NAB 2025

OTT at YouTube Scale

Live Workflow - Sean McCarthy

Live Scale, Quality and Latency - Kirk Haller

Synthetic Experience Metrics - Chas Mastin

Stable Volume and Industry Consensus - Steven Robertson

Live Workflow - Sean McCarthy

OTT Purpose

Provide a personalized destination for transformative media experiences, connecting users with their passions on a global scale.

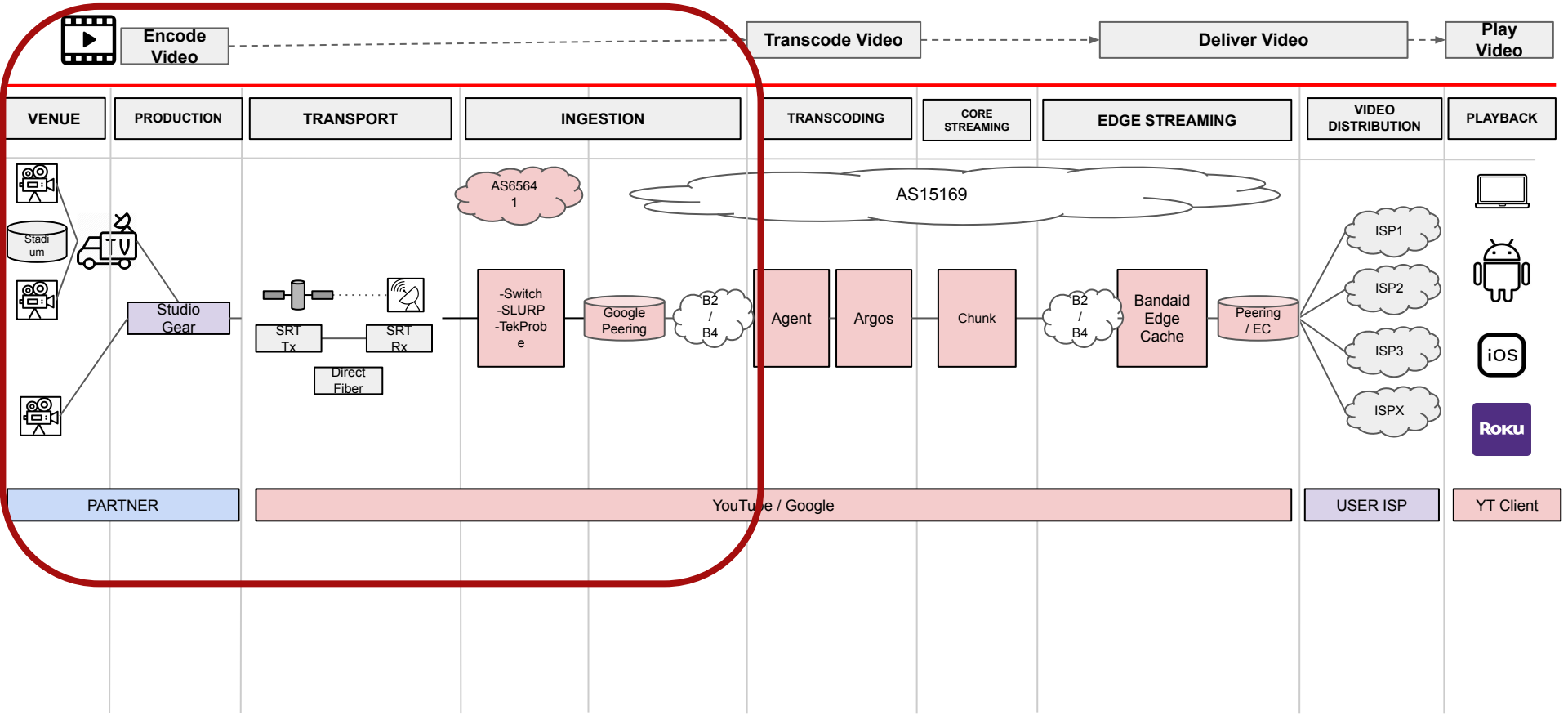
YouTube TV

Primetime Channels

Team Mission

Deliver the highest quality and most reliable premium live video streams to enable YouTube products.

OTT Specific Infrastructure



An OTT Timeline

First OTA antenna on HQ roof

"When YouTube TV started out, a group of engineers climbed onto the roof of YouTube headquarters while holding an antenna in order to build a prototype" - blog

Covered 3 DMAs and head-ends in CBF and AUS

First NOC established

24x7 signal ingestion

YouTube TV launches

5 DMAs - blog

Marquee monitoring starts. EOG expands to adjust schedules for EPG / DVR accuracy

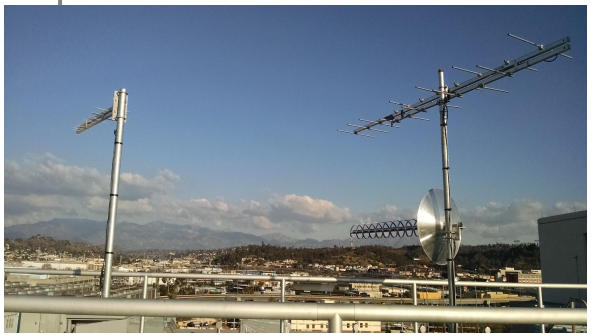
PSold DAI launches

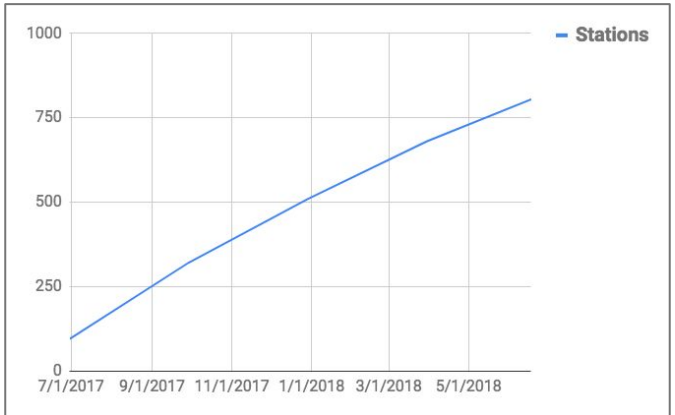
YouTube TV expands to 100 DMAs

- DAI launches on YTTV

- 750 Linear Channels offered

2015 | **2016** | **2017** | **2018**





Growing the Service

YouTube TV goes Nationwide - [blog](#) and 1 million subs!

[MLB GOTW](#) launches

Service Continues to Grow

PTC launches - [blog](#)

YTV crosses 5 million subs

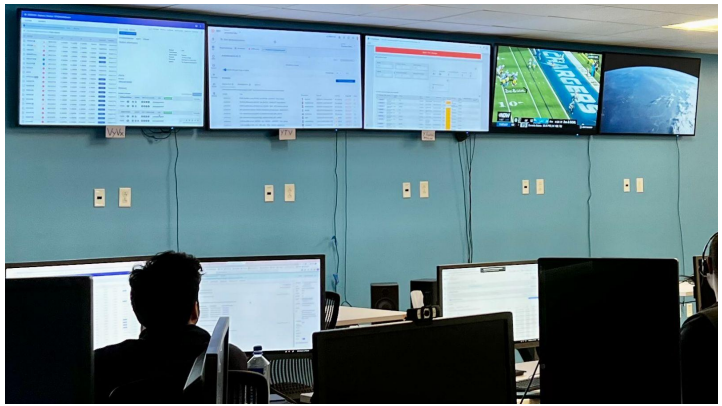
First SRT partner launches (BeIN Sports)

Paulistão launches - [blog](#)

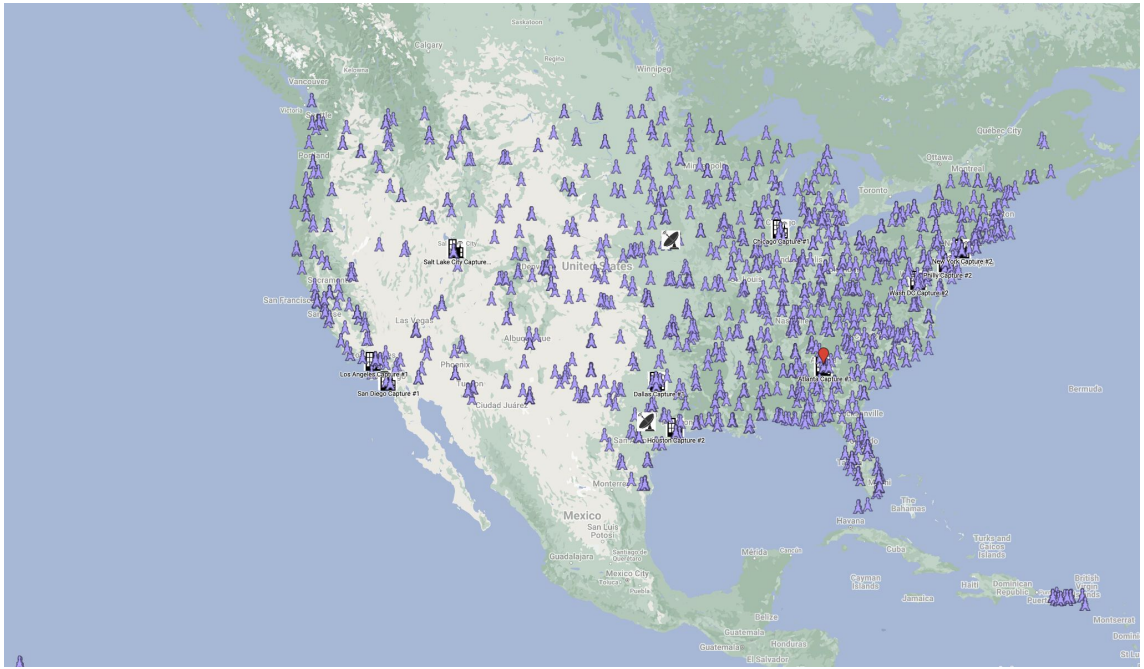
NFL Sunday Ticket launches - [blog](#)

