2013

DETAILED COMPARATIVE ANALYSIS OF PESQ AND VISQOL BEHAVIOUR IN THE CONTEXT OF PLAYOUT DELAY ADJUSTMENTS INTRODUCED BY VOIP JITTER BUFFER ALGORITHMS

Andrew Hines
Technological University Dublin, andrew.hines@tudublin.ie

Peter Pocta
University of Zilina, Slovakia

Hugh Melvin
National University of Ireland, Galway

DETAILED COMPARATIVE ANALYSIS OF PESQ AND VISQOL BEHAVIOUR IN THE CONTEXT OF PLAYOUT DELAY ADJUSTMENTS INTRODUCED BY VOIP JITTER BUFFER ALGORITHMS

Andrew Hines[‡*], *Peter Počta*[†] *and Hugh Melvin*[⋆]

[‡] Sigmedia, Trinity College Dublin, Ireland
[†] Department of Telecommunications and Multimedia, FEE, University of Žilina, Slovakia
[⋆] College of Engineering & Informatics, National University of Ireland, Galway, Ireland

ABSTRACT

This paper undertakes a detailed comparative analysis of both PESQ and VISQOL model behaviour, when tested against speech samples modified through playout delay adjustments. The adjustments are typical (in extent and magnitude) to those introduced by VoIP jitter buffer algorithms. Furthermore, the analysis examines the impact of adjustment location as well as speaker factors on MOS scores predicted by both models and seeks to determine if both models are able to correctly predict the impact on quality perceived by the end user from earlier subjective tests. The earlier results showed speaker voice preference and potentially wideband experience dominating subjective tests more than playout delay adjustment duration or location. By design, PESQ and VISQOL do not qualify speaker voice difference reducing their correlation with the subjective tests. In addition, it was found that PESQ scores are impacted by playout delay adjustments and thus the impact of playout delay adjustments on a quality perceived by the end user is not well modelled. On the other hand, VISQOL model is better in predicting an impact of playout delay adjustments on a quality perceived by the user but there are still some discrepancies in the predicted scores. The reasons for those discrepancies are particularly analysed and discussed.

Index Terms— Adaptive jitter buffer algorithm, playout adjustments, PESQ, VISQOL, speech quality, VoIP

1. INTRODUCTION

The default best-effort Internet presents significant challenges for delay-sensitive applications such as VoIP. To cope with non determinism, receiver playout strategies are utilised in VoIP applications that adapt to network condition. Such strategies can be divided into two different groups, namely per-talkspurt and per-packet. The former make use of silence periods within natural speech and adapt such silences to track network conditions, thus preserving the integrity of active speech talkspurts. Examples of this approach are described in [1, 2]. Per packet strategies are different in that adjustments are made both during silence periods and during talkspurts by time-scaling of packets, a technique also known in the literature as time-warping. This approach is more effective in coping with short network delay changes because the per talkspurt approach can only adapt during recognized silences even though the duration of many delay spikes may be less than that of a talkspurt. This approach however introduces potential degradation caused by the scaling of speech packets. Examples of this approach are described in [3, 4] and such techniques are frequently deployed in popular VoIP applications such as GoogleTalk and Skype. In this research, we focus on applications that deploy per talkspurt strategies, which are commonly found in current telecommunication networks.

It has been shown in [5] that the impact of silence period adjustments (playout delay adjustments introduced by VoIP jitter buffers) on subjective quality scores is insignificant. To the best of our knowledge, the subjective results presented in [5] are a first proof of the assertion published frequently in the literature [1, 2, 6] that playout adjustments, typical of those introduced by jitter buffers in VoIP scenarios do not have a noticeable effect on quality perceived by the end user. It has been also found in [5] that such playout adjustments have however a significant impact on objective MOS scores predicted by the PESQ model [7, 8]. In particular , a strong negative correlation between the extent of adjustments and objective scores predicted by PESQ has been noted. Finally in [5], the impact of the position in the sample where adjustments are made has been reported to be insignificant for subjective scores but significant for objective scores predicted by PESQ model, especially for higher range of adjustment magnitudes.

