Internet Engineering Task Force (IETF)
Request for Comments: 7874
Category: Standards Track
ISSN: 2070-1721

JM. Valin
Mozilla
C. Bran
Plantronics

May 2016

WebRTC Audio Codec and Processing Requirements

Abstract

This document outlines the audio codec and processing requirements for WebRTC endpoints.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7874.

Table of Contents

1. Introduction

An integral part of the success and adoption of Web Real-Time Communications (WebRTC) will be the voice and video interoperability between WebRTC applications. This specification will outline the audio processing and codec requirements for WebRTC endpoints.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

3. Codec Requirements

To ensure a baseline level of interoperability between WebRTC endpoints, a minimum set of required codecs are specified below. If other suitable audio codecs are available for the WebRTC endpoint to use, it is RECOMMENDED that they also be included in the offer in order to maximize the possibility of establishing the session without the need for audio transcoding.

WebRTC endpoints are REQUIRED to implement the following audio codecs:

- Opus [RFC6716] with the payload format specified in [RFC7587].
- PCMA and PCMU (as specified in ITU-T Recommendation G.711 [G.711]) with the payload format specified in Section 4.5.14 of [RFC3551].

- [RFC3389] comfort noise (CN). WebRTC endpoints MUST support [RFC3389] CN for streams encoded with G.711 or any other supported codec that does not provide its own CN. Since Opus provides its own CN mechanism, the use of [RFC3389] CN with Opus is NOT RECOMMENDED. Use of Discontinuous Transmission (DTX) / CN by senders is OPTIONAL.

- the 'audio/telephone-event' media type as specified in [RFC4733]. The endpoints MAY send DTMF events at any time and SHOULD suppress in-band dual-tone multi-frequency (DTMF) tones, if any. DTMF events generated by a WebRTC endpoint MUST have a duration of no more than 8000 ms and no less than 40 ms. The recommended default duration is 100 ms for each tone. The gap between events MUST be no less than 30 ms; the recommended default gap duration is 70 ms. WebRTC endpoints are not required to do anything with tones (as specified in RFC 4733) sent to them, except gracefully drop them. There is currently no API to inform JavaScript about the received DTMF or other tones (as specified in RFC 4733). WebRTC endpoints are REQUIRED to be able to generate and consume the following events:

Event Code	Event Name	Reference
0	DTMF digit "0"	[RFC4733]
1	DTMF digit "1"	[RFC4733]
2	DTMF digit "2"	[RFC4733]
3	DTMF digit "3"	[RFC4733]
4	DTMF digit "4"	[RFC4733]
5	DTMF digit "5"	[RFC4733]
6	DTMF digit "6"	[RFC4733]
7	DTMF digit "7"	[RFC4733]
8	DTMF digit "8"	[RFC4733]
9	DTMF digit "9"	[RFC4733]
10	DTMF digit "*"	[RFC4733]
11	DTMF digit "#"	[RFC4733]
12	DTMF digit "A"	[RFC4733]
13	DTMF digit "B"	[RFC4733]
14	DTMF digit "C"	[RFC4733]
15	DTMF digit "D"	[RFC4733]

For all cases where the endpoint is able to process audio at a sampling rate higher than 8 kHz, it is RECOMMENDED that Opus be offered before PCMA/PCMU. For Opus, all modes MUST be supported on the decoder side. The choice of encoder-side modes is left to the implementer. Endpoints MAY use the offer/answer mechanism to signal a preference for a particular mode or ptime.

For additional information on implementing codecs other than the mandatory-to-implement codecs listed above, refer to [RFC7875].

4. Audio Level

It is desirable to standardize the "on the wire" audio level for speech transmission to avoid users having to manually adjust the playback and to facilitate mixing in conferencing applications. It is also desirable to be consistent with ITU-T Recommendations G.169 and G.115, which recommend an active audio level of -19 dBm0. However, unlike G.169 and G.115, the audio for WebRTC is not constrained to have a passband specified by G.712 and can in fact be sampled at any sampling rate from 8 to 48 kHz and higher. For this reason, the level SHOULD be normalized by only considering frequencies above 300 Hz, regardless of the sampling rate used. The level SHOULD also be adapted to avoid clipping, either by lowering the gain to a level below -19 dBm0 or through the use of a compressor.

Assuming linear 16-bit PCM with a value of +/-32767, -19 dBm0 corresponds to a root mean square (RMS) level of 2600. Only active speech should be considered in the RMS calculation. If the endpoint has control over the entire audio-capture path, as is typically the case for a regular phone, then it is RECOMMENDED that the gain be adjusted in such a way that an average speaker would have a level of 2600 (-19 dBm0) for active speech. If the endpoint does not have control over the entire audio capture, as is typically the case for a software endpoint, then the endpoint SHOULD use automatic gain control (AGC) to dynamically adjust the level to 2600 (-19 dBm0) +/- 6 dB. For music- or desktop-sharing applications, the level SHOULD NOT be automatically adjusted, and the endpoint SHOULD allow the user to set the gain manually.

The RECOMMENDED filter for normalizing the signal energy is a second-order Butterworth filter with a 300 Hz cutoff frequency.

