Formal Subjective Test Results of VP8 compared to AVC Constrained Baseline (WebVC)

Introduction

- Results from a formal video subjective assessment test are presented.
- The test was conducted in March 2013 as part of a study aimed at providing MPEG with more details about how VP8 compares to IVC.
- The test was conducted by Vittorio Baroncini (MPEG's test chair), as requested by MPEG.
- VP8 proponents also requested that the test include AVC constrained baseline. The results for this part of the test were not presented to MPEG.
- This presentation provides the results from that test.

Test Details

- There were 4 classes in the test, however AVC constrained baseline streams were only available for three of those classes (these streams were generated for MPEG's Internet Video Coding Technology Call for Proposals by members of Fraunhofer FHG).
- This presentation provides results for:
  - Class A (1920x1080): 2 sequences.
  - Class B (832x480 – WVGA): 4 sequences.
  - Class C (416x240 – WQVGA): 4 sequences.
  - All material was progressive.
  - All sequences were part of the first IVC Call for Proposals.
- VP8 encoder version 0.9.6 was used.
- AVC JM 16 was used.

Test Methodology

- Conducted using a test protocol derived from the ITU-R DSIS (Double Stimulus Impairment Scale) test method. The scale used was from 0 to 10.
- The viewers were screened for visual acuity and color blindness. A total of 24 viewers were used in the test (20 males and 4 females).
- The viewers were trained.

Test Scale

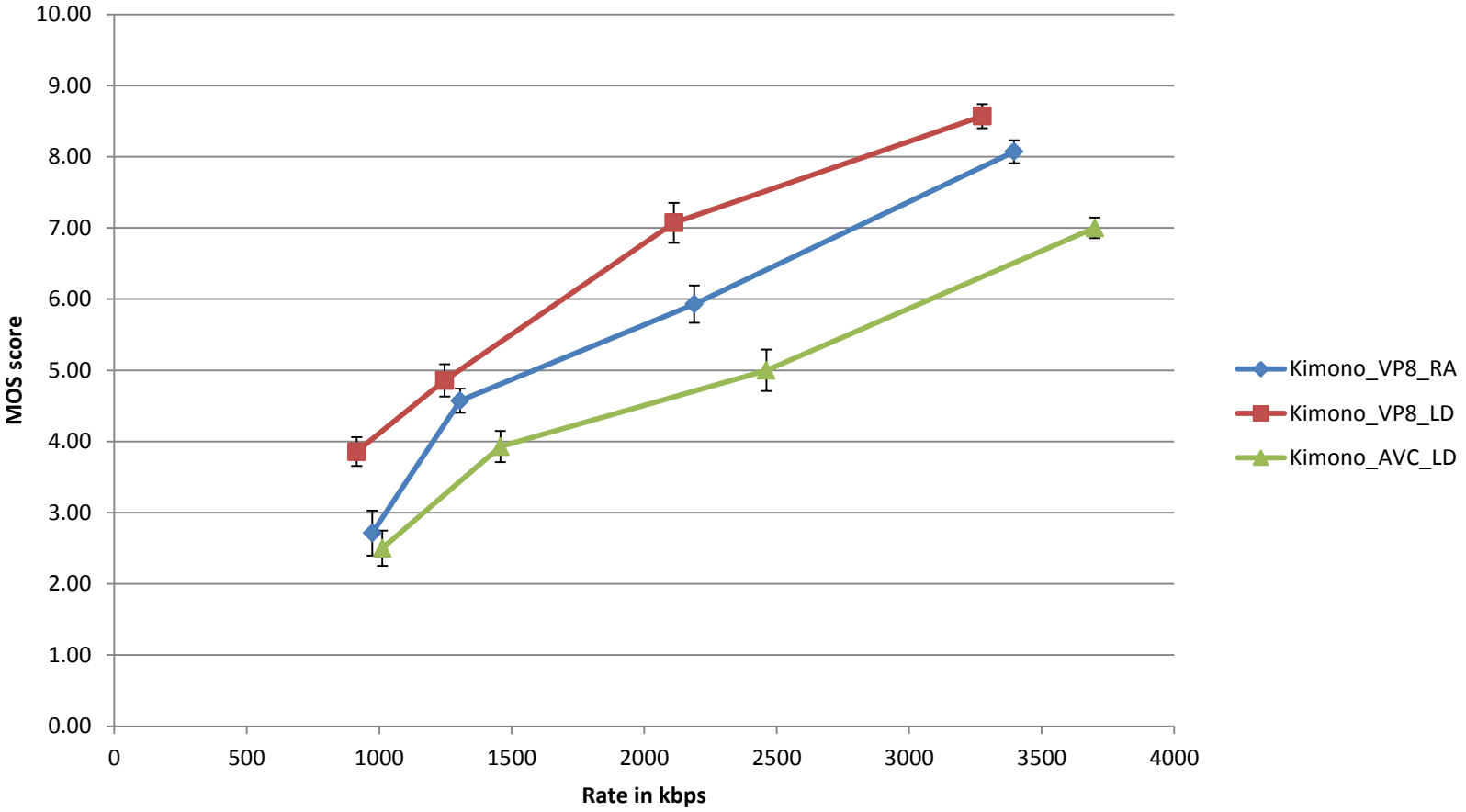
- 9 to 8: some degradation is seen (the subject was asked to use this even if they are not sure or after a careful inspection of the image).
- scores 7 and 6 :The subject is sure he/she has seen some degradation.
- Scores 5 and 4: Degradation is clearly visible
- Scores 3 and 2: Degradation are immediately evident;
- Scores 1 and 0: The video is widely and deeply corrupted.