Arising from these findings, which indicate significant differences between subjective and PESQ results, this paper further investigates the dominant factors influencing the subjective and objective tests in the context of playout delay adjustments introduced by VoIP jitter buffers. Furthermore, we investigate the behaviour of an alternative objective test method, termed VISQOL [9, 10] and compare findings with PESQ.

In summary, we identified a number of key research questions, namely:

- What are the dominant factors in MOS scores, arising from subjective tests, and objective tests using both PESQ and VISQOL?
- Is the impact of playout delay on quality perceived by the end user being addressed correctly by full reference objective speech models?
- Which of the objective metrics examined is better suited to predicting the impact of playout delay adjustments?

The remainder of this paper is structured as follows. Section 2 introduces the VISQOL model. Section 3 outlines our simulator-based approach to generating the impaired speech samples used for the subjective and objective comparisons. Section 4 presents and discusses experimental results. Section 5 concludes the paper and suggests some areas for future research arising from this paper.

*Thanks to Google, Inc. for funding. Email: andrew.hines@tcd.ie

2. VISQOL MODEL

The Virtual Speech Quality Objective Listener (VISQOL) is an objective speech quality model of human sensitivity to degradations in speech quality [9, 10]. It is a signal based full reference metric that uses a spectro-temporal measure of similarity between a reference and a test speech signal and has been trained and tested with narrowband speech under a wide variety of degradations. VISQOL aims to predict the overall quality of experience for the end listener whether the cause of speech quality degradation is due to codec choice, ambient noise, or transmission channel degradations. The model has three major processing stages: pre-processing, alignment and comparison. The pre-processing stage scales the test signal to match the reference signal's sound pressure level. Short-term Fourier Transform (STFT) spectrogram representations of the reference and test signals are created using critical bands between 150 and 3,400 Hz. A 256 sample, 50% overlap Hamming window is used for signals with 8 kHz sampling rates. The reference signal is segmented into patches 0.48 seconds long (30 frames, each 16 msec) by 16 critical frequency bands as illustrated in Figure 1. VISQOL can be configured to use more bands (21 bands up to 8kHz) if the signals to be tested are wideband. Each reference patch is aligned with the corresponding area from the test spectrogram. The Neurogram Similarity Index Measure (NSIM) [11] is used to align and measure the similarity between the reference and a test spectrogram patches, as in Figure 1. The NSIM score per patch is averaged over the patches to yield the similarity metric for the test signal. It is transformed into a MOS-LQO as the signal similarity estimate.

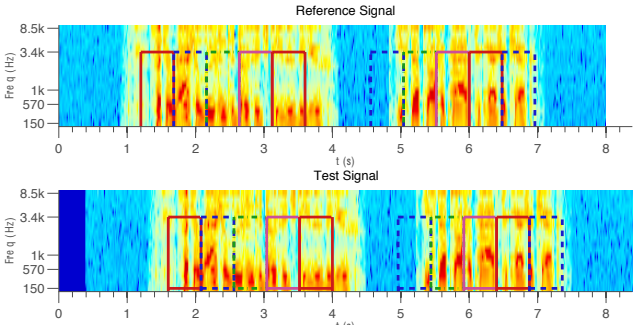

Fig. 1. VISQOL patch alignment and comparison example

3. METHODOLOGY

In this section, the methodology used to generate the speech samples for the analysis is described and then details of the testing process are presented. The overall methodology consists of a number of stages as follows:

- Generate a series of network packet delays (D), consistent with varying network conditions
- Using these delays (D), and simulated voice patterns (V), generate a series of playout adjustments (A) that are typical of adaptive VoIP playout strategies under the above network conditions
- Apply these set of adjustments (A) to different locations within reference speech samples

Using the derived test samples, an objective test programme using both PESQ and VISQOL models was carried out.

Figure 1 depicts the playout adjustment simulator implemented in Matlab. As can be seen in Figure 1, the simulator consists of three separate module blocks, namely:

- Voice Simulator Block
- Delay Simulator Block
- Playout Algorithm Simulator Block.

More information about the simulator and its blocks can be found in [5, 12].

