RTC @scale conference 03/20/24

BERYL - full software AEC (by Sriram Srinivasan & Hoang Do)

☆ META did 20% reduction in "No Audio" or "Audio device reliability" issue on iOS & Android

☆ 15% reduction in P50 mouth to ear latency on Android

☆ Revamp of Audio processing stack core for whatsapp Instagram & messenger

☆ Very diverse userbase
- Different kinds of handsets
- Different Geography
- Noisy conditions
- Both high end & Low end phones (more than 20% low end ARMv7)

Leave No user behind

☆ Based on telemetry & user feedback Meta decided to tackle
(1) ECHO & (2) Audio Quality under low bitrate network

☆ High end devices use ML to suppress echo

☆ To accomodate low end devices which cannot run ML, a baseline solution for echo cancellation is needed

Welcome BERYL

- ☆ Beryl replaces webRTC's AEC3, AECM on all devices
- ☆ Interestingly users experiencing echo issues are also on low end devices which cannot run ML
- ☆ Meta's scale is too large.
  - High end phones have hardware AEC.
  - Low end phones do not
  - Stereo & spatial audio only possible in s/w.
  - H/w only does mono AEC

AECM = AEC Mobile (low compute)

AEC3 = higher compute full AEC

- ☆ Beryl was needed because AECM either leaves lot of residual echo or degrades quality of double-talk
  - AECM
    - Not scalable for Millions of users
    - Quality not best
- ☆ Beryl AEC = Low compute - DSP based s/w AEC
  - Lite mode for low end devices
  - Full mode for high end

- Both modes adaptive vs. AECM being simple echo suppressor
- Near instant adaptation to changes
- Better double talk performance
- Multi-channel capture & render 16kHz & 48kHz
- Tuned using 3000 music + speech (mono + stereo) on 20+ devices
- CPU usage increase of less than 7% compared to webRTC AEC

★ Beryl Components

① Delay Estimator

★ Clock drift when using external mic & speaker as they do not share common clock

★ Delay estimator, estimates delay between far-end reference signal (speaker) & near end capture signals (mic)

★ Beryl full mode can handle non-causal delays (-ve delay)

★ Can handle delay up to 1 sec

② Linear AEC

- Estimate echo & subtract from capture signal
- Beryl AEC is normalized least mean squared (NLMS) frequency domain dual filter algo
- One fixed & one adaptive filter
- Coefficients can be copied between filters
  - relative difference in the powers of error signal between two filters and input mic signal
  - coupling factor between echo estimate & error signal
- Adaptation step size is configurable & depends on coherence between mic & reference signals, power and SNR
- Great double talk performance compared to WebRTC AEC

③ Acoustic Echo Suppressor (AES)

- Non linear distortions are introduced by amplifiers before speaker and after microphone
- AES removes this non-linear echo (residual echo)
- AES removes stationary echo noise, distortion, applies perceptual filtering & ambient noise matching

☆ Implementation

- ☆ Reduce memory, CPU & latency
- ☆ Synchronization needed due to work on audio from input & output devices from different threads
  1. mutex in functions (Good safety but worse realtime performance)
  2. Low level locks on shared data structures
  3. Thread safe low level data structures (ok safety, great realtime performance)
- ☆ Neon on ARMv7 & ARM64
- ☆ AVX2 on Intel
- ☆ CPU <110 % of webRTC AEC