Results

- In the results that follow LD means IPPP structure, RA means Random Access (an I frame every 1.1 seconds). There were only LD sequences available for AVC.
- The Label “MOS” is used for the test score.
- The bitrates for the test points do not align exactly for the three cases shown, however the trend of each curve is clear.

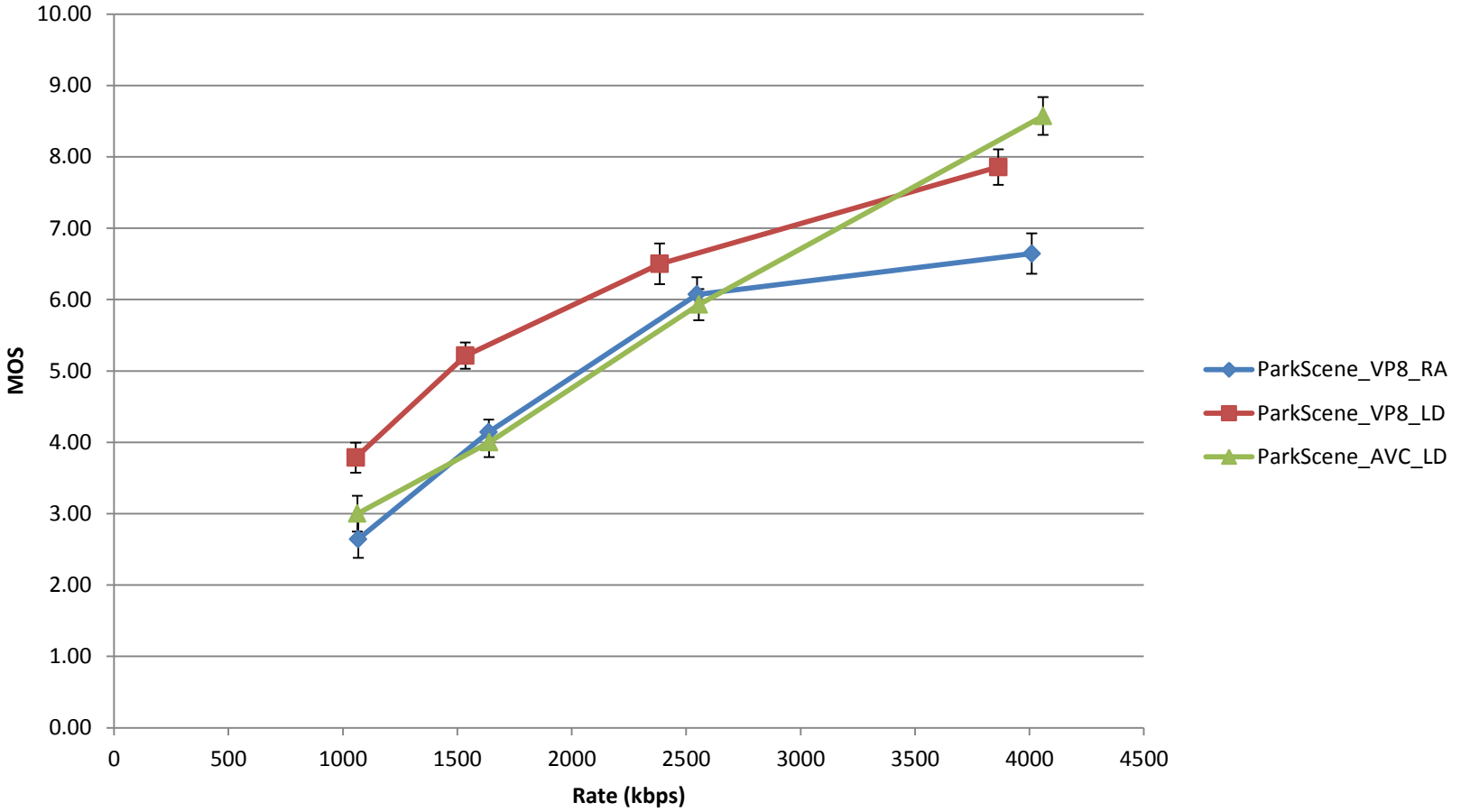
Class A (1920x1080)

- Kimono (24 fps)



