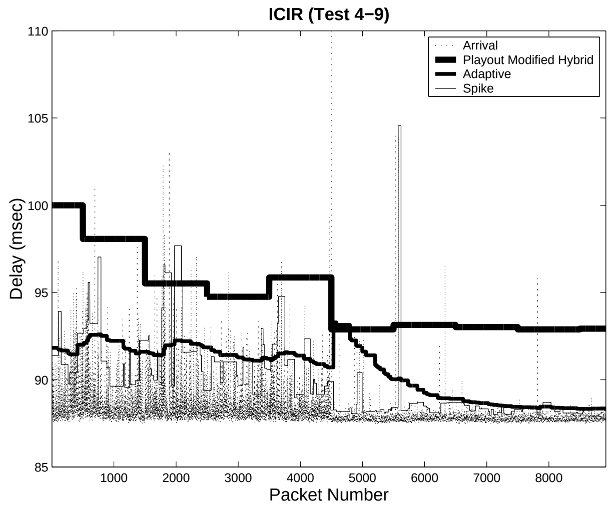

Fig. 3. ICIR: Measured Data

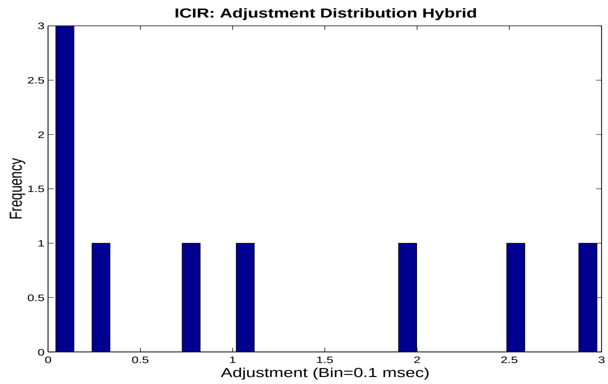

Fig. 4. ICIR: Adjustment Distribution Hybrid

adjustments. Typically, the hybrid mode had a significant step change in playout delay at the start of the session which is included in the above figures, with smaller subsequent adjustments.

Figures 4, 5 and 6 outline the adjustment distributions for the hybrid along with Alg. A/B within the ICIR trace. It is clear that the frequency of adjustments for the hybrid was much less than for either A or B. Furthermore, the extent of adjustments were quite small for both the hybrid and Alg. A (No adjustment above 3 msec) whereas for Alg. B, the extent of adjustments was much greater.

Other than the initial step change for the hybrid, the playout adjustments for the hybrid were generally of the same order as the larger adjustments for Alg. A with the magnitude of adjustments for Alg. B typically much higher. Furthermore, it is worth noting that the *recency* effect described by Clark in [17] dictates that because the large delay step change for the hybrid is typically at the start, its impact on MOS scores will be reduced. However, the scale of the adjustments for all the measured traces was minimal in the context of average silence period duration as discussed in section II. This was due to the underlying lack of congestion on these network paths. As such, it is unlikely that adjustments of the order of magnitude

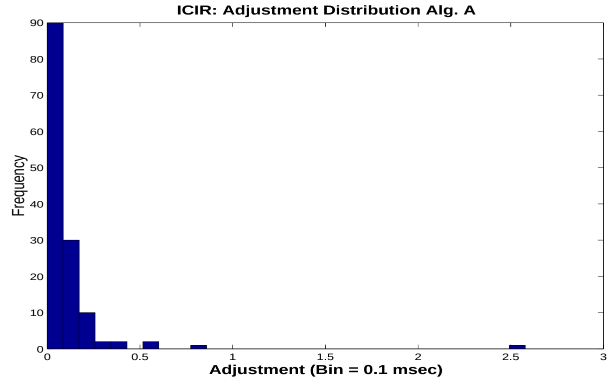

Fig. 5. ICIR: Adjustment Distribution Alg. A

Model	*Avg. BAD Burst (msec)*	*BAD State Prob.*	*Jitter (%)*	*BAD State Multiplier*	*Spike Threshold (msec)*
A	90	0.2	20	3	10
B	90	0.2	20	3	20
C	90	0.2	20	3	30
D	30	0.2	20	3	20
E	90	0.2	10	3	10
F	90	0.05	20	3	10
G	90	0.05	20	5	10
H	90	0.05	5	5	10