It is common for the audio output on some devices to be "calibrated" for playing back pre-recorded "commercial" music, which is typically around 12 dB louder than the level recommended in this section. Because of this, endpoints MAY increase the gain before playback.

5. Acoustic Echo Cancellation (AEC)

It is plausible that the dominant near-to-medium-term WebRTC usage model will be people using the interactive audio and video capabilities to communicate with each other via web browsers running on a notebook computer that has a built-in microphone and speakers. The notebook-as-communication-device paradigm presents challenging

echo cancellation problems, the specific remedy of which will not be mandated here. However, while no specific algorithm or standard will be required by WebRTC-compatible endpoints, echo cancellation will improve the user experience and should be implemented by the endpoint device.

WebRTC endpoints SHOULD include an AEC or some other form of echo control. On general-purpose platforms (e.g., a PC), it is common for the analog-to-digital converter (ADC) for audio capture and the digital-to-analog converter (DAC) for audio playback to use different clocks. In these cases, such as when a webcam is used for capture and a separate soundcard is used for playback, the sampling rates are likely to differ slightly. Endpoint AECs SHOULD be robust to such conditions, unless they are shipped along with hardware that guarantees capture and playback to be sampled from the same clock.

Endpoints SHOULD allow the entire AEC and/or the nonlinear processing (NLP) to be turned off for applications, such as music, that do not behave well with the spectral attenuation methods typically used in NLP. Similarly, endpoints SHOULD have the ability to detect the presence of a headset and disable echo cancellation.

For some applications where the remote endpoint may not have an echo canceller, the local endpoint MAY include a far-end echo canceller, but when included, it SHOULD be disabled by default.

6. Legacy VoIP Interoperability

The codec requirements above will ensure, at a minimum, voice interoperability capabilities between WebRTC endpoints and legacy phone systems that support G.711.

7. Security Considerations

For security considerations regarding the codecs themselves, please refer to their specifications, including [RFC6716], [RFC7587], [RFC3551], [RFC3389], and [RFC4733]. Likewise, consult the RTP base specification for RTP-based security considerations. WebRTC security is further discussed in [WebRTC-SEC], [WebRTC-SEC-ARCH], and [WebRTC-RTP-USAGE].

Using the guidelines in [RFC6562], implementers should consider whether the use of variable bitrate is appropriate for their application. Encryption and authentication issues are beyond the scope of this document.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, DOI 10.17487/RFC3551, July 2003, <http://www.rfc-editor.org/info/rfc3551>.

[RFC3389] Zopf, R., "Real-time Transport Protocol (RTP) Payload for Comfort Noise (CN)", RFC 3389, DOI 10.17487/RFC3389, September 2002, <http://www.rfc-editor.org/info/rfc3389>.

[RFC4733] Schulzrinne, H. and T. Taylor, "RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals", RFC 4733, DOI 10.17487/RFC4733, December 2006, <http://www.rfc-editor.org/info/rfc4733>.

[RFC6716] Valin, JM., Vos, K., and T. Terriberry, "Definition of the Opus Audio Codec", RFC 6716, DOI 10.17487/RFC6716, September 2012, <http://www.rfc-editor.org/info/rfc6716>.

[RFC6562] Perkins, C. and JM. Valin, "Guidelines for the Use of Variable Bit Rate Audio with Secure RTP", RFC 6562, DOI 10.17487/RFC6562, March 2012, <http://www.rfc-editor.org/info/rfc6562>.

[RFC7587] Spittka, J., Vos, K., and JM. Valin, "RTP Payload Format for the Opus Speech and Audio Codec", RFC 7587, DOI 10.17487/RFC7587, June 2015, <http://www.rfc-editor.org/info/rfc7587>.

[G.711] ITU-T, "Pulse code modulation (PCM) of voice frequencies", ITU-T Recommendation G.711, November 1988, <http://www.itu.int/rec/T-REC-G.711-198811-I/en>.

8.2. Informative References

[WebRTC-SEC] Rescorla, E., "Security Considerations for WebRTC", Work in Progress, draft-ietf-rtcweb-security-08, February 2015.

[WebRTC-SEC-ARCH]
Rescorla, E., "WebRTC Security Architecture", Work in Progress, draft-ietf-rtcweb-security-arch-11, March 2015.

[WebRTC-RTP-USAGE]
Perkins, C., Westerlund, M., and J. Ott, "Web Real-Time Communication (WebRTC): Media Transport and Use of RTP", Work in Progress, draft-ietf-rtcweb-rtp-usage-26, March 2016.

[RFC7875] Proust, S., Ed., "Additional WebRTC Audio Codecs for Interoperability", RFC 7875, DOI 10.17487/RFC7875, May 2016, <http://www.rfc-editor.org/info/rfc7875>.

Acknowledgements

This document incorporates ideas and text from various other documents. In particular, we would like to acknowledge, and say thanks for, work we incorporated from Harald Alvestrand and Cullen Jennings.

Authors' Addresses

Jean-Marc Valin
Mozilla
331 E. Evelyn Avenue
Mountain View, CA 94041
United States

Email: jmvalin@jmvalin.ca

Cary Bran
Plantronics
345 Encinial Street
Santa Cruz, CA 95060
United States

Phone: +1 206 661-2398
Email: cary.bran@plantronics.com