Multiview launches

2019 | **2020 - 2021** | **2022** | **2023+**



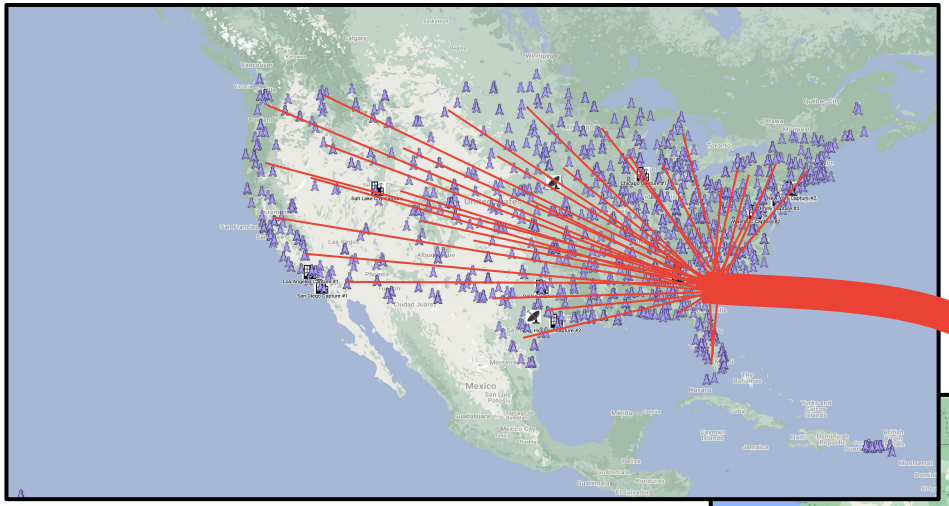
Legacy Infrastructure



Infrastructure Sites:

- **3** NHE's & **8** RHE's
- **12** IHE's (international)
- **101** OTA Sites
- **30+** Direct Fiber Installations
- **270+** Network Switches
- **215+** IRDs
- **3** Satellite Sites
- **200** Appliance Ecoders