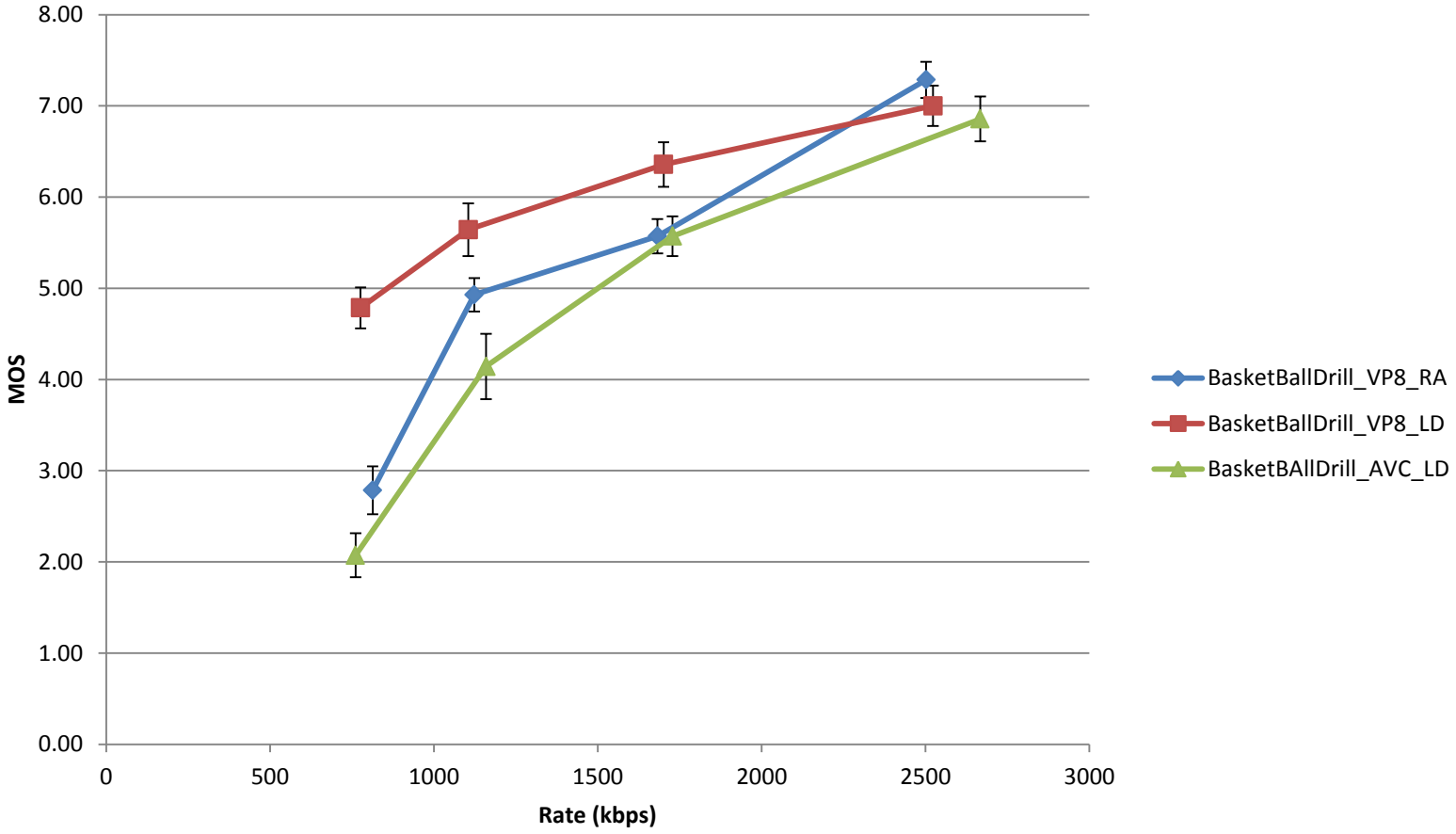
Class A (1920x1080)

- ParkScene (24 fps)