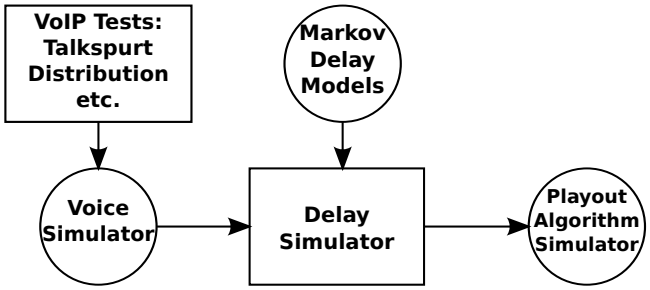

Fig. 2. Playout adjustment simulator

The overall simulator works as follows. First of all, user specifies a talkspurt distribution file which is extracted from live speech and loaded as an array (V) e.g. [1,0,0..]. Secondly, user defines network conditions for test. An array of network delays (D) corresponding to the specified network conditions is returned by delay block of the simulator. The input parameters are defined in second column of Table 1. Third column of Table 1 specifies exact values of input parameters used in the simulations. As can be clearly seen in Table 1, parameters 1–4 were kept constant while parameters 5–7 were varied to give different network characteristics, ranging from stable to unstable.

No.	Name of input parameter	Exact value/s used in the simulations
1	Number of packets (-)	4000
2	Packet interval (ms)	20 ms
3	Base delay (ms)	50 ms
4	BAD state burst length (ms)	200 ms (10 packets)
5	BAD state probability (%)	20, 40, 60, 80%
6	GOOD state jitter (% of base delay)	25, 50%
7	BAD state jitter multiplier (-)	2, 3, 4

Table 1. Input parameters of playout adjustment simulator and their exact values used in the simulations

The input parameters defined in Table 1 and used in the simulations, particularly those relating to jitter, were derived following a series of detailed delay and jitter measurements between Ireland and the US/mainland Europe. Further details can be found in [13]. More recent measurements described in [13, 14] have highlighted the particular problem of very high jitter/delay in congested IEEE 802.11 networks.

As a next step, delay values (D) generated by the delay block of the simulator were applied to adaptive jitter buffering (AJB) algorithms 1 and 2 (Alg. 1 and 4 from [1]) using the talkspurt distribution

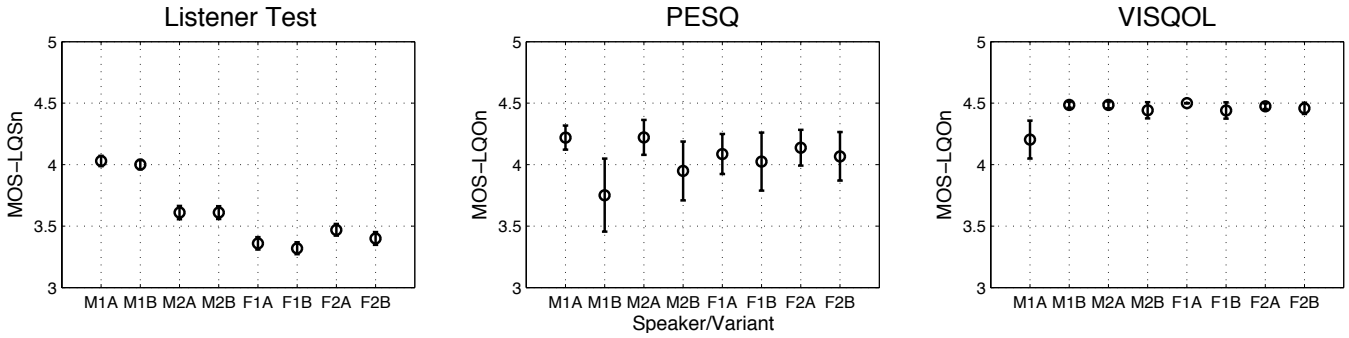

Fig. 3. Dominant Experimental Factors. The results are aggregated by speaker (e.g. M1 is male speaker) 1 and by playout variant (i.e. A or B). The subjective scores and objective estimates from PESQ and ViSQOL are presented along with the 95% confidence intervals.