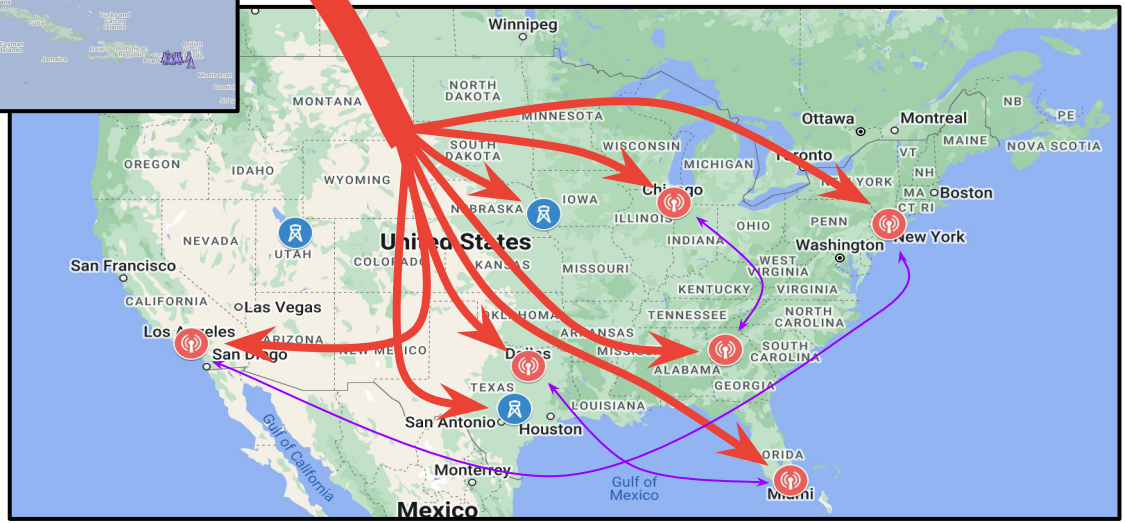
SHE Stop Gap to Reduce complexity



US Infrastructure Convergence

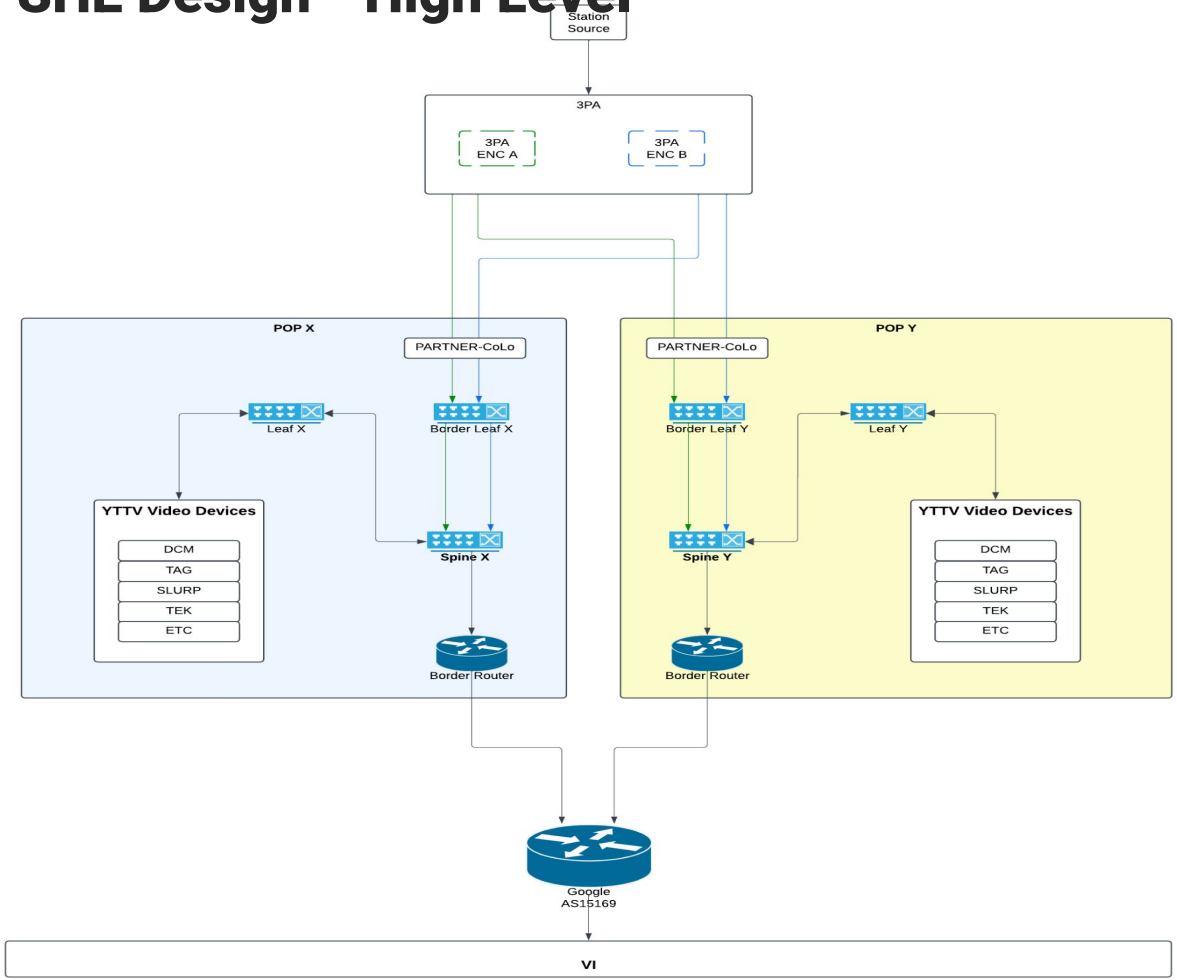
6 Super Headends

2 National Headends



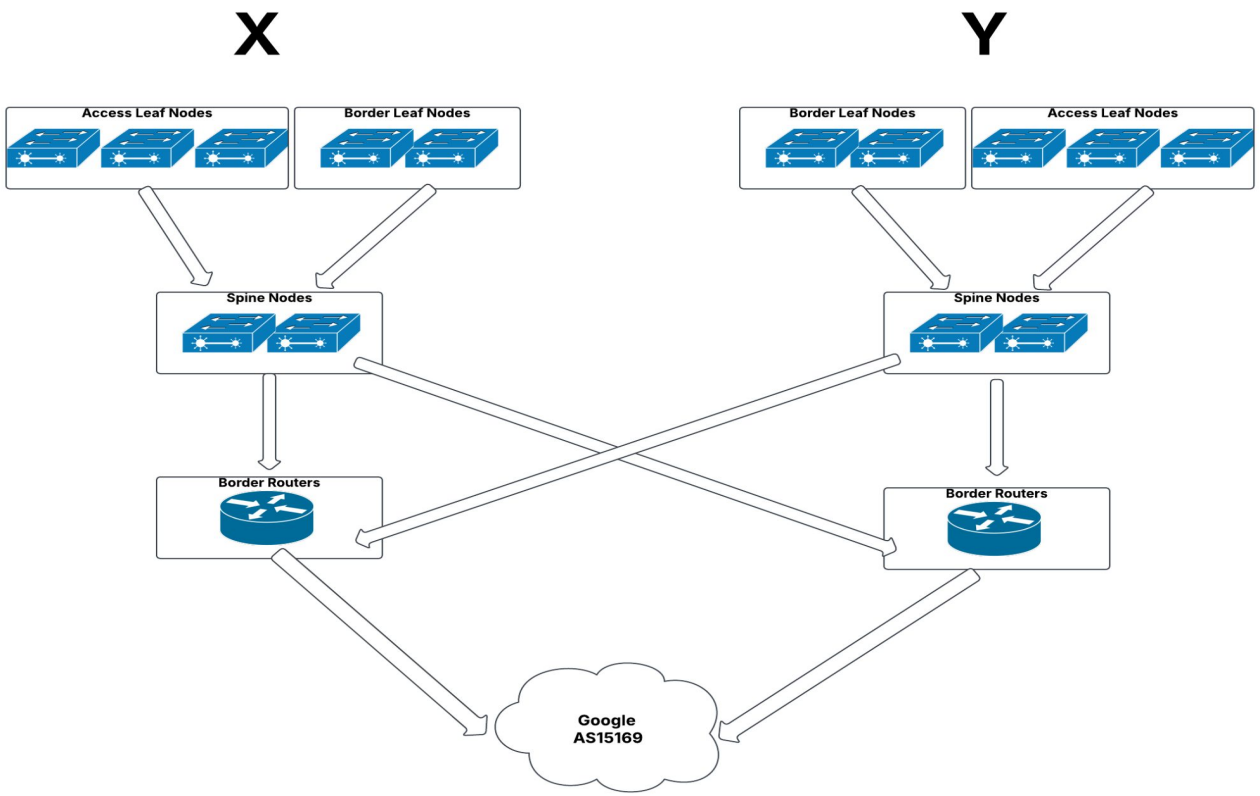
- Converge 300+ sites to 8
- Improved Q&R
- Significantly less maintenance
- Lower infrastructure management costs

SHE Design - High Level



- 6 Geo Diverse PoPs
- 2 PoPs (one cluster) per stream
- 2 Zones per PoP
- 4 redundant copies per stream
- High capacity throughput
- Fiber circuits from media co's and 3PAs

Inside a PoP - Spine-Leaf Architecture



- Redundant X,Y Paths
- Border Leafs for external connections (3PAs, Cross connects)
- Access Leafs are for internal devices (monitoring, normalization, etc)
- Spine Nodes connect all leaf nodes as well as Border Routers
- Border Routers are direct connected with Google Prod network

SHE Assessment

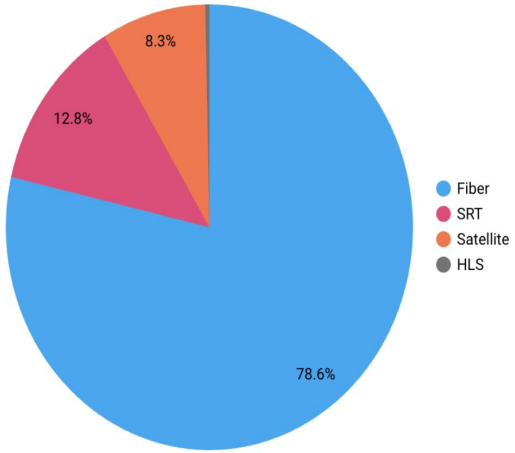
The good

- Maintainable and supportable, heterogeneous
- Lots of headroom in terms of capacity, bandwidth and server expansion capabilities
- Extremely redundant and able to support high quality streams
- Fully capable of video normalization/processing via 3rd party devices

The not-so-good

- Expensive and timely build outs (18-24 months design to install)
- Non elastic/finite scaling, difficult manual capacity planning
- Fiber ingest only
- Difficult to monitor
- Strong network engineering expertise required for design, implementation and maintenance

YTTV Acquisition Landscape



	acquisition_type	Percent Total
1.	Fiber	78.63%
2.	SRT	12.77%
3.	Satellite	8.28%
4.	HLS	0.33%

Format	Pros	Cons
MPEGTS	• Most widely used in broadcast • Typically high throughput and low latency • Easy for multi-partner distribution	• Physical vulnerabilities (fiber cuts) • Not supported in cloud environments (can explore direct connect w/ multicast support) • Expensive, fixed bandwidth • Often multicast only
SRT	• Internet native • Resilient (FEC) • Open Source • Strong ecosystem support/tooling • Strong adoption for live events, growing support for linear • Low latency	• More difficult server to implement than HTTP • Multiple modes/implementations • More expensive & difficult to scale • No video encryption • Although internet native, better with dedicated direct connects
CMAF over CDN for Syndication	• Passthrough opportunity for distributors (no transcoding required) • New Codecs already supported (AV1) • Full DRM support • Extremely cost effective and simple to implement • Can converge D2C and partner distribution processing • Better geo targeting/content replacement control	• CMAF packaging not ideal for re-transcoding • More latency
Media over Quic (MoQ)	• Low latency, target latency • Head-of-line blocking, better congestion control & tput • Multi-format support • Secure, Scaleable • Able to advertise media metadata	• New emerging standard, low adoption

The Metadata Challenge

gracenote. A NIELSEN COMPANY → How TV Metadata is shared today

- Can be Expensive
- Main API based on a linear data model , not great for discrete live events
- New services and providers aren't in the ecosystem
- Time to propagate data/changes

OR

Disney + Paramount+ + WB WARNER BROS. DISCOVERY

OR

SCTE 224

- Event-driven ESNI
- Flexible metadata
- Real-time signaling

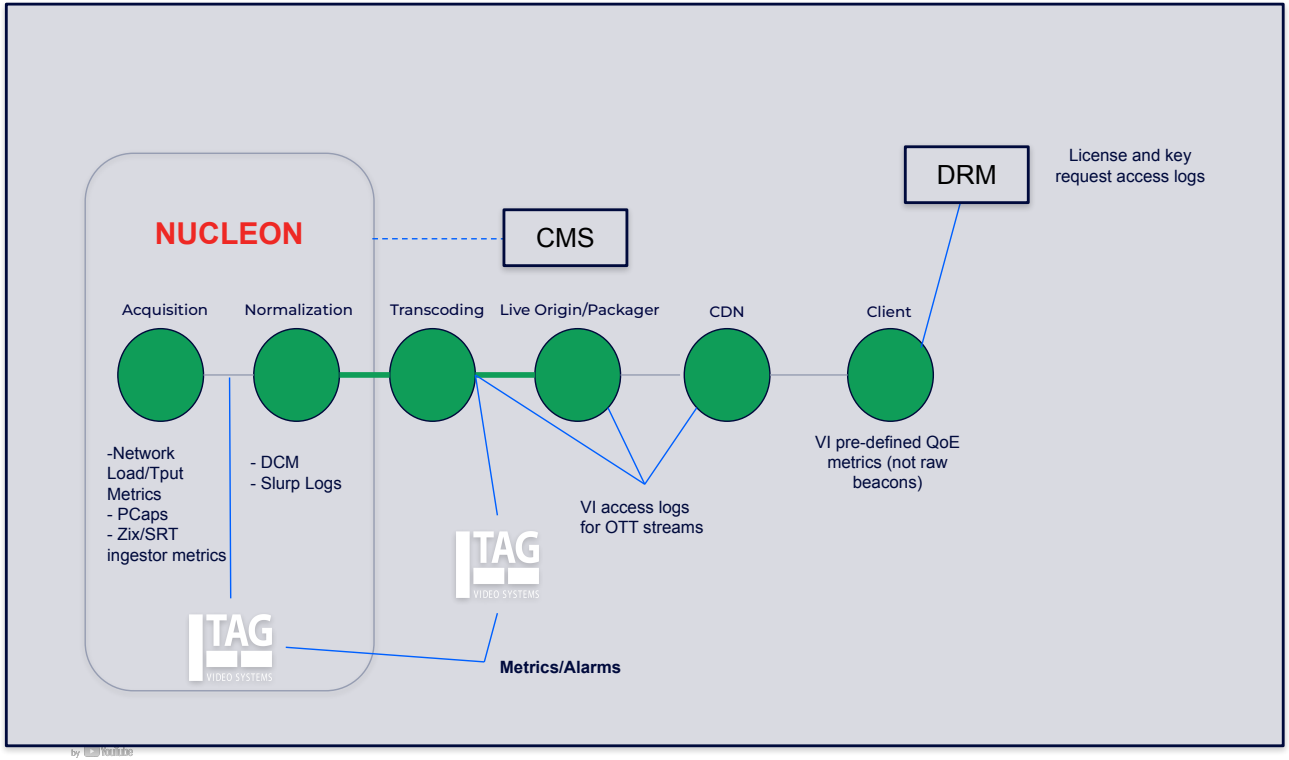
+

EIDR

- Registry for a Universal Stream Identifier
- Metadata reconciliation
- schema.org open spec

