YouTube Ads and Video

KPI and QoE Metrics

Test Environment

- Motorola E2 in India (Bangalore, Outer Ring Road, 1pm IST)

1. Lenovo-Motorola has 9% market share in India
2. Android 5.0.1
3. Vodafone 4G (Vodafone has 18% market share in India)

- Nexus 6 in India (Bangalore, Outer Ring Road, 1pm IST)

1. Android 7.1.1
2. Airtel 3G (Airtel has 25% market share in India)

- Samsung S7 in United States (Mountain View Downtown, 1pm PST)

1. S7 has a 10% market share in USA
2. AT&T LTE (AT&T has 33% market share in USA)
3. Android 7.0

- Number of Iterations - 10

Test Case

- Load a YouTube Video that has an embedded AD
- Measure the time to load the AD
- Measure the time to load the video following the AD
- Measure the Video QoE Metrics for AD
- Measure the Video QoE Metrics for Video
- Measure the Audio QoE Metrics for AD
- Measure the Audio QoE Metrics for Video

Test Results

- All Test results collected and stored in MySql
- All the results exported to Google Sheets
- Results Available at: https://goo.gl/fVWj96
- Test Samples in Google Drive: https://goo.gl/kbse5W

Video Quality of Experience Measurements

- Using PEVQ (Perceptual Evaluation for Video Quality) for end to end measurement
- For this proof of concept tests, the following video metrics were captured
  - Mean Opinion Score (MOS)
  - Differential MOS
  - Frame Freeze (%)
  - Jerkiness
  - Blur
  - Blockiness
  - Peak Noise to Signal Ration (PSNR)

Audio Quality of Experience Measurements

- Using POLQA (Perceptual Objective Listening Quality Analysis) for end to end measurement
- For this proof of concept tests, the following audio metrics were captured
  - MOS-LQ0
  - R-Factor
  - SNR (Signal to Noise Ratio)

Video Mean Opinion Score and Differential MOS

- MOS Score is on a scale of
  - 1 Unacceptable
  - 2 Poor
  - 3 Fair
  - 4 Good
  - 5 Excellent

- Differential MOS is the difference between reference and processed MOS. The score is on a scale of
  - 3.1 - 4.0 Most Users Dissatisfied
  - 2.1 - 3.0 Many Users Dissatisfied
  - 1.1 - 2.0 Some Users Satisfied
  - 0.7 - 1.0 Most Users Satisfied
  - 0.0 - 0.6 Very Satisfied

Audio Mean Opinion Score and R-Factor

- MOS-LQ0 Score is on a scale of
  - 0– Worst
  - 4.5 - Best

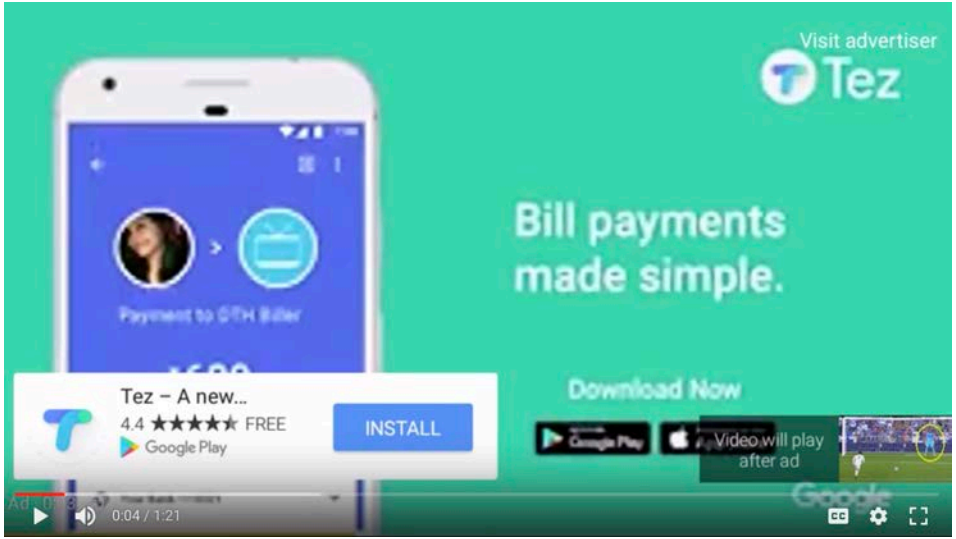
- R-Factor is on the following scale
  - < 60 Unacceptable
  - 0 – Bad
  - 100 - Best