array (V). It creates a series of playout adjustments (A) for every network test condition and playout algorithm. Details of the complex tests using the simulator are presented in [15]. In total 24 different network delay models were used, producing 24 network delay arrays (D). These delay values were fed to 2 different playout algorithms. For each playout algorithm, tests were repeated using different parameters such as alpha (history weighting - varied from 0.8 to 0.998), beta (jitter multiplier - varied from 4 to 6) and spike mode threshold (only Alg. 2). Each combination generated a distinct set of playout adjustments (A) for each test scenario. It is worth noting that voice samples (V) were based on 80 seconds of active speech with 40 talkspurts (Marker bits = 1) whereas the speech samples chosen for this experiment were 8 seconds long thus this also had to be taken into account. Basically a pro-rata approach was deployed in this case. In other words, the 80 seconds of speech used as input to the Voice Simulator block contained 40 playout adjustments so for our 8 second ITU-T speech samples, we implemented 4 adjustments. Arising from the full range of test combinations described above, which numbered 96, a subset of 12 sets of playout adjustments (A') containing 4 adjustments each were taken to represent a spectrum of network conditions ranging from a stable network to an unstable network. Table 2 presents the actual playout adjustments selected that were applied to the speech samples.

Four reference speech samples were used as is the norm for quality testing. The English subset of ITU-T P.23 [16] database was used for speech material, consisting of a pair of utterances with a small pause between the utterances. Two male and two female speakers were included in the stimuli. The speech samples used were 8 seconds in length and stored in 16 bit, 8000 Hz linear PCM.

All speech samples were altered by inserting and removing silence periods to reflect the adjustments typical of those introduced by adaptive jitter buffering playout algorithms. As can be seen in Table 2, the adjustments were a mix of positive and negative adjustments summing to zero (adding and removing silence periods).

As the final step, the four adjustments were applied to the samples in two different locations (referred to hereafter as variant A and B). The location of the adjustments is considered a further experimental variable. The difference between variant A and B is that the impairments in variant B were applied in the latter part of each sample. A free sound editor was deployed in order to introduce the adjustments into the samples. Details can be found in [5, 15]. The overall result of this sampling created 96 speech samples (4 voices x 12 test conditions x 2 variants). It should be noted here that the same reference and degraded samples were used in [5].

In the final step, the 100 samples (essentially the 96 degraded samples (using test conditions No.1-12) and 4 reference samples (Ref condition)) were processed by PESQ and VISQOL model in order to get objective quality scores. The output of PESQ model (raw PESQ scores) was converted to MOS-Listening Quality Objective narrowband (MOS-LQOn) values as defined in [17].

The subjective values presented in this paper were obtained from ACR subjective listening test performed in accordance with ITU-T Recommendation P.800 [18]. In every case, up to 2 listeners were seated in a small listening room (acoustically treated) with a background noise well below 20 dB SPL (A). All together, 30 naïve listeners (16 male, 14 female, 20-55 years, mean 34.43 years) participated in the test. All subjects were Irish Nationals whose first language was English. Participants were remunerated for their efforts. The samples (96 degraded samples + 4 reference samples) were played out using high quality studio equipment in a random order and diotically presented over Sennheiser HD 455 headphones (presentation level: 73 dB SPL (A)) to the test subjects. The results of the opinion scores from 1 (bad) to 5 (excellent) were averaged to obtain MOS-Listening Quality Subjective narrowband (MOS-LQSn) values for each sample.

4. RESULTS AND DISCUSSION

Previously reported results in [5] showed that PESQ scores exhibited poor correlation with the scores obtained from the subjective test. This fact has motivated us to realise a detailed analysis of PESQ behaviour in this context. By repeating the experiment here with an alternative objective speech quality prediction model, VISQOL, allowed comparison of sensitivity to playout delay for two full reference objective metrics.

Three key research questions defined in section 1 are addressed in following subsections.

4.1. Dominant Experimental Factors

Analysis of the subjective listener test results in [5] highlighted a significant preference to male speaker 1 (M1). As described in section 3, four speakers were used in the tests, with adjustments applied in two different locations , resulting in variant A tests and variant B tests. An analysis of the results broken down by speaker and variant was carried out for the subjective and objective metric results to establish the dominant factor in the MOS score trends.