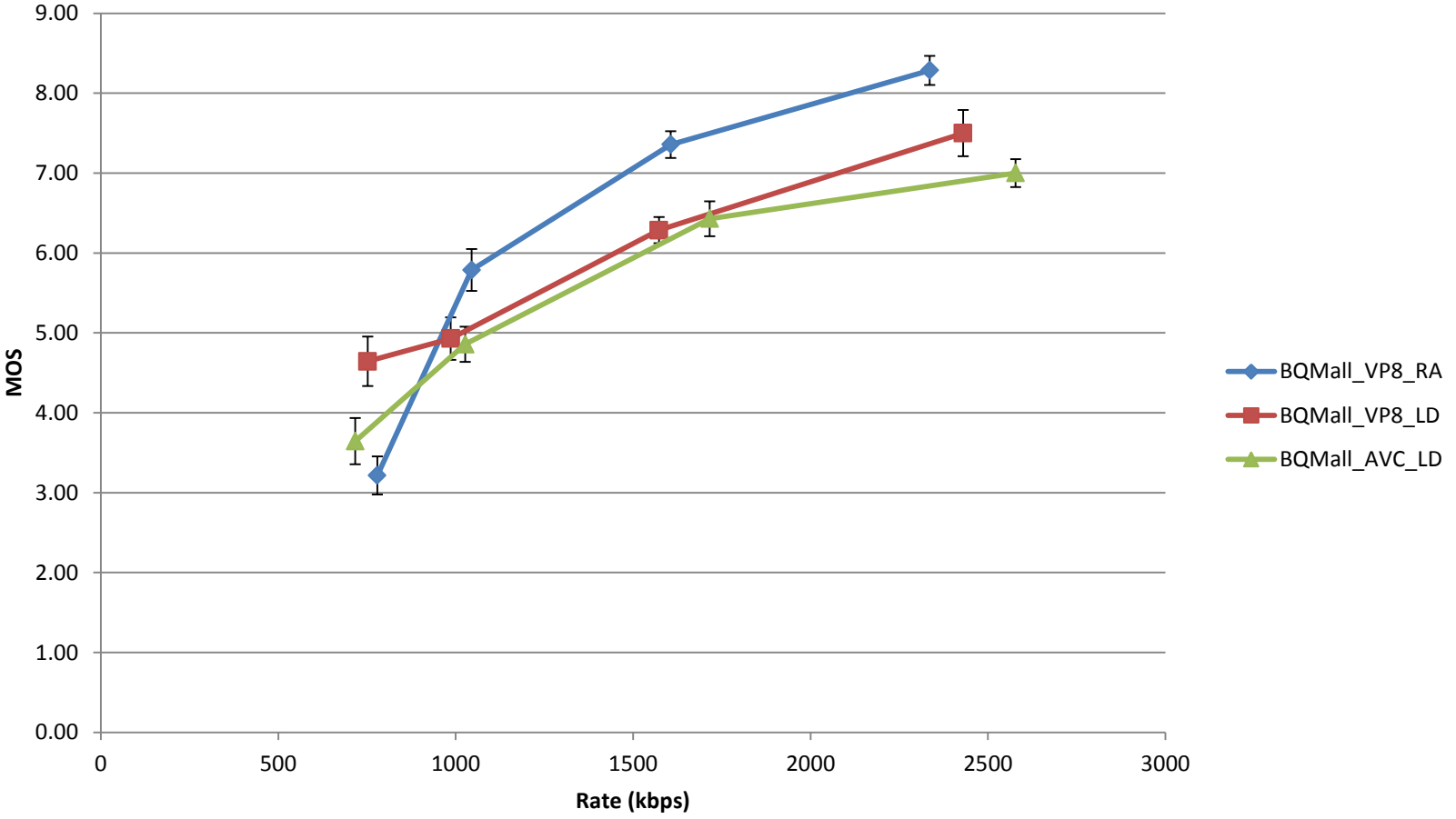
Class B (832x480)

- BasketBallDrill (50 fps)



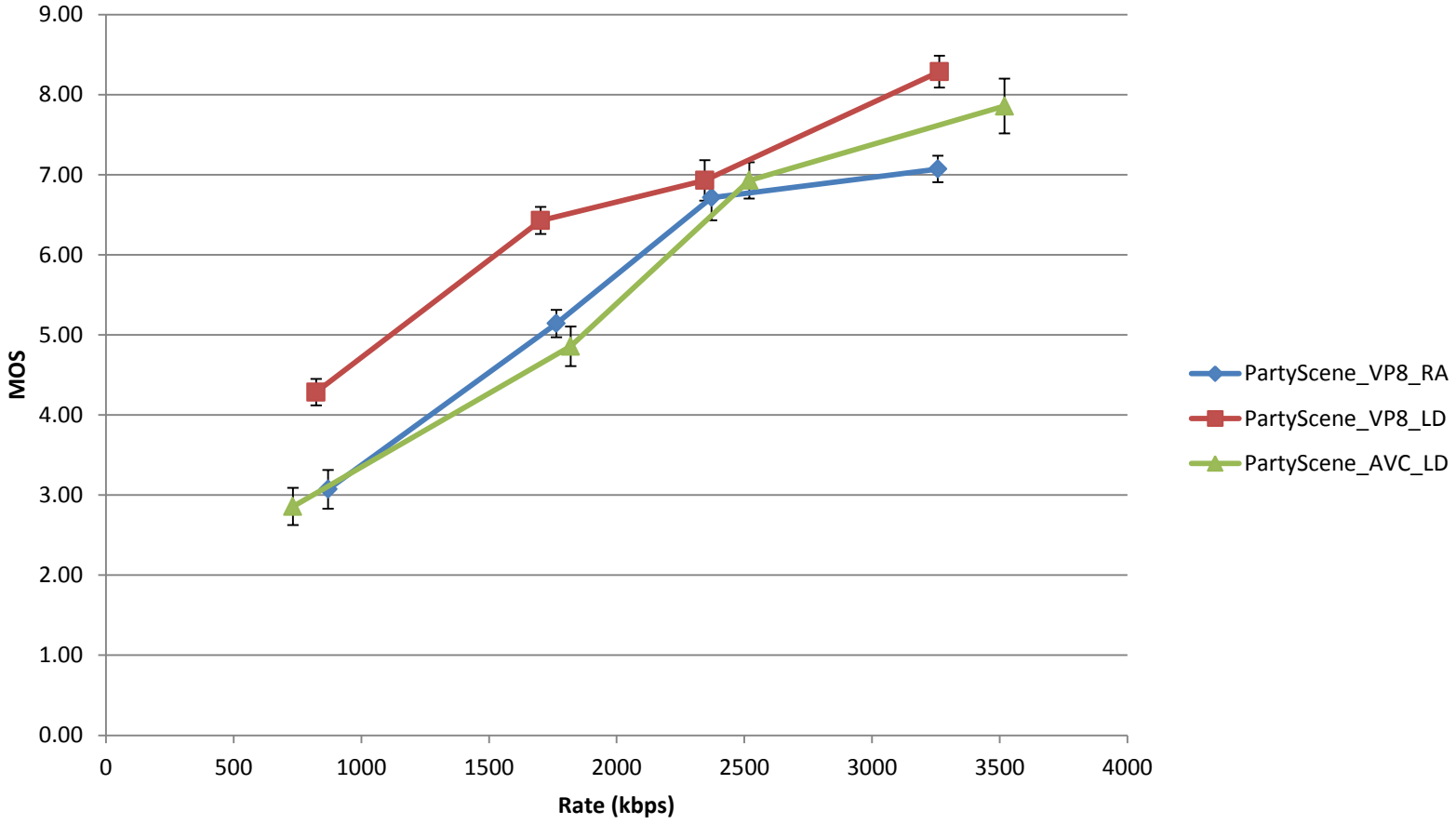
Class B (832x480)