Observability Challenge

A Need for End-to-End Monitoring



Challenges

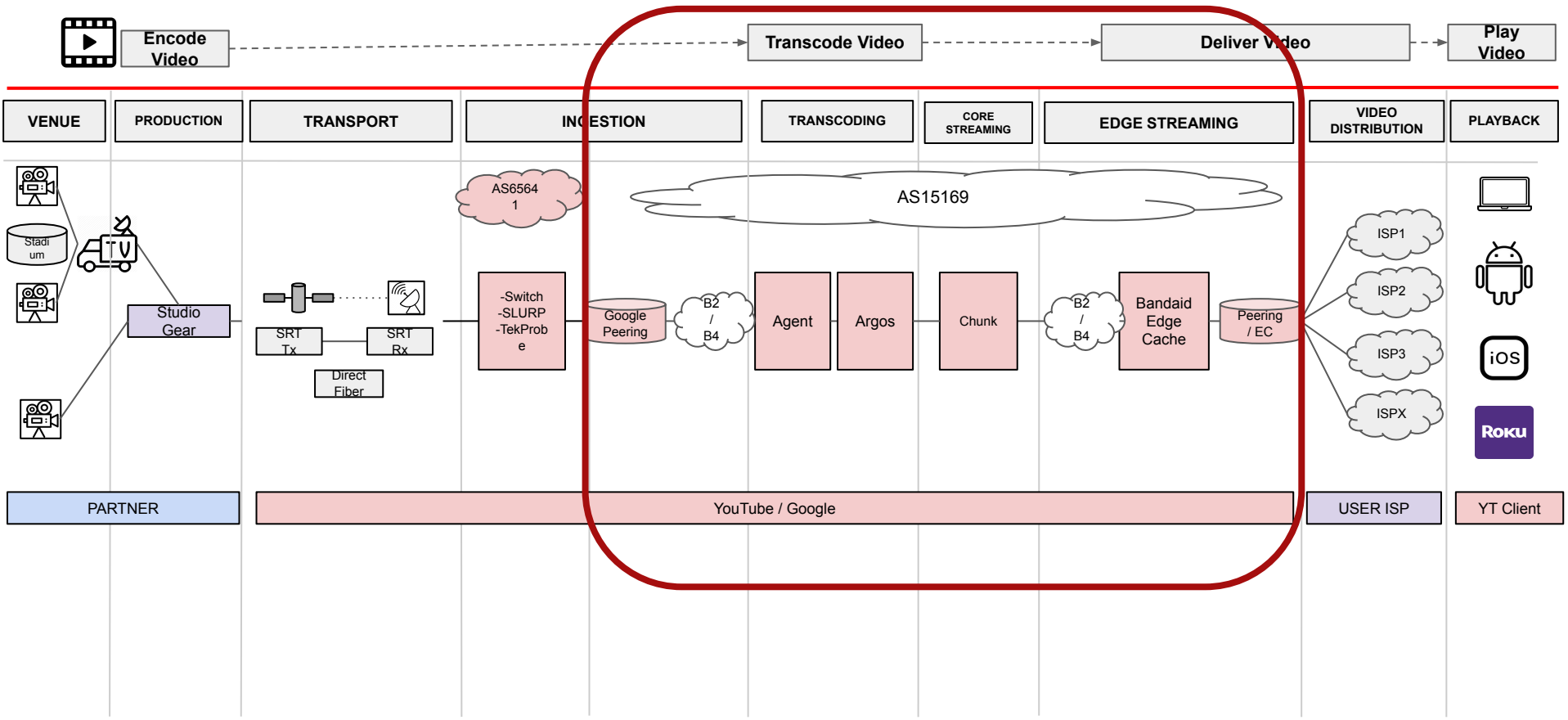
- Correlating ingestion events with QoE playback issues
- Monitoring the bitflow of a channel/stream – distributed request tracing
- Differentiating component, transport level and content level issues
- Realtime Raw log collection

Goals

- Realtime – high volume, high dimensionality, high cardinality, data pipelines (sub 30 seconds)
- Fast querying
- Curated, well understood metrics for operations
- End to end observability

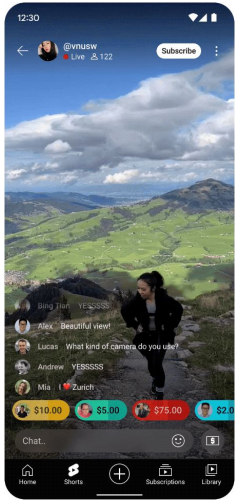
Quality Real Time Video at Scale - Kirk Haller

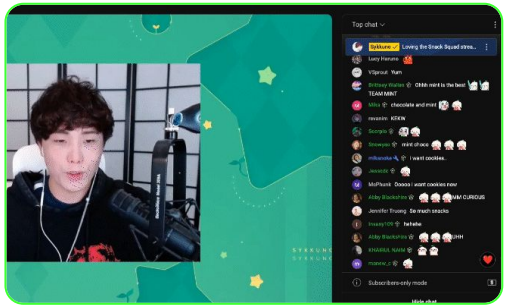
YT Live


Encode Video
Transcode Video
Deliver Video
Play Video
VENUE
PRODUCTION
TRANSPORT
INGESTION
TRANSCODING
CORE STREAMING
EDGE STREAMING
VIDEO DISTRIBUTION
PLAYBACK
Stadium
Studio Gear
SRT Tx
SRT Rx
Direct Fiber
AS65641
-Switch
-SLURP
-TekProbe
AS15169
Google Peering
B2 / B4
Agent
Argos
Chunk
B2 / B4
Bandaid Edge Cache
Peering / EC
ISP1
ISP2
ISP3
ISPX
iOS
Roku
PARTNER
YouTube / Google
USER ISP
YT Client

Diversity and Scale

YouTube Live: from casual mobile to professional 4K HDR





Streams range from Mobile to Gaming and Desktop to Broadcast.

Beyond YouTube TV, sports are a big international presence on YouTube.

YouTube Live Streams can be running for years

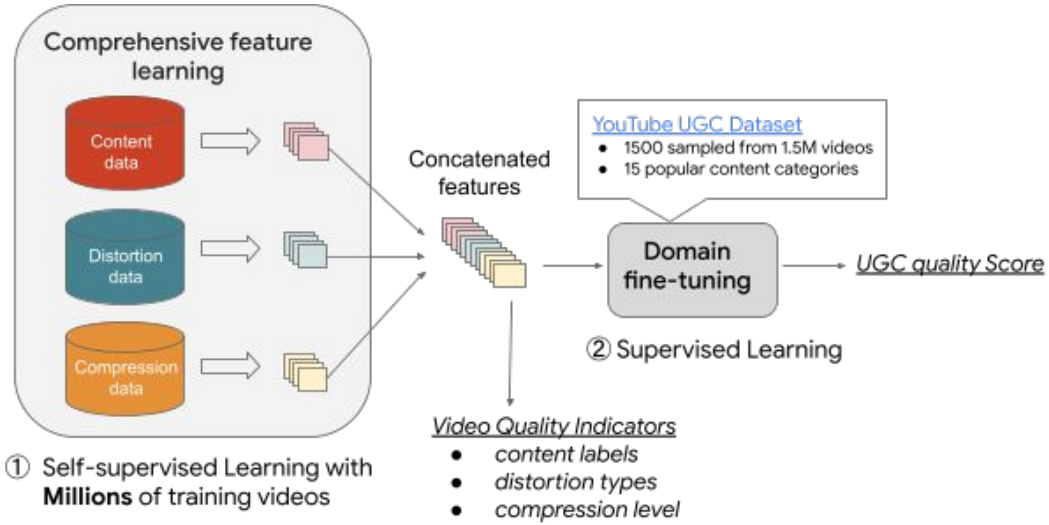
Quality

Video quality drives viewer engagement

In 2024, Living Room accounted for >50% of Coachella's total livestream watchtime, the highest of any year to date.

In 2025, we will also be using “Watch With”, allowing creators to commenting on the stream.

Measuring Audio/Video Quality



In 2022, YouTube and Google research opened source UVQ

- ML Model trained on Mean Opinion Scores (MOS)
- Based on content, distortions and compression data
- “No reference” metric

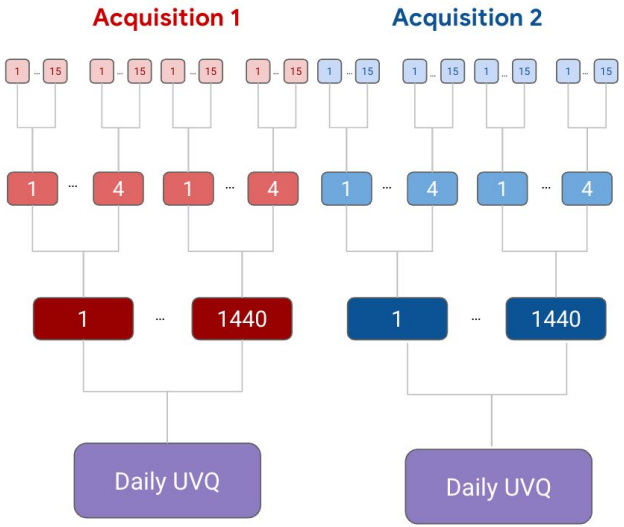
Measuring Audio/Video Quality

UVQ score	Perceptual quality
[1.0, 3.5)	relatively low
[3.5, 4.1)	fair
[4.1, 5,0]	relatively high
0.05~0.1 UVQ delta	Just noticeable difference (JND)