Figure 3 shows the results of :

	Adjustments (ms)				Absolute sum of adjustments (ms)
Test conditions	**1st**	**2nd**	**3rd**	**4th**	
Ref	0	0	0	0	0
1	2	-2	3	-3	10
2	4	-4	-4	4	16
3	3	-3	-6	6	18
4	5	-5	-5	5	20
5	3	-6	-7	10	26
6	16	-12	-8	4	40
7	10	-17	-6	13	46
8	10	-15	-10	15	50
9	8	-23	-3	18	52
10	5	10	-30	15	60
11	-15	15	-15	15	60
12	-25	22	-8	11	66

Table 2. Playout adjustments applied to speech samples

- Subjective test averaged by speaker (e.g. M1 is male speaker 1) , for both playout variant (i.e. A or B) across all conditions. 95% interval also shown.
- PESQ test averaged by speaker as above
- VISQOL test averaged by speaker as above

Examining the subjective results first, the clear preference for the M1 speaker over all others is very evident, although there is also smaller preference towards M2 relative to F1 and F2 speakers. The subjective test results also show a small but statistically insignificant preference for variant A over variant B which was analysed in [5]. The 95% confidence intervals show that the influence of playout delays was much smaller than between speakers. This points towards the speaker voice being the dominant quality factor perceived in the listener tests. In addition to the speaker factor and as already discussed in [5], we speculate that perhaps the subjects opinion has been also affected by their previous long-term experience with wideband telephony (wideband speech), though this was not validated.

Looking at the PESQ results next, the most striking difference is the size of the confidence intervals. These indicate that across all speakers and variants A/B there was a large variation in the predicted quality scores. The results do not show a significant difference between speakers, except in the case of M1B. This could have pointed towards an issue with the playout delay locations although as it was not consistent with the VISQOL results, it points towards it being a metric rather than data issue. Overall, the results show that PESQ is much more sensitive to playout delay adjustments than real listeners.

In contrast with the PESQ quality predictions, the VISQOL model predictions has 95% interval ranges that are of the same order as the subjective tests, indicating negligible impact of playout adjustment. Aside from the M1A results, there was very little variance predicted across speaker or variant A/B.

Overall, this analysis points towards speaker voice being one of the dominant factors for the listeners in the subjective test. Playout delay condition and location are the dominant factors for PESQ and VISQOL does not display sensitivity to either factor, although the M1A results indicated a potential delay-related factor that is addressed below in section 4.3.

4.2. Capture of Playout Delay by Objective Metrics

The results presented in [5] showed that PESQ had a low correlation with subjective scores when assessed by network condition ($R = 0.17$ and $R = -0.15$ for conditions A and B respectively). Correlation with subjective results for VISQOL proved equally low ($R = -0.61$ and $R = 0.20$ for conditions A and B respectively). The dominant experimental factor analysis hints that this is due to subjective listeners being significantly more sensitive to voice type/quality than relatively small playout delay adjustments. Moreover and as already speculated in [5], exposure to wideband telephony (discussed also in section 4.1) probably also influenced the subjective results and thus the performance of PESQ and VISQOL model (correlation between the objective and subjective data) in this experiment.

Our experiments tested the reference signals and 12 playout delay adjustment conditions (with progressively increasing absolute adjustment sums, as shown in Table 2). Figure 4 presents the predicted MOS-LQO for PESQ and VISQOL across each condition, separately shown for each speaker and playout variant (A and B).

A decreasing quality trend from left to right is apparent across conditions in the PESQ scores, but not in the VISQOL scores, where the M1A variant is the major anomaly in results that otherwise show an insensitivity to playout delay. This explains the large confidence intervals exhibited by PESQ in Figure 3, being as a result of the playout delays. It further shows that not only is playout delay the dominant factor, but PESQ is also ranking the progressive delays in its predicted quality scores. PESQ is thus somehow sensitive to playout delay and ranks it closely with the actual delays. However, as the listener tests showed that listener were insensitive to these types of degradation, the VISQOL results are preferable from a QoE perspective.