Video Mean Opinion Score (MOS), Differential MOS, Frame Freeze and Jerkiness (India)

Test Device/Carrier	MOS	DMOS	Frame Freeze (%)	Jerkiness
Moto E2 Vodafone 4G (Video)	2.627	2.373	0	1.522
Moto E2 Vodafone 4G (Ads)	3.122	1.614	96	3.968
Moto E2 Vodafone 3G (Video)	2.799	2.201	32	3.102
Moto E2 Vodafone 3G (Ads)	2.893	2.107	12.8	0.428
Nexus 6 Airtel 3G (Video)	2.723	2.277	124	3.419
Nexus 6 Airtel 3G (Ads)	2.958	2.042	78.4	1.007

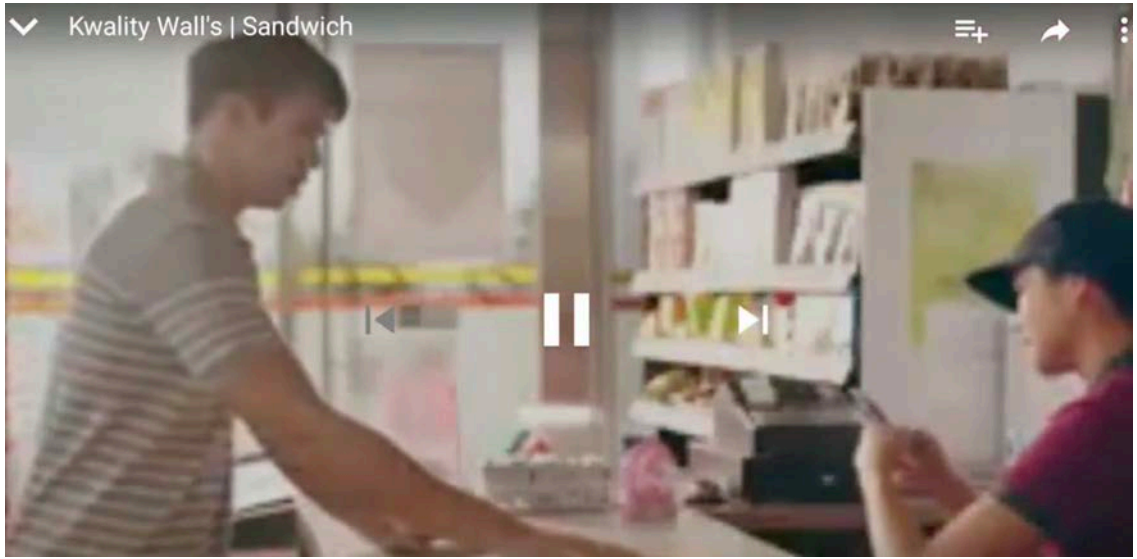
Poor experience highlighted in RED

Video Example (India)



Blurry Video Quality

Ad Example (India)



Blurry Video

Ad Example (India)

Grainy Video with Jerkiness

Video Mean Opinion Score (MOS), Differential MOS, Frame Freeze and Jerkiness (USA)

Test Device/Carrier	MOS	DMOS	Frame Freeze (%)	Jerkiness
S7 AT&T 4G (Video)	3.752	1.248	124	2.65
S7 AT&T 4G (Ads)	3.799	1.201	0	1.115
S7 AT&T HSPAP (Video)	3.386	1.614	96	3.968
S7 AT&T HSPAP (Ads)	3.631	1.369	152.35	1.899

Video Example (USA)

Grainy Video

Video Example (USA)

Grainy and Blurry Video

What is Blockiness, Blur and PSNR?