- BQMall (60 fps)



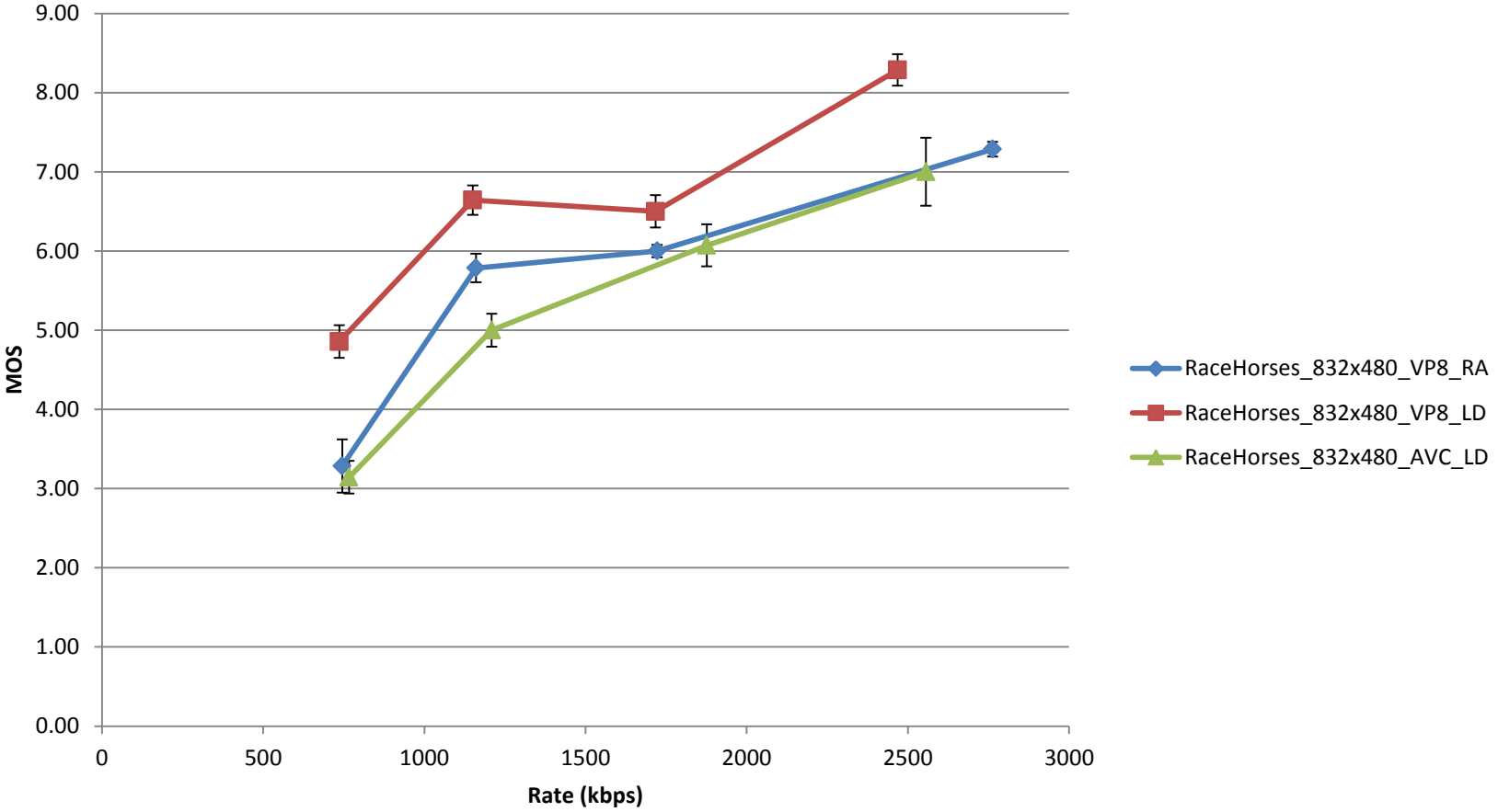
Class B (832x480)

- PartyScene (50 fps)