Due to the nature of full reference objective models, they can only perceive quality relative to a reference. Thus, for all speakers, the reference condition will inherently be at the maximum in the range and all degraded versions will be at or below this value. The voice quality is not taken into account in any way. Accordingly, we suggest that for the subjective tests, the impact of voice quality, the possible exposure to wideband telephony, and the insensitivity to playout delay leading to MOS scores in the 3 to 4 range (even for the reference conditions) would appear to account for the low correlation between subjective scores achieved by both models, particularly PESQ.

4.3. VISQOL Model Results

The discrepancy in the male speaker 1, variant A (M1A) results was investigated in detail as it was at odds with all of the other speaker and playout delay location variations which displayed insensitivity to playout delay. In Figure 4 it is clear that M1A conditions 7,8,9 and 12 are the major contributors to the low quality scores seen from VISQOL.

Figure 5 shows the signals in the time domain with the VISQOL patch time boundaries marked. It should be stressed the similarity is done on spectro-temporal similarity but the speech signal is shown in the figure for explanation purposes. As outlined in the section 2, VISQOL calculates the predicted speech quality aligning patches from the reference signal spectrogram to patches from the degraded signal and calculating their similarity. The mean similarity of all

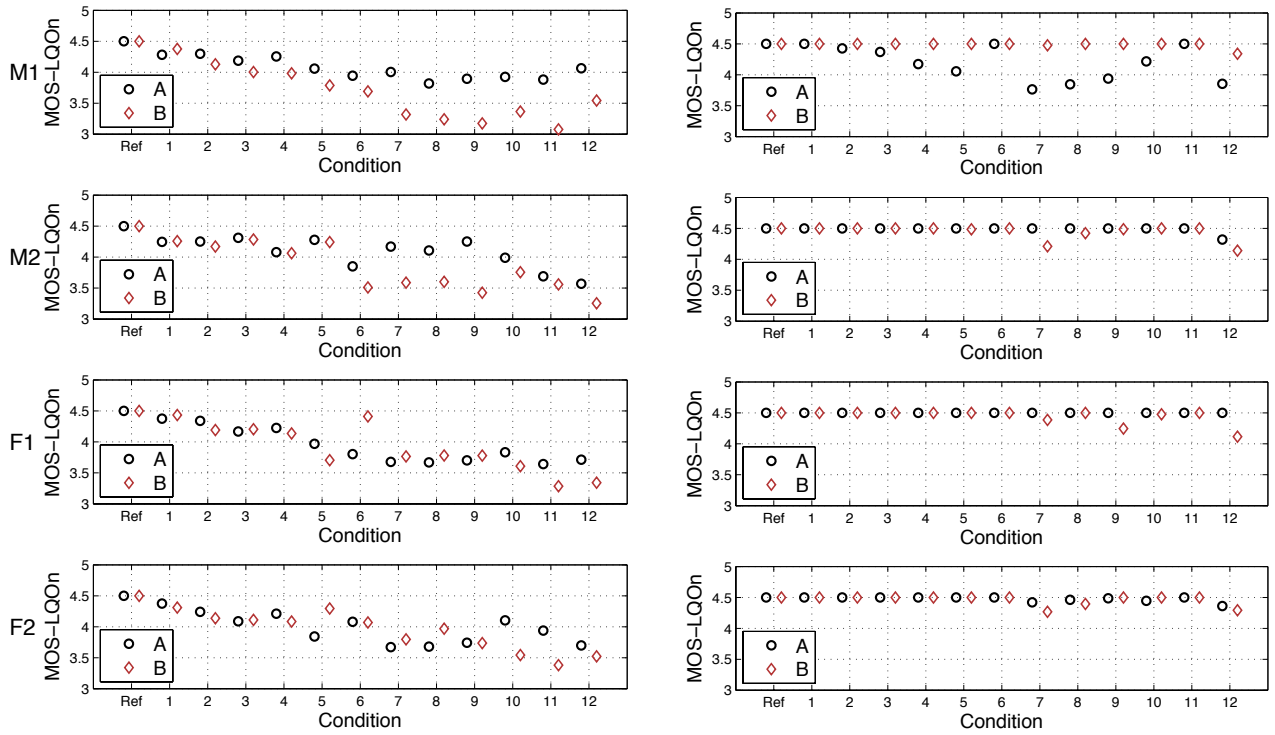

Fig. 4. Sensitivity of Objective Metrics to playout delay. PESQ (left) and VISQOL (right) results broken down by speaker (M1,M2,F1,F2) and condition (A,B). A decreasing quality trend from left to right is apparent across conditions in the PESQ scores, but not in the VISQOL scores, where the M1A variant is the major anomaly in results that otherwise show an insensitivity to playout delay.