TABLE IV
ICIR-BASED DELAY MODELS

shown in the above traces would impact significantly on MOS scores. Furthermore, the response of the various playout strategies shown in Fig. 3 was derived from a simulator. In reality, implementing playout delay adjustments of the order of single figure msec or less is often unfeasible in the context of Operating System (OS) scheduling uncertainties and other software-based delays. A mismatch between sound card driver fragment size and codec frame size can also introduce large and varying delays which dwarf playout adjustments of this magnitude. These issues are dealt with in greater detail in [16]. Nonetheless, if the effects were measurable, then the impact for both adaptive approaches would be more significant than for the hybrid strategy, particularly for Alg. B.

B. Results: Model-derived Network Data

As outlined above, the measured delay data was limited both by the extent of testing but moreso by the fact that the underlying network paths were well provisioned and suffered little from congestion. As such a series of delay models were developed; these used the baseline delays as measured to both LKN and ICIR but jitter characteristics were varied significantly from the measured values. Table IV outlines the characteristics of the ICIR-based models developed.

Full results arising from these tests are available in [16]. In this paper, we focus on the playout adjustment distribution arising from models B and F. From Table IV, it is evident that model B had a much higher BAD state probability than F but similar BAD state characteristics.

Figures 7 and 8 illustrate respectively, the delay characteristics of models B and F along with the response of the various playout strategies. It is clear that the extent and magnitude of jitter from both models was greater than that encountered in the real delay measurement data, shown in Fig. 3.

Figures 9 and 10 outline for models B and F, the distribution of playout adjustments for Alg. A whereas Figures 11 and 12 outline the distribution for Alg. B. As with the playout distributions for the real ICIR data, the distribution for Alg.B is much broader than for Alg. A. The extent of distributions is however much greater with both models than for the real data. This is particularly so for Alg. B and Model B where 70 adjustments were in the range 10-50 msec. The hybrid playout strategy has very few adjustments relative to either Alg. A or B, similar to the findings from the real network data.

As outlined in section II, the extent to which short silence periods (in the range 10-50 msec) occur in packetised speech depends on hangover and VAD settings. If such short silence periods are present, we suggest that playout adjustments in the same range will have a much more significant impact on speech quality than if they occur during longer silences. Although little research has been done to quantify this impact, the work of Voran [8] deals peripherally with the issue and is discussed in the next section.

IV. Likely Impact on MOS of Playout Adjustments

In [8], Voran evaluated the impact of *temporal discontinuities* and packet loss on speech quality. Temporal discontinuities included both silence period insertion and removal (without compensation) of active speech packets within speech. Voran suggests that the distortion caused by such impairments should be quantified and added to the overall impairment determined by the PESQ algorithm. Furthermore, he argues that due to the persistence of best effort networks, receiver jitter buffer under and overflows will continue to occur in VoIP applications resulting in voice impairment. In this context a range of experiments were carried out to quantify the impact on MOS of such impairments.

He introduced three impairments termed *loss, jump and pause* to speech where *loss* refers to conventional packet loss and was compensated for through PLC, *jump* refers to temporal contraction of speech by dropping packets and *pause* refers to temporal elongation of speech through silence packet insertion. As such the *pause* impairment is of most interest as it most closely reflects the type of impairment caused by per-talkspurt playout strategies in that it simply adds silences. The impact of both magnitude and frequency of each of the three impairment were examined independently as well as a combination of *pause* and *jump*. Impairments were added at random locations within G.723-encoded speech. His findings are summarised as follows:

- For a given frequency and magnitude of impairment, the impact of the four impairments (loss, pause, jump, pause and jump) on MOS scores was found to be roughly similar.
- As the magnitude of impairment increased, the reported MOS scores decreased at an almost linear rate. For