MOS score on TV	Satisfaction	Resolution
4.17	95%	1080
4.0	95%	720

Quality

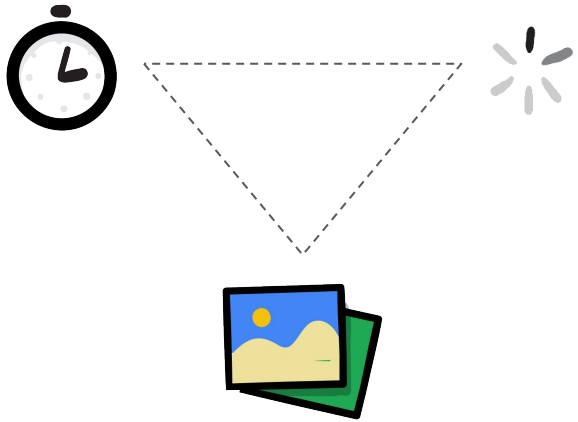
Measuring Bad Minutes



- Use Sampling w UVQ
- Aggregate by minute
- Count bad minutes

Latency

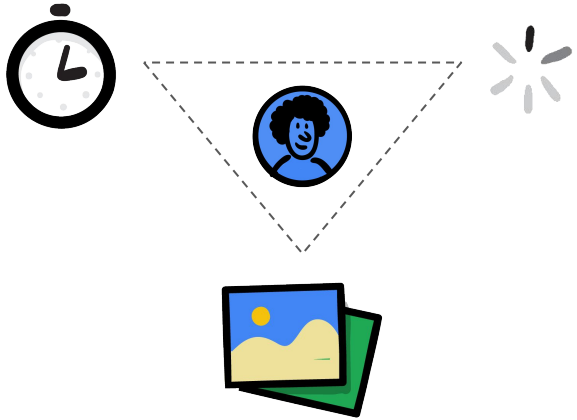
The tradeoff



- Stream Latency
- VIdeo Quality
- Playback Quality

Latency

Personalization



Adjust tradeoff to fit viewer

Signals:

- Content nature
- Viewer signal
  - Interaction
  - Preference
- Client bandwidth
- Network health
- Stream health

Synthetic Experience Metrics - Chas Mastin

Playback Experience (Px)

The Best Experience Per Bit

People > Product > Process

How to improve Experience for >2B Users?

The Naive View of QoE

Naive QoE Rank (Me, circa 2020)

1. Video Start Failures
2. Playback Failure
3. Rebuffer
4. Bitrate
5. Start Latency
6. Black Screens, A/V synch, long tail issues

The Naive View of QoE

QoE Rank (Me, circa 2025)

1. Start Latency (Survival Style)
2. Everything else…

The Naive View of QoE

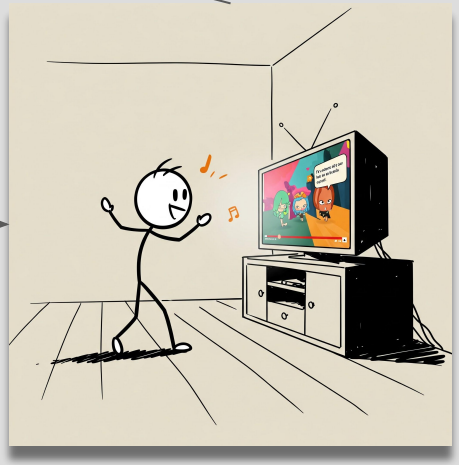
QoE Data

Streaming Systems, transcoders, CDN, etc

Internet, ISP

Device, OS, Client, Player, Media

Device Buffer



Source of Most Fixable Issues

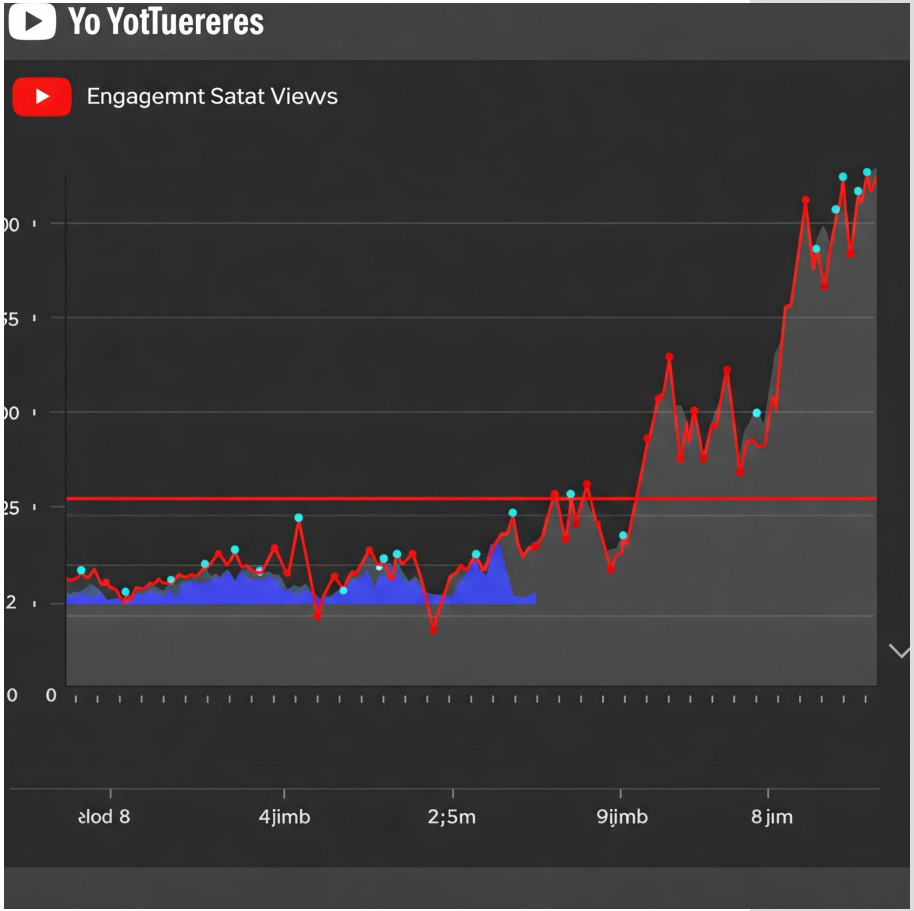
Synthetic Experience Metrics

- A Synthetic Experience Metric is a metric built out of QoE metrics time-normalized in some way to allow comparison and improvements. It can include combining multiple categories of QoE, or showing the likelihood of a bad experience.

- Simple: % of video starts > 1000ms 7DA
- Complex: Survival-based

- Hire a Data Scientist to build these

Synthetic Experience Metrics - What is an SLO?



SLO

Synthetic Experience Metrics

Metric SLOs *either* protect user experience, *or* allow you to set improvement goals.

- Rebuffer is an excellent metric to set **defensive** SLOs on, but it's difficult to set improve goals on due to changes in audience and networking conditions.
- You end up having to spend a lot of time comparing things by hand, not a great use of engineering time.

    - MTBR slightly better
    - Likelihood to rebuffer within x minutes the best…

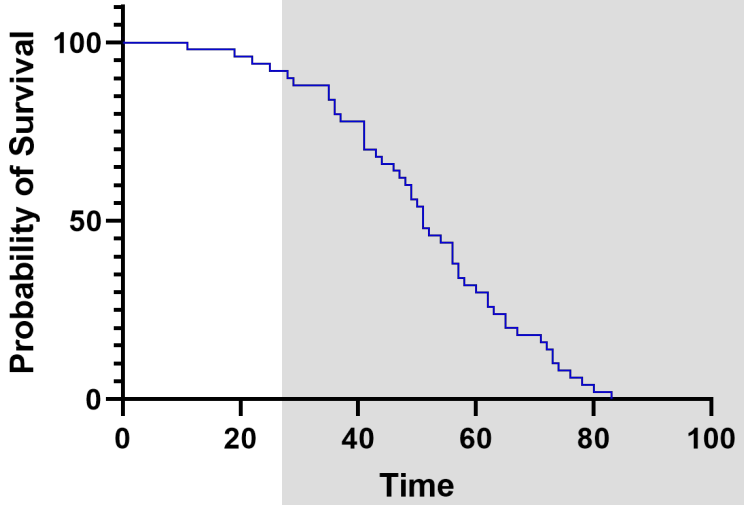
Synthetic Experience Metrics p2

Metrics based on Survival Curves let you make intelligent tradeoffs with other Defensive Metrics.

- Survival metrics measure the likelihood of surviving a Bad User Experience event over time (latent startup event, rebuffer, fatal error)

- Create a human perceptible threshold
  - >1s is “slow”
  - % of playbacks per day going over that threshold

- Sessions with a Slow Playback Rate



It's Latency All the Way Down

Latency is *the* best proxy for Experience, and the secret of improving the long tail, on device and server-side.

- You don't *need* Survival metrics, you can set appropriate user threshold % playbacks (ie: % of playbacks start > 1000ms)
- You may find that diagnosing shifts in your latency metrics take a lot of engineering time - tooling becomes important.
- Opportunities for AI analysis of trends and opportunities.
- Problems will show in latency before they show in experience. Improvements will make your product magic.
- 1) Set Improve OKRs 2) Task engineers with improving performance and not just building features

Experience Codecs

Start Fast, Play On.