patches is then mapped to a MOS-LQO prediction. The sentence in Figure 5 is one of the M1A results and it appears that alignment problems have caused the drop in similarity scores seen in the four conditions that were significantly lower than the rest of the tests. The patches adjacent to patches with adjustments have similarity scores as low or lower than those where the adjustments occur. This implies the quality drop is as a result of progressive misalignment rather than purely as a result of the playout delay alterations to the signal.

Figure 6 illustrates an alternative scenario where as a result of the adjustments all occurring close together, the impact is confined to a small number of patches. The impact of the playout delay adjustments has been reduced as VISQOL calculates the overall similarity score as a mean of the patch similarities.

The patch alignment strategy and mean similarity across patches calculation used by VISQOL does appear to be the reason it is insensitive to playout delays. However, the mis-alignment problem illustrated in Figure 5 does appear to be a design issue that needs to be investigated further.

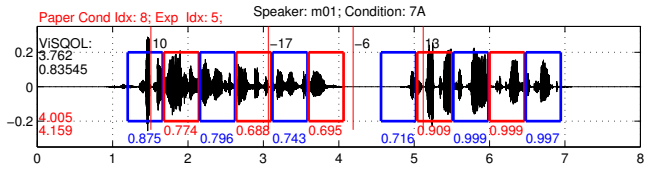

Fig. 5. Male 1 Speaker, Condition 7A. First and second adjustments (+10 ms and -17 ms), causes a knock on effect to next three patches alignments. Third adjustment impacts next patch but 4th doesnt impact final 3 patches. It appears that a mis-alignment of patches has occurred as a result of the first playout delay but not as a result of the 4th playout delay.

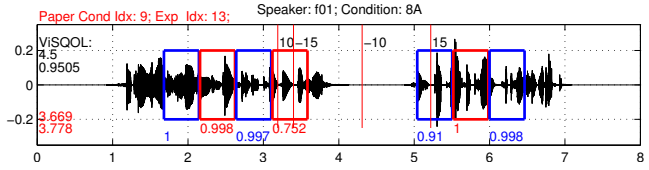

Fig. 6. Female 1 Speaker, Condition 8A. No misalignments have occurred in this sample but as two of the playout delays occur within the same patch and one occurred during a silence where there was no patch comparison, the overall impact of the 4 adjustments is only computed within two of the patches.

5. CONCLUSIONS & FUTURE WORK

In this paper, we have done a detailed analysis of both PESQ and VISQOL model behaviour, when tested against speech samples modified through playout delay adjustments. Three specific questions, outlined in section 1 were addressed in this study. Regarding the first question, we report that the speaker voice and potentially wideband experience were the dominant factors for listeners in the subjective test. On the other hand, the extent of playout delay adjustment and its location were dominant factors for PESQ. Like PESQ, VISQOL's reported correlation with subjective tests was low due to

the voice quality factor. However, VISQOL's insensitivity to playout adjustments was comparable to the subjective test aside from the one discrepancy that was analysed in detail in section 4.3. Regarding the second question, we show that PESQ is somehow sensitive to playout delay and there is negative correlation with the adjustment magnitude. However, the earlier subjective test has shown that listeners were insensitive to these types of degradation, and this should be reflected in the scores provided by PESQ model. In other words, it seems that an impact of playout delay adjustments on a quality perceived by the end user is not properly modeled by PESQ. Although the PESQ model is designed with a time-alignment module, it has to be emphasised here that the model was not explicitly verified for playout delay adjustments resulting from VoIP applications with adaptive buffering over congested networks. As such our research represents a somewhat out-of-domain use case for this model. On the other hand, the predictions provided by VISQOL model are more in line with the playout adjustments quality perception of listeners involved in the subjective test (the auditory ratings). Moving to the third question, and on the basis of the presented results, we can conclude VISQOL model seems to be more sufficient than PESQ for predicting an impact of playout delay adjustments introduced by VoIP jitter buffers on a quality perceived by the end user. However the current VISQOL model also appears to have some alignment shortcomings impacting on results under certain conditions.