- Blockiness is often the result of a low bit rate coding that uses a block matching algorithm for the motion estimation and a coarse quantization for the image blocks.

- Blur is a distortion characterized by reduced sharpness of contour edges and spatial detail.

- Peak Signal to Noise Ratio (PSNR): To allow for a coarse analysis of distortions in different domains the PSNR is provided for the Y, Cb and Cr components separately.

- Source - http://www.opticom.de/technology/pevq.php

Audio QoE Scores (India)

Test Device/Carrier	MOS-LQ0	R-Factor
Moto E2 Vodafone 4G (Video)	2.9778	58
Moto E2 Vodafone 4G (Ads)	1.0064	7
Moto E2 Vodafone 3G (Video)	1.0161	8
Moto E2 Vodafone 3G (Ads)	1	0
Nexus 6 Airtel 3G (Video)	2.32	45
Nexus 6 Airtel 3G (Ads)	1	0

Poor experience highlighted in RED

Audio QoE Scores (USA)

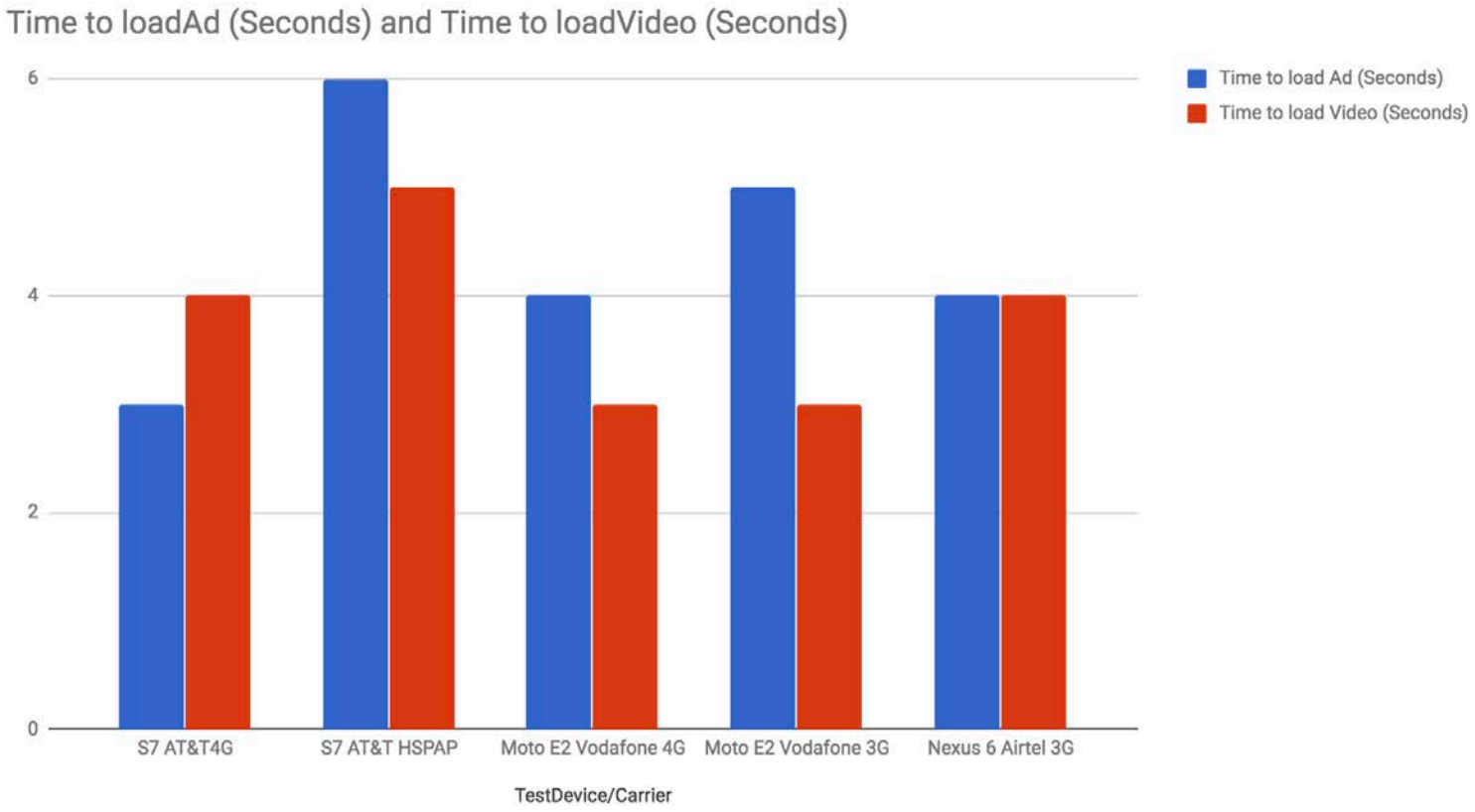
Test Device/Carrier	MOS-LQ0	R-Factor
S7 AT&T 4G (Video)	1.12255	15
S7 AT&T 4G (Ads)	4.2314	86
S7 AT&T HSPAP (Video)	1.0744	13
S7 AT&T HSPAP (Ads)	4.5	100