Not everything has a clear user experience metric:

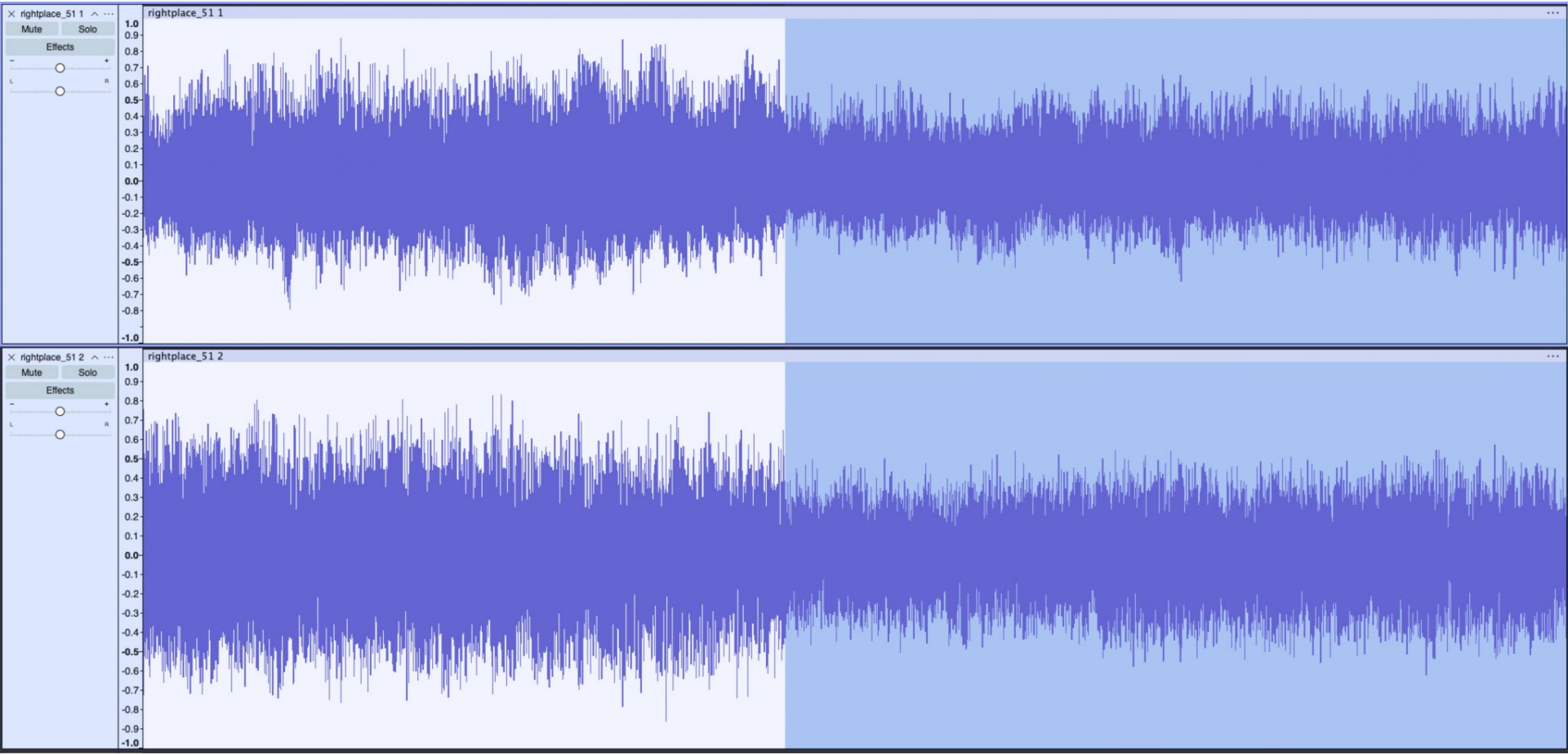
ie: Audio quality…

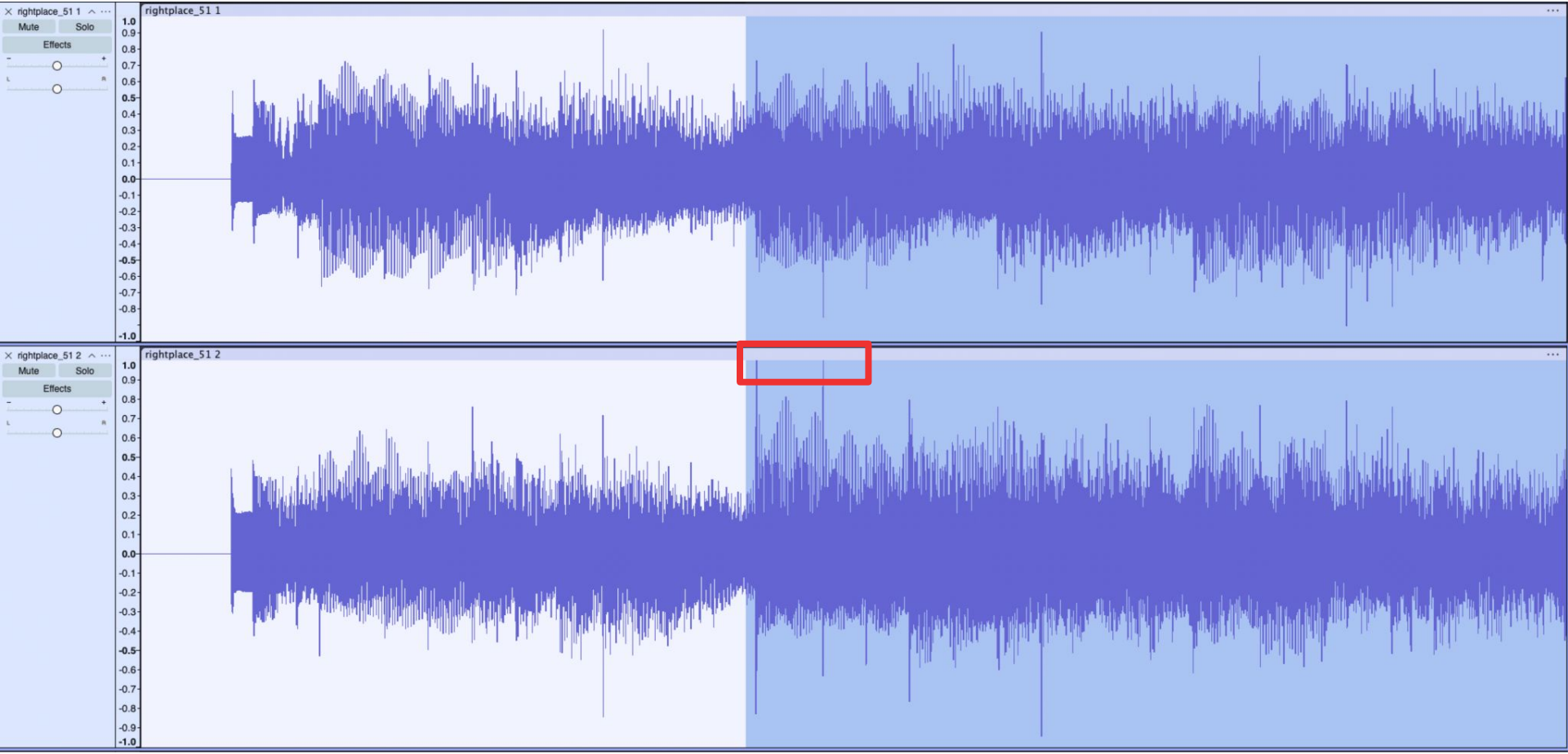
Stable Volume: Viewers vs Creators - Steven Robertson

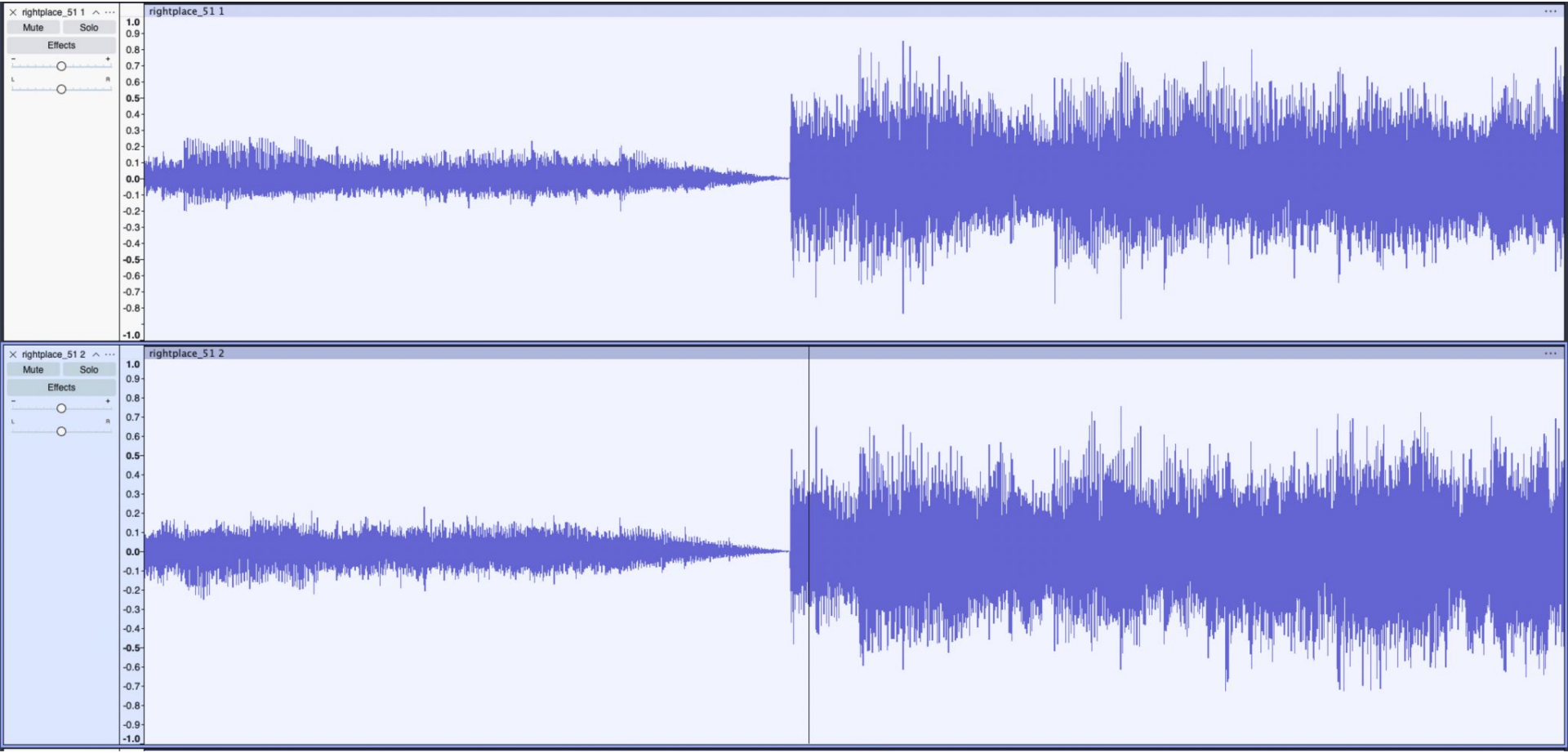
Loudness management is an accessibility feature