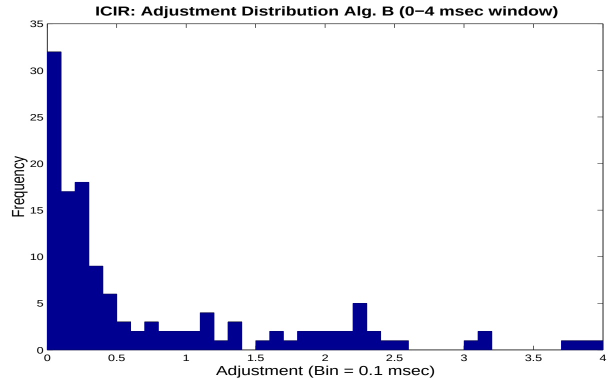

Fig. 6. ICIR: Adjustment Distribution Alg. B

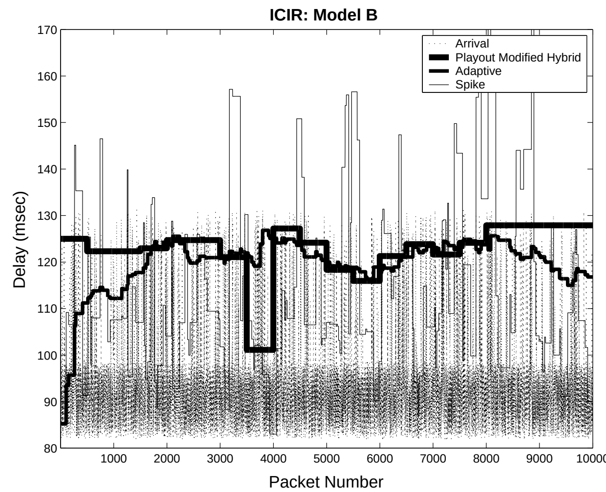

Fig. 7. ICIR Model B

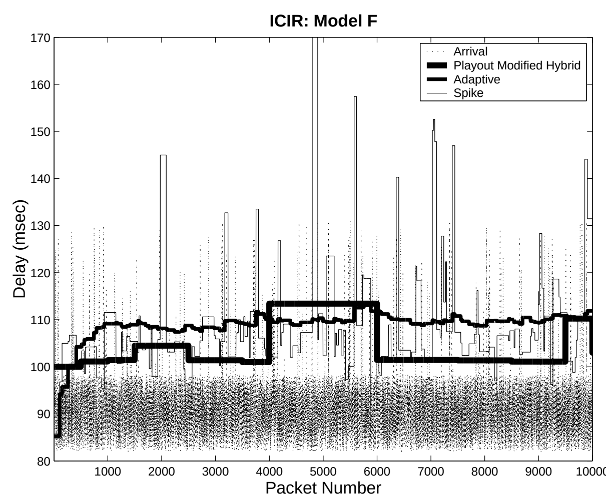

Fig. 8. ICIR Model F

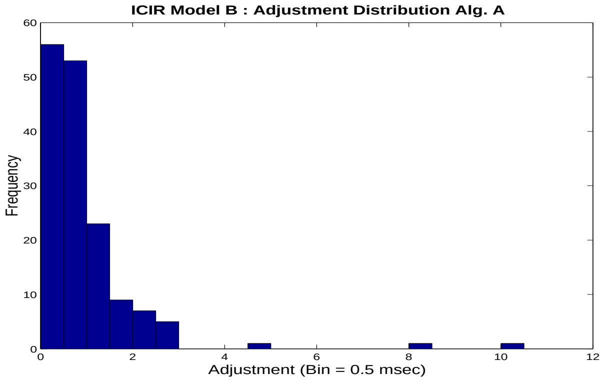

Fig. 9. ICIR Model B: Alg. A

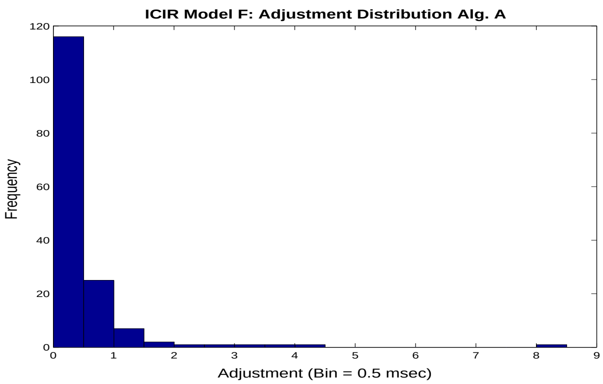

Fig. 10. ICIR Model F: Alg A

example, at a frequency of one impairment per 100 frames, a 30/60/120 msec *pause* impairment resulted in the MOS score dropping by 0.21/0.41/1.15 respectively.