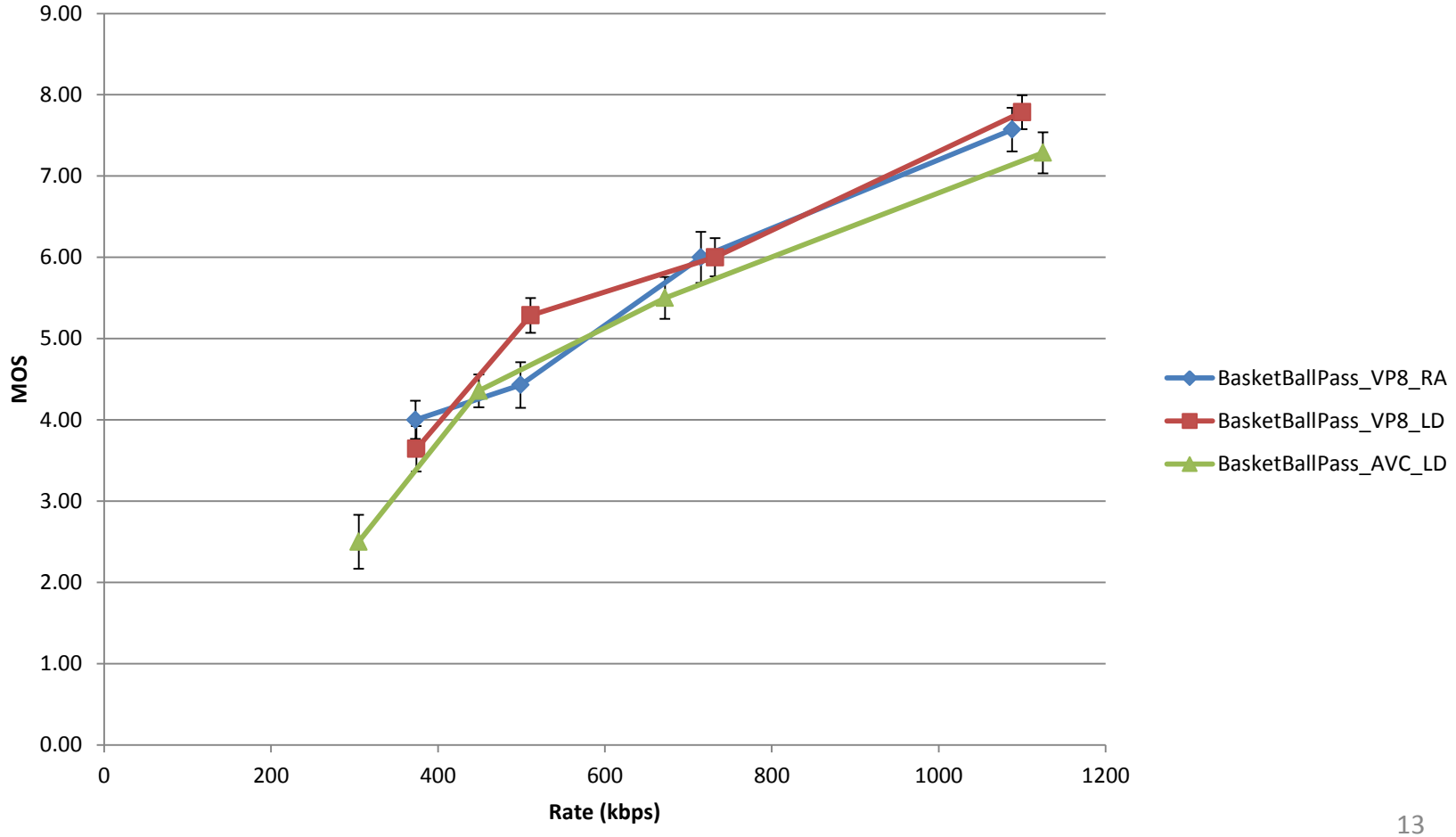
Class B (832x480)

- RaceHorses (30 fps)