Poor experience highlighted in RED

YouTube Ads and Video KPI

Test Device/Carrier	Time to load Ad (Seconds)	Time to load Video (Seconds)
S7 AT&T 4G	3	4
S7 AT&T HSPAP	6	5
Moto E2 Vodafone 4G	4	3
Moto E2 Vodafone 3G	5	3
Nexus 6 Airtel 3G	4	4

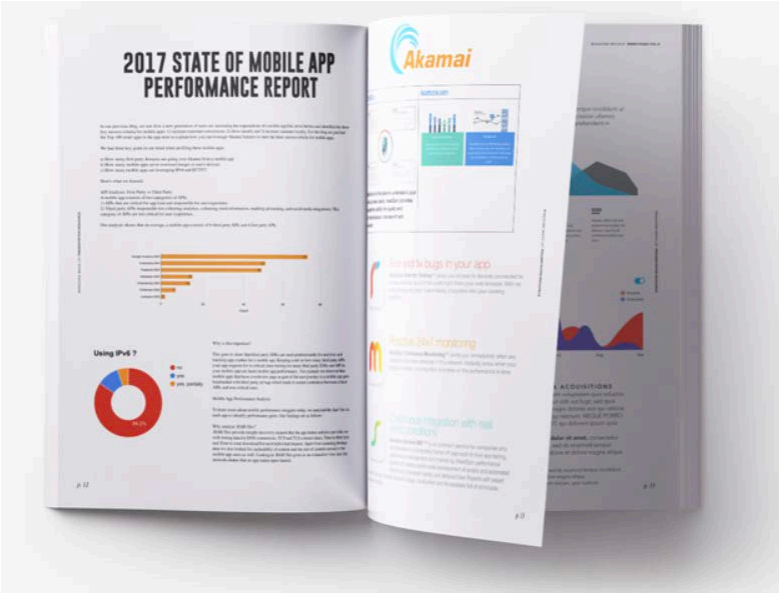
YouTube Ads and Video KPI



Future Work

- Test and measure QoE KPI/Metrics across multiple iOS and Android SKUs in different cities
  - Devices, OS and Carriers coverage based on market share %
- Test every build, code change and release
- Identify problems before launching app
- Supercharge Development and troubleshooting
  - Use USBoverIP software bridge to import devices into XCODE and Android Studio
- Automatically get issues from HeadSpin's AI Engine
  - Collect data to identify the problems with Image management, slow backend and third party
- Run the tests continuously in multiple locations on Stable hardware
  - HeadSpin has POPs in 150+ major worldwide cities
- Proactive alerts when the production app QoE issues
- Find bugs early, launch high quality version of the app!
- Compare with other apps like FB Video, Netflix, Amazon Prime Video, Disney etc.

2017 State of Mobile Performance Report



https://blogs.akamai.com/2017/04/the-state-of-mobile-app-performance.html