- As the frequency of impairment increased, the reported MOS scores decreased at a non-linear rate. For example, at an impairment magnitude of 30 msec, an impairment frequency of 1/2/4 impairments per 100 frames resulted in the MOS score dropping by 0.21/0.22/0.45 respectively. This means that users found a single large impairment more noticeable than multiple small impairments. Note that from section III-B the response of Alg. B to model-

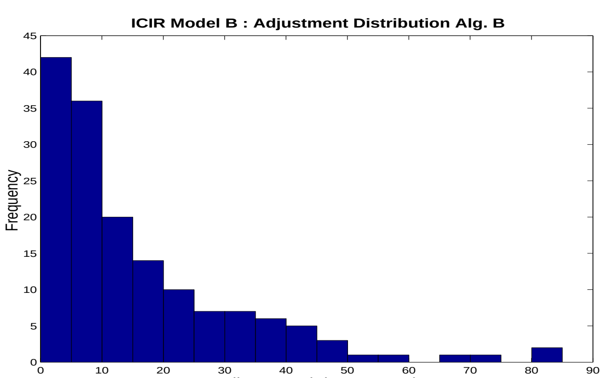

Fig. 11. ICIR Model B: Alg. B

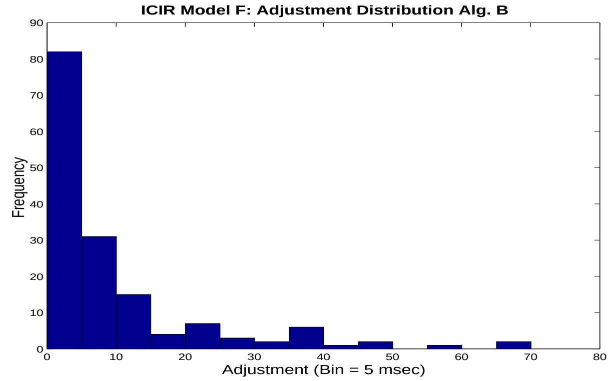

Fig. 12. ICIR Model F: Alg B

derived delay data in Fig. 11 resulted in 70 adjustments in the range 10-50 msec over a test size of 10,000 packets which is slightly less than 1 impairment per 100 packets. Note that with regard to packet loss, Cox et al. in [18] reported somewhat different results in that for high average loss rates, bursty loss resulted in better MOS scores than for random loss, whereas for lower average rates, the converse was found to be true.

As Voran's work randomly distributed the impairments throughout active speech (talkspurts) only, his results of *pause* impairments cannot be directly applied to per talkspurt playout strategies where the impairments only occur during silences (i.e. the silence period is contracted or elongated). However, we apply his work indirectly as follows:

- As outlined in section II, the distribution of talkspurts and silence periods is very much dependent on VAD and hangover settings. In order to ensure that per-talkspurt approaches react quickly to rapidly changing network conditions (such as delay spikes) and thus minimise late packet loss, it is vital that such settings are correctly set. Where such settings result in many very short silence periods, a per-talkspurt playout strategy will respond quickly to changing network delay. From section III-B, we showed that under conditions of significant jitter, the extent of adjustments using adaptive approaches (particularly one similar to Alg. B) can be of a similar order of magnitude to many of the silence periods. This in turn will result in many such silences being significantly distorted. On the other hand, the same original speech with different VAD/hangover settings applied will result in much larger silence periods/talkspurts with very few short silences. This will limit the effectiveness of adaptive playout strategies in that many silence periods from the first scenario are represented as active speech segments (talkspurts) in the second.
- A comparison of both speech signals (after processing by an adaptive playout strategy) will thus show distortion (due to the adaptive playout strategy) of the first signal relative to the second. As such, the additional silence period adjustment within the first signal is effectively similar to the *pause/jump* impairment being introduced into active speech segments by Voran, though Voran

applied them randomly.

- Voran's results therefore represent a very worst-case scenario of the impact of silence period distortion on MOS scale. Undoubtedly, the precise impact of such distortion is influenced by the scale of the adjustment relative to the overall silence period which complicates the issue. More specific research is thus required to accurately determine the precise impact of playout adjustments (for a given network delay characteristic and adaptive playout strategy), taking into account factors such as the relative size of adjustments, which in turn is influenced by VAD/hangover settings.
- A further outcome of Voran's work that needs to be reconsidered in the context of adaptive playout strategies is that large step adjustments should be avoided and should be spread out over a number of talkspurts.

