



Kobe, Japan & online
10–14 November 2025

Worker QoS reboot

Markus Handell
Google

Problem - why QoS of media resources

- External & internal workloads causing latency jitter, even on WebWorkers
- Actors
 - xPU Media workloads, for x in ['C', 'G', 'N', 'T']
 - GC
 - xPU Inference
 - ...
- Concrete issues/related
 - <https://g-issues.chromium.org/issues/456006123?pli=1&authuser=0>
 - <https://github.com/webmachinelearning/webnn/issues/898>
 - <https://github.com/webmachinelearning/webnn/issues/902>
- Ensure media workloads prioritized

Bad Consequences with poor QoS

- The usual suspects
 - robo-sound
 - audio glitches
 - dropped video frames
 - janky video
 - Increased E2E delay (JB impact)

History

- Intel presented proposal 2023 (work not active atm)
 - <https://www.w3.org/2023/Talks/TPAC/breakouts/web-worker-qos/>
 - <https://www.w3.org/2023/09/tpac-breakouts/47-minutes.pdf>
- Power-focused

Proposal

:= Intel's proposal

- “high”/“low”/“default” classes
- + “video-capture”/“display”/“record”/“playout”/“batch” (preference hint)
- + Abuse control
- + Influence other APIs work on behalf of worker (WebGPU, WebNN)

Abuse control

If it walks like a duck & quacks like a duck...

- Playout + monopolizing CPU -> 👊 (or OS-provided mechanism)
- Playout + woken by Audio context + sleeps +... (?) -> 👍

Innocent until proven guilty vs. Guilty until proven innocent.

Alternatives

- Heuristically classify workload automatically
 - Default-in, no opt out: could backfire

Feedback & next steps

- Feedback?
- Next step suggestions?