Arising from this paper , we see the need for further research to address three open questions. What precise mechanisms within PESQ cause it to react significantly to both the magnitude and location of playout adjustments, leading to poor correlation with subjective results? Why were subjective test scores so low, even for reference samples from the ITU-T P.23 database? Can the mis-alignment of VISQOL patches seen in some combinations of speaker and playout adjustments be addressed?

6. REFERENCES

[1] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide-area networks," in *INFOCOM, Proceedings of IEEE*, 1994, pp. 680–688.

[2] S. B. Moon, J. Kurose, and D. Towsley, "Packet audio playout delay adjustment: performance bounds and algorithms," *Multimedia Systems*, vol. 6, no. 1, pp. 17–28, 1998.

[3] Y. J. Liang, N. Farber, and B. Girod, "Adaptive playout scheduling using time-scale modification in packet voice communications," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*. IEEE, 2001, vol. 3, pp. 1445–1448.

[4] F. Liu and C-C J. Kuo, "Quality enhancement of packet audio with time-scale modification," in *ITCom 2002: The Convergence of Information Technologies and Communications*. International Society for Optics and Photonics, 2002, pp. 163–173.

[5] P. Pocta, H. Melvin, and A. Hines, "An analysis of the impact of playout delay adjustments introduced by VoIP jitter buffers on speech quality," *Under review with Speech Communication*, 2013.

[6] W. Montgomery, "Techniques for packet voice synchronization," *Selected Areas in Communications, IEEE Journal on*, vol. 1, no. 6, pp. 1022–1028, 1983.

[7] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 749–752.

[8] J. G. Beerends, A. P. Hekstra, A. W. Rix, and M. P. Hollier, "Perceptual evaluation of speech quality (pesq): The new itu standard for end-to-end speech quality assessment part ii-psychoacoustic model," *Journal of the Audio Engineering Society*, vol. 50, no. 10, pp. 765–778, 2002.

[9] A. Hines, J. Skoglund, A. Kokaram, and N. Harte, "VISQOL: The Virtual Speech Quality Objective Listener," in *IWAENC*, 2012.

[10] A. Hines, J. Skoglund, A. Kokaram, and N. Harte, "Robustness of speech quality metrics to background noise and network degradations: Comparing ViSQOL, PESQ and POLQA," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013.

[11] A. Hines and N. Harte, "Speech intelligibility prediction using a neurogram similarity index measure," *Speech Commun.*, vol. 54, no. 2, pp. 306 – 320, 2012.

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

[13] H. Melvin, P. O. Flaithearta, J. Shannon, and L. B. Yuste, "Time Synchronisation at Application Level: Potential Benefits, Challenges and Solutions," in *International Telecom Synchronisation Forum (ITSF) 2009, Rome (Italy)*, 2009.

[14] P. O. Flaithearta and H. Melvin, "E-model based prioritization of multiple voip sessions over 802.11e," in *Digital Technologies 2010 conference, Žilina (Slovakia)*, 2010.

[15] O. Stapleton, "Quantifying the effectiveness of PESQ (Perceptual Evaluation of Speech Quality), in coping with frequent time shifting," *Master Thesis, NUI Galway/Regis University Colorado*, 2010.

[16] ITU, "ITU-T coded-speech database," Int. Telecomm. Union, Geneva, Switzerland, ITU-T Rec. P.23, 2003.

[17] ITU, "Mapping function for transforming P.862 raw result scores to MOS-LQO," Int. Telecomm. Union, Geneva, Switzerland, ITU-T Rec. P.862.1, 2003.

[18] ITU, "ITU-T methods for subjective determination of transmission quality," Int. Telecomm. Union, Geneva, Switzerland, ITU-T Rec. P.800, 1996.