Loudness variation was
a top user complaint

-14dB on Main App


rightplace_51 1
Mute
Solo
Effects
rightplace_51 1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-1.0
rightplace_51 2
Mute
Solo
Effects
rightplace_51 2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0


rightplace_51 1
Mute
Solo
Effects
rightplace_51 1
rightplace_51 2
Mute
Solo
Effects
rightplace_51 2


rightplace_51 1
Mute
Solo
Effects
rightplace_51 1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-1.0
rightplace_51 2
Mute
Solo
Effects
rightplace_51 2
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0

Stable Volume

Decreased viewer complaints

Increased overall usage

Loud ads = less revenue

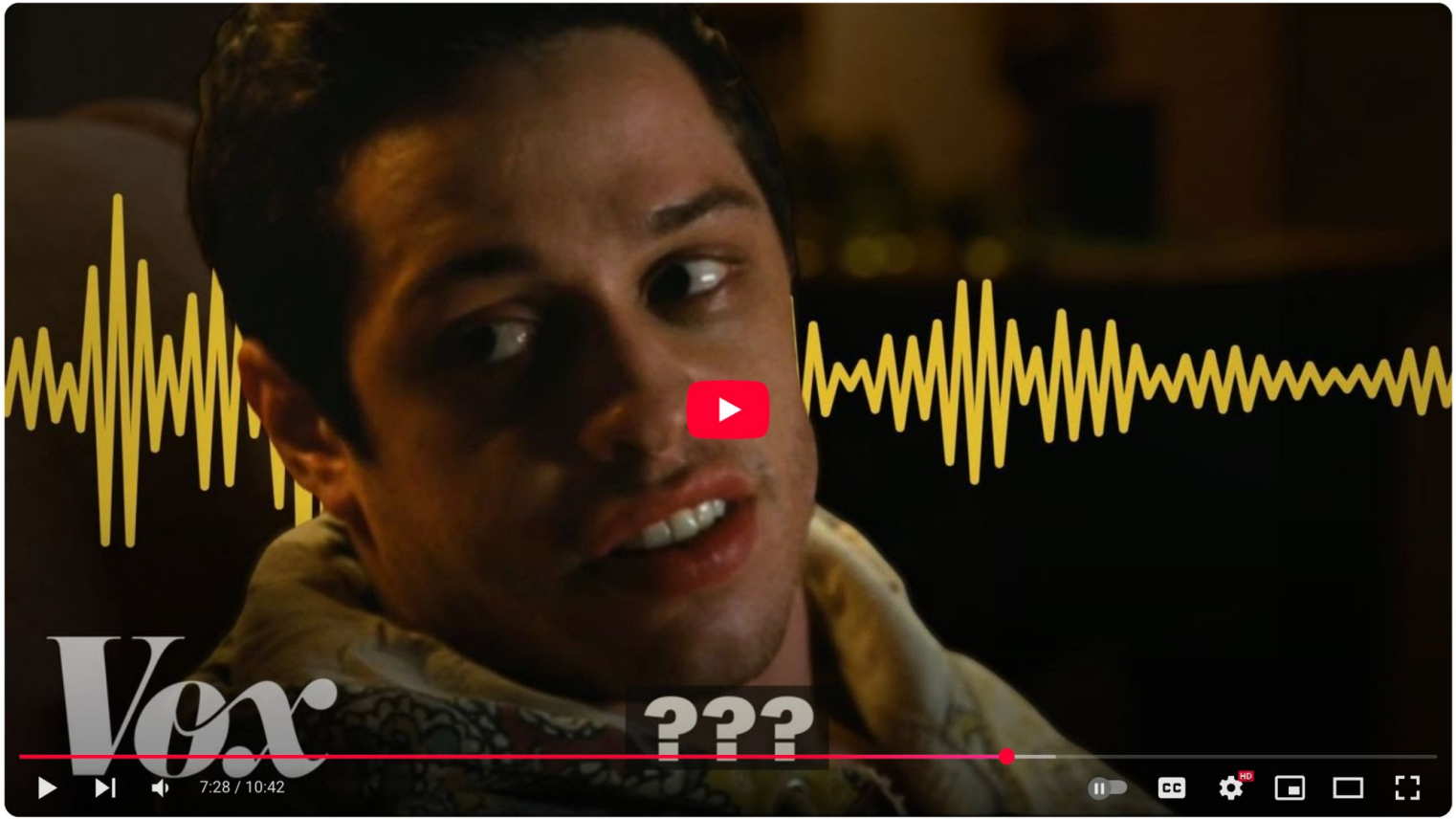
Increased *creator* complaints

YouTube ruined all my videos - UNLESS you disable this setting

Production Advice
37.6K subscribers
Subscribe

869 Share Download Clip

Why do viewers and creators disagree?

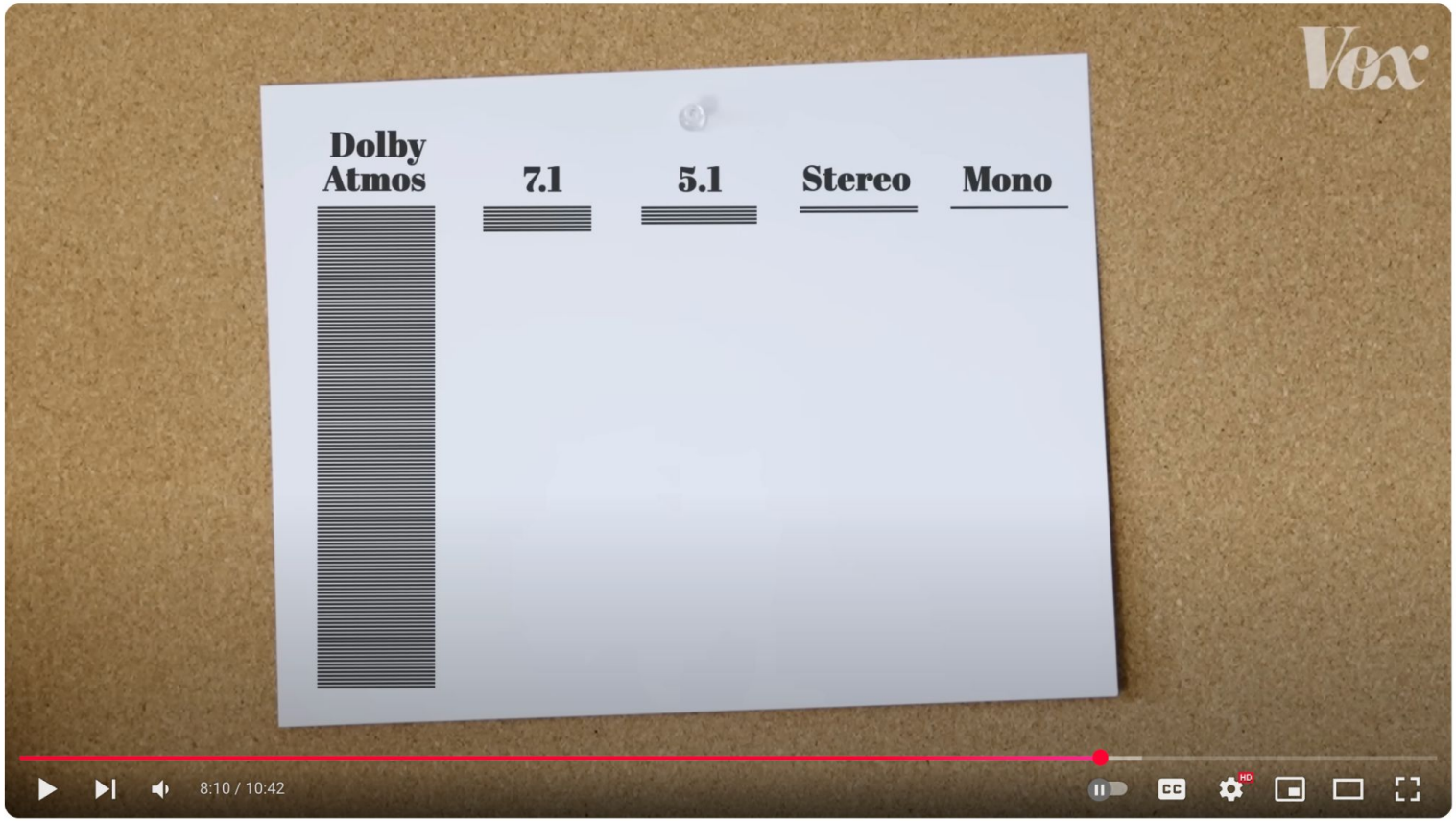


Why we all need subtitles now

Vox
12.5M subscribers

YouTube

vox why we all need subtitles



Why we all need subtitles now

Vox
12.5M subscribers

Subscribe

506K

Share

Download

Save

Why don't movies look like *movies* anymore?