V. Conclusions

In this paper, the extent and impact of playout adjustments, characteristic of many adaptive playout approaches is examined. The distribution of talkspurts and silence periods is greatly influenced by VAD and hangover settings which in turn greatly limit the effectiveness of adaptive approaches in tracking network conditions. Implementing responsive settings will result in many short silence periods which will limit late losses but which will also most likely result in more significant MOS degradation. We show that where networks are well provisioned, the extent of adjustments is unlikely to impact noticeably on the MOS scale. For congested networks, adaptive algorithms (particularly those with spike compensation mechanisms such as Alg. B) will result in frequent and large adjustments. With responsive VAD/hangover settings, these adjustments may constitute a significant percentage of many of the silence periods. Although little research has examined the precise impact of such adjustments on the MOS scale, we review Voran's work and apply it to a limited degree in this area. Further research is however required to formally quantify the impact of playout adjustments. In any event, we also show that the hybrid playout strategy proposed by the authors greatly limits the extent of adjustments relative to adaptive approaches and thus will result in minimal MOS degradation.

References

[1] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide-area networks," *Proc. Conf. Comp. Comm. (IEEE Infocom)*, IEEE CS Press, Los Alamitos, Calif., June 1994, pp. 680-688.

[2] S. Moon, J. Kurose, and D. Towsley, "Packet audio playout delay adjustment:performance bounds and algorithms," *ACM/Springer Multimedia Systems*, vol. 6, pp. 17-28, Jan. 1998.

[3] Y. Liang, N. Farber, and B. Girod, "Adaptive Playout Scheduling using Time-scale Modification in Packet Voice Communications," *Proceedings of IEEE International Conference on Ascoustics, Speech, and Signal Processing ICASSP 2001*, Salt Lake City, Utah, May. 2001.

[4] F. Liu, J. Kim, and C. Jay Kuo, "Quality Enhancement of Packet Audio with Time-Scale Modification," *Proceedings of SPIE ITCOM 2002 Multimedia Systems and Applications*,Boston, Jul. 2002.

[5] H. Melvin and L. Murphy, "An evaluation of the potential of synchronized time to improve VoIP quality", *IEEE Proceedings of International Conference on Communications (ICC 2003)*, Anchorage, May 2003.

[6] W. Montgomery, "Techniques for Packet Voice Synchronization," *IEEE Journal on Selected Areas in Comm.*, vol. SAC-1, no. 6, Dec. 1983.

[7] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," ITU, 2001.

[8] S. Voran, "Perception of Temporal Discontinuity Impairments in Coded Speech- A Proposal for Objective Estimators and Some Subjective Test Results," *Proceedings of the 2nd International Conference on Measurement of Speech and Audio Quality in Networks*, Prague, May 2003.

[9] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning," ITU, Mar. 2003.

[10] ITU-T Recommendation P.59, "Telephone transmission quality objective measuring apparatus: Artificial conversational speech," ITU, Nov. 1996.

[11] P. Brady, "A technique for investigating on-off patterns of speech," *Bell System Technical Journal*,44(1), Jan. 1965.

[12] W. Jiang and H. Schulzrinne, "Analysis of On-Off Patterns in VoIP and Their Effect on Voice Traffic Aggregation," *Proceedings of International Conference on Computer Communications and Networks (ICCCN 2000)*, Las Vegas, Oct. 2000.

[13] H. Schulzrinne, "Voice Communication across the Internet," *Technical Report TR 92-50*, Department of Computer Science, Univ. of Massachusetts, Jul. 1992.

[14] ITU-T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality, Annex D- Degradation Category Rating (DCR) Method," ITU, Aug. 1996.

[15] H. Melvin and L. Murphy, "An Evaluation of Delay-Aware Receiver Playout Strategies for VoIP Applications," *Proceedings of International Federation for Information Processing (IFIP Networking 2004)*, Athens, May. 2004, and published in *Proceedings of LNCS*, Springer-Verlag 2004.

[16] H. Melvin, "The Use of Synchronised Time in Voice over IP (VoIP) Applications," *PhD Thesis*, University College Dublin, Oct. 2004.

[17] A. Clark, "Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality," *Proceedings of IP Telephony Workshop*, NY, Mar. 2001.

[18] R. Cox, R. Perkins, "Results of a Subjective Listening Test for G.711 with Frame Erasure Concealment," TIA1.7/99-016, May 1999.