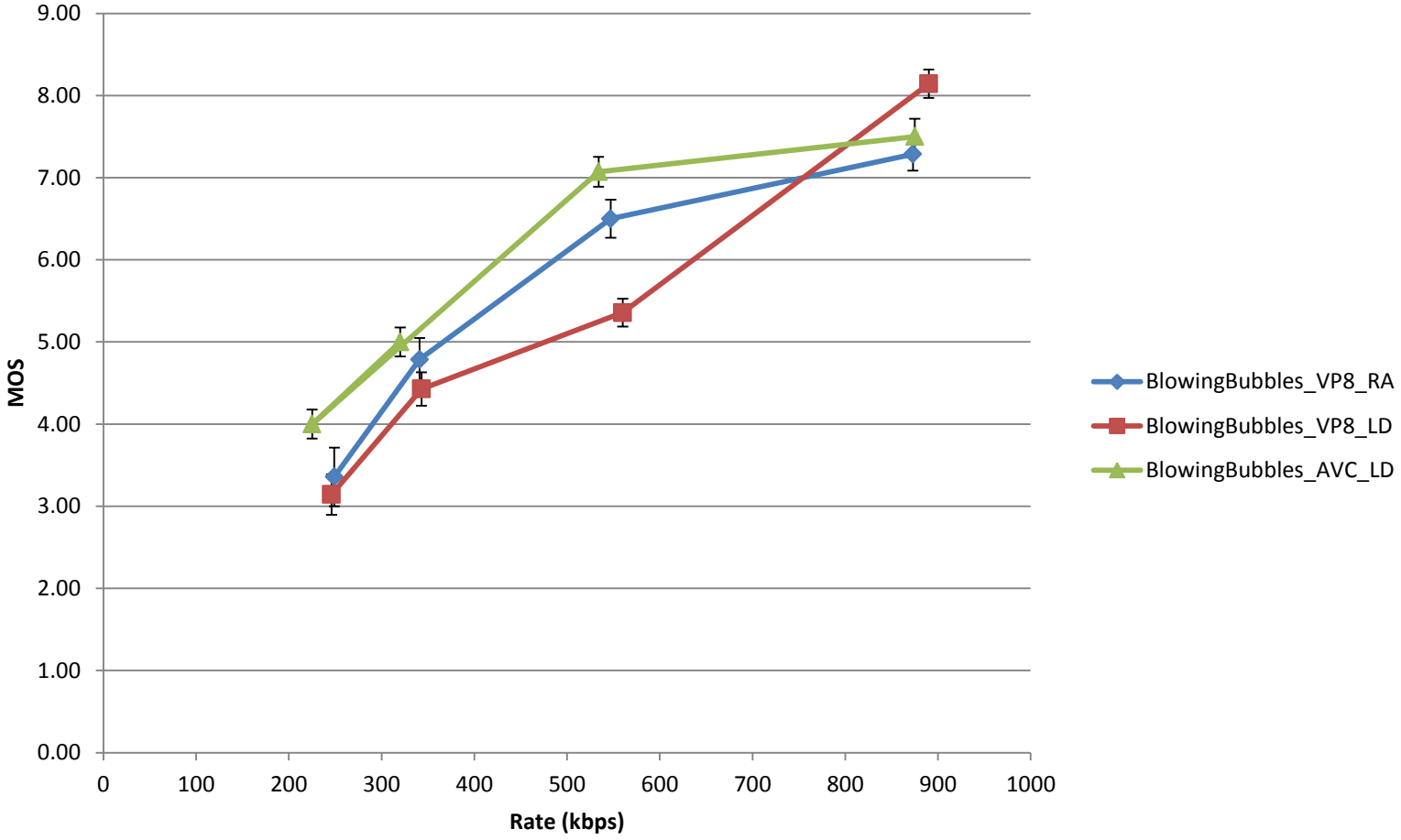
Class C (416x240)

- BasketBallPass (50 fps)



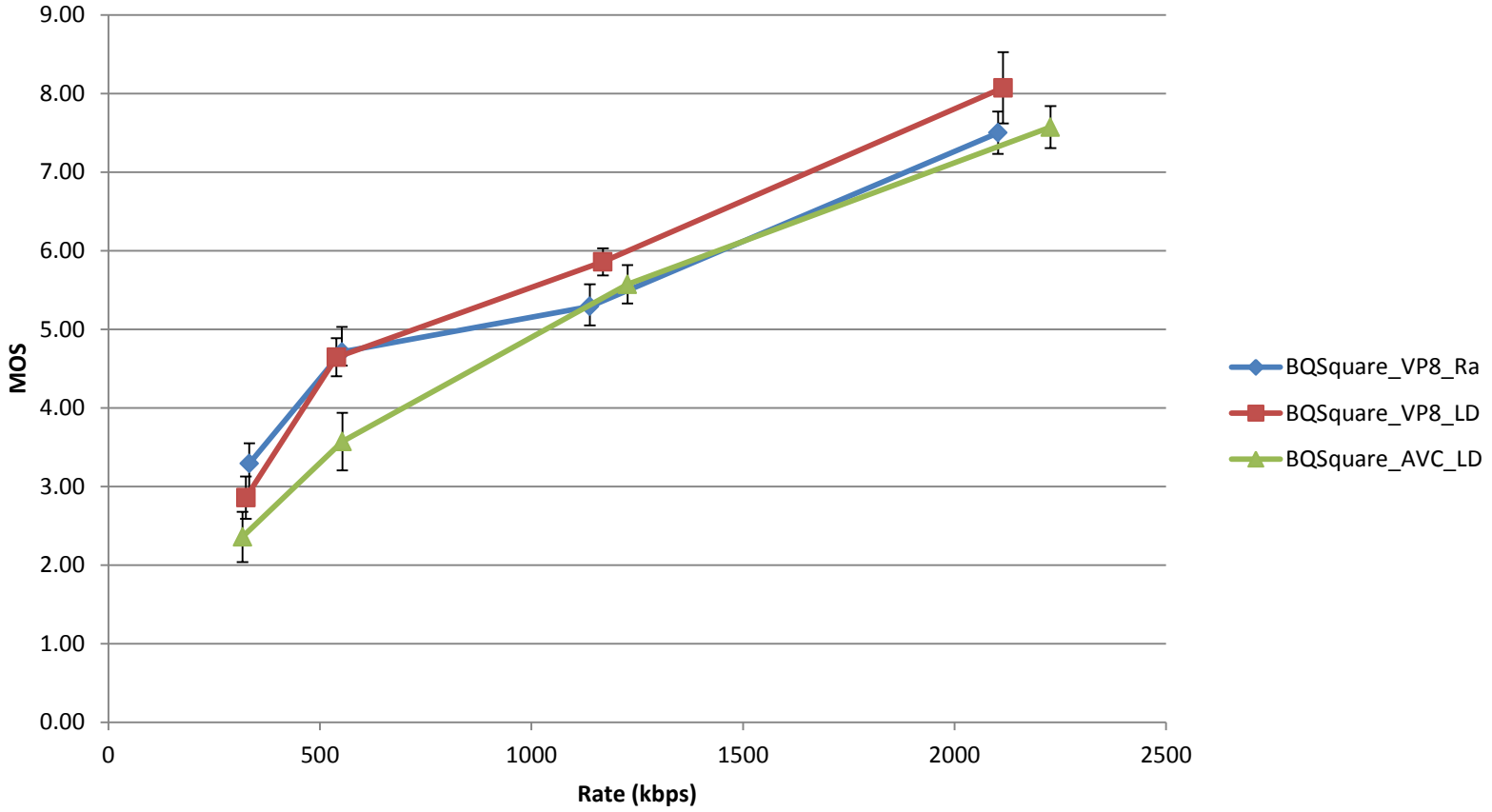
Class C (416x240)