Patrick Tomasso
167K subscribers

Subscribe

39K

Share

Save

I think this is wrong.

I think this is ***scurvy.***

The cure for scurvy has been found several times.

The cure for scurvy has been **lost** several times.

"The lime juice issued by the British was almost totally ineffective, probably because it came into contact with copper (which oxidizes vitamin C) when it was manufactured."

https://www.md-a.co/p/solving-scurvy

"This is because at the same time, the advent of the steam engine made voyages much shorter, meaning that sailors usually no longer spent enough time between ports to develop scurvy."

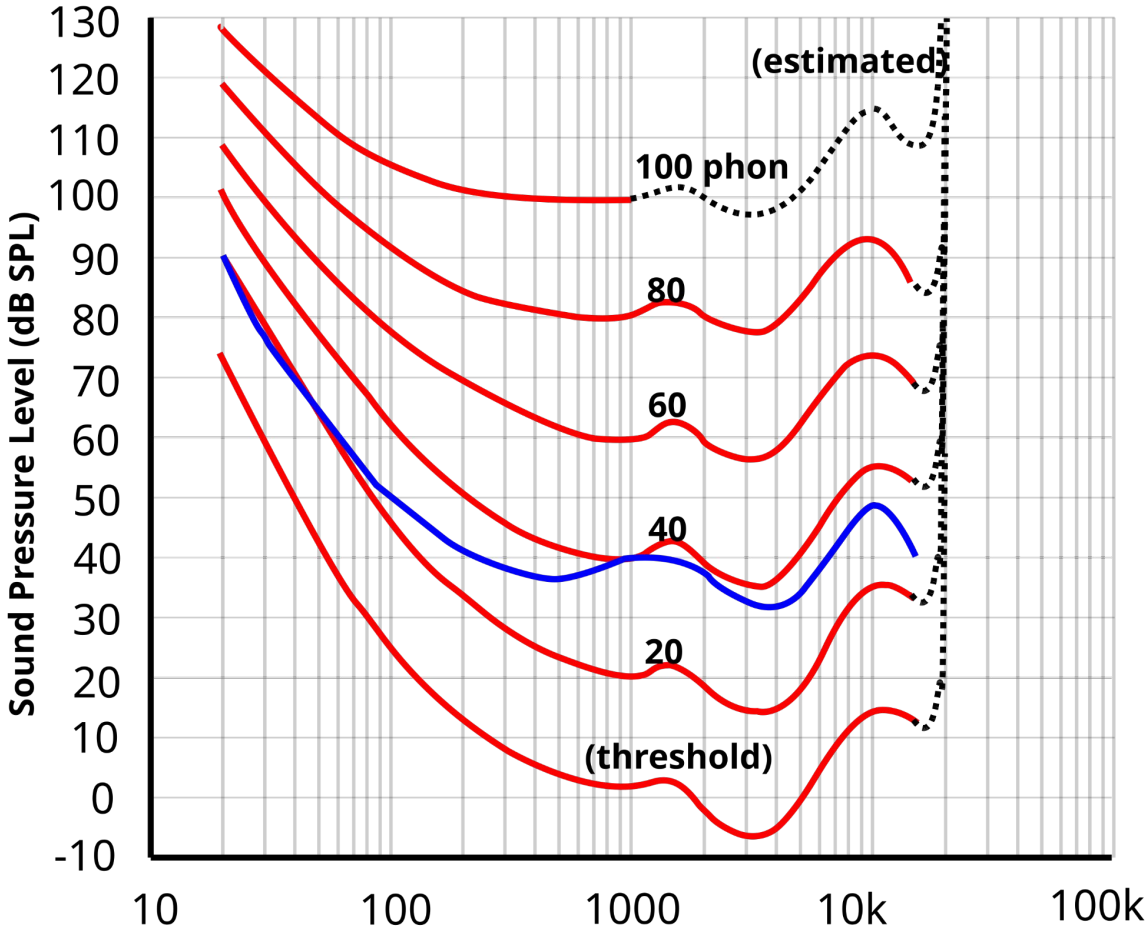
https://www.md-a.co/p/solving-scurvy

Context is key.

Identify the empathy gap.

Table D.1 Reference Sound Pressure Level

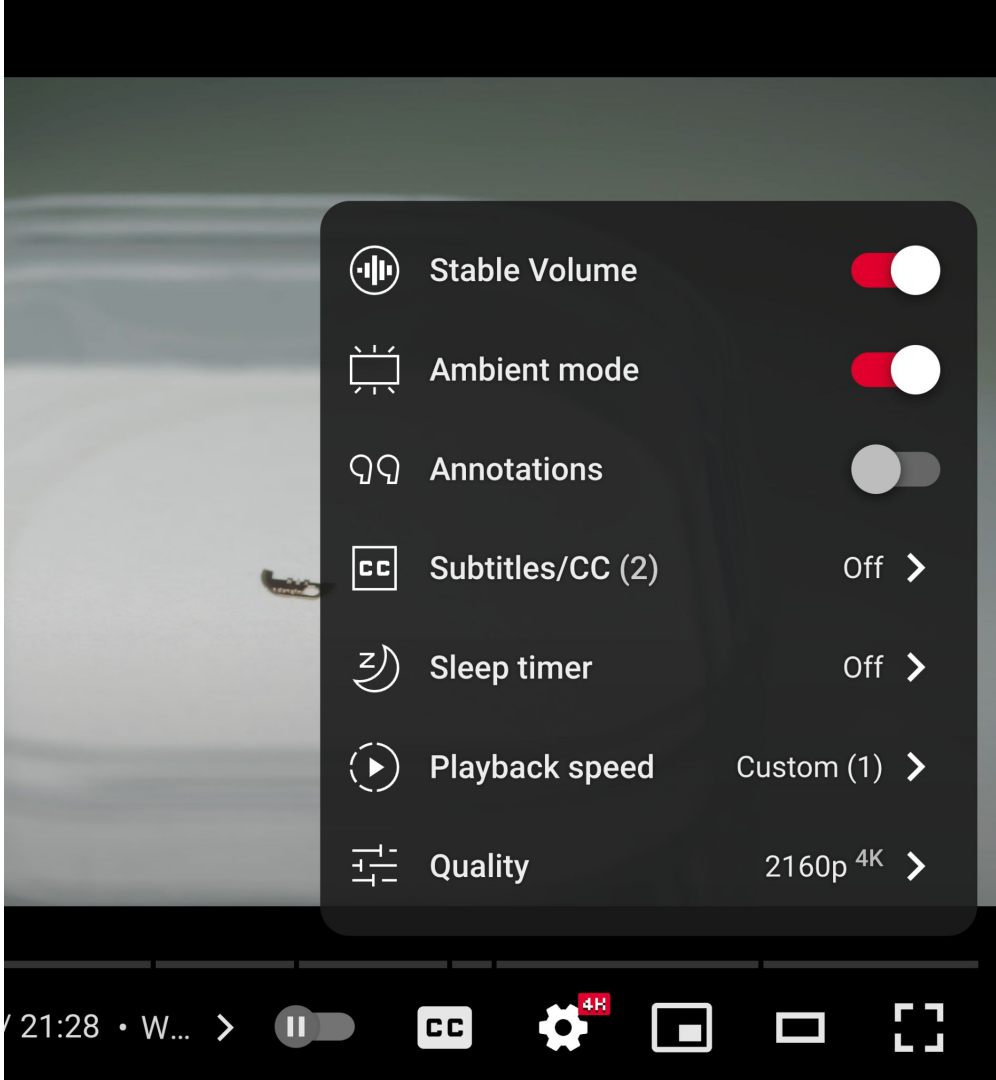
Categories	**Room Volume in Cubic Feet**	**SPL in dB re 20 μN/m^2**
I, II (Mix rooms)	> 20,000	85*
	10,000 < 19,999	82
	5,000 < 9,999	80
	1,500 < 4,999	78
	< 1,499	76
III (Edit rooms sometimes used for mixing)	Depends on room usage. For editing purposes, may be controlled by the editor for use with the material at hand. For final program mixing, follow the recommendations for categories I, II above.	
IV (Booths, vans)	< 1,500	76
V (Headphones)	Use 2 cc. Coupler and set 400 Hz level to 78 dB.	
* Per SMPTE RP 200 [6]		



Equal-loudness contours (red) (from ISO 226:2003 revision)
Original ISO standard shown (blue) for 40-phons

Viewers vs creators: a false choice

Personal,
transparent,
user controllable.



Client-controlled DRC is not new (especially in Living Room).

Closed-loop iteration enables a more complete solution.

We'll keep iterating.

Thank you!