- BlowingBubbles (50 fps)



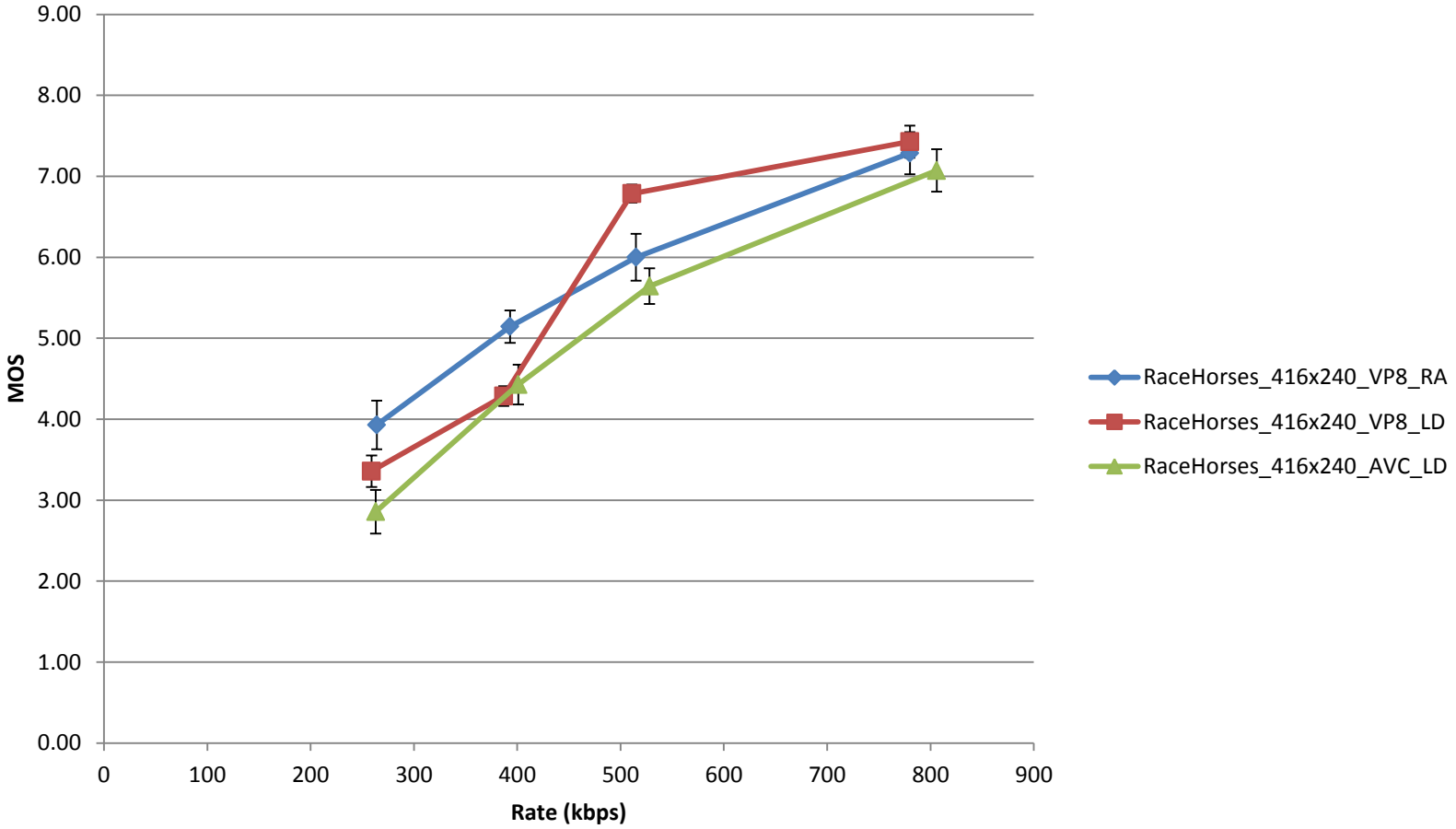
Class C (416x240)

- BQSquare (60 fps)



Class C (416x240)

- RaceHorses (30 fps)



Conclusion

- In 7/10 sequences tested VP8 LD was clearly better than AVC constrained baseline.
- The difference was most visible in classes A and B.
- In the majority of cases VP8 scores more than 8 on the test scale for the highest rate. This only happens twice in the case of AVC constrained baseline.
- Overall, VP8 LD provided better quality than AVC